Looking at Online Language Comprehension with the Visual World Paradigm: A Systematic and Cross-Linguistic Review of Three Decades*

İpek Pınar Uzun[1] Alper Kumcu[2]

ORCID: [1]0000-0003-3103-0758, [2]0000-0003-0844-3562

[1]Ankara Üniversitesi, Dilbilimi Bölümü, Ankara, Türkiye

[2]Hacettepe Üniversitesi, Mütercim ve Tercümanlık Bölümü, Ankara, Türkiye

[1]pinarbekar@gmail.com [2]alperkumcu@gmail.com

(Received 9 May 2025; Accepted 8 July 2025)

Abstract

The Visual World Paradigm (VWP) is one of the most widely used experimental paradigms in psycholinguistic research for studying online word and sentence processing. Over the past three decades, a wealth of research has shown that eye movements can reveal much about the systematic and temporal relationship between fixations on visual information and online language processing. The present systematic review takes stock of the current state of the VWP as an experimental paradigm, outlining critical issues that should guide future research. In this paper, we provide a systematic overview of the role of language functions using the VWP in speech comprehension from a cross-linguistic perspective. Our findings contribute to the VWP literature by evaluating existing methodological approaches and outlining directions for future research. The review reveals that although methodological sophistication has increased, especially through the adoption of methods such as LMMs and GAMMs, a significant gap remains in linguistic diversity.

Keywords: Visual World Paradigm, comprehension, word processing, sentence processing, word recognition

* This study was supported by the Scientific and Technological Research Council of Türkiye (TÜBITAK), 1002-A Rapid Support Program [No:122K664], Coordinator: İpek Pınar Uzun, Researcher: Alper Kumcu.

https://doi.org/10.18492/dad.1692960
Dilbilim Araştırmaları Dergisi, 2025/2, 185–223.

Görsel Dünya Paradigması ile Süreç-içi Dil Anlamaya Bakış: Otuz Yıllık Sistematik ve Dillerarası Bir Derleme

Öz

Görsel Dünya Paradigması (GDP), süreç-içi sözcük ve tümce işlemeyi incelemek amacıyla psikodilbilimsel araştırmalarda en yaygın ve en bilgilendirici şekilde kullanılan deneysel paradigmaların başında gelmektedir. Tanenhaus ve diğerleri (1995) çalışmasından bu yana geçen otuz yıllık süreçte, çok sayıda araştırma, göz hareketlerinin görsel bilgilere yönelik bakışlarla çevrim-içi dil işlemleme arasındaki sistematik ve zamansal etkiye ilişkin önemli veriler sunduğunu göstermiştir. Bu sistematik derleme çalışması, çok yönlü bir deneysel paradigma olarak GDP'nin mevcut durumunu ortaya koymakta; son otuz yıla dönük bir değerlendirme yaparak, gelecekteki araştırmaları yönlendirebilecek önemli sorunları öne çıkarmaktadır. Bu çalışmada, GDP'nin konuşma anlama bağlamında dilsel işlevlerle ilişkisini dillerarası bir bakış açısıyla sistematik biçimde ele almayı amaçlamaktayız. Sonuçlar, alanyazının LMM ve GAMM gibi zamana duyarlı analizlerin benimsenmesiyle yöntemsel olarak gelişmiş olmasına karşın dilsel çeşitlilik konusunda önemli bir boşluğun var olmaya devam ettiğini ortaya koymaktadır.

Anahtar sözcükler: Görsel Dünya Paradigması, anlamlandırma, sözcük işleme, tümce işleme, sözcük tanıma

1 Introduction

The Visual World Paradigm (VWP) is a powerful experimental method, typically implemented with eye-tracking, in online language comprehension. In a typical VWP experiment, language users are presented with a "visual world" containing potential references (i.e., scenes of real objects, pictures, or words) to language. Participants typically view visual stimuli on a four-by-four grid while listening to auditory input, and their eye movements are monitored (see Figure 1). As seen in Figure 1, VWP represents two displays: a four-object (left panel) for 'look-and-listen' or "look-listen-and-click-response" tasks, and a single printed word (right panel) for "click-on-the-word-you-heard" tasks. The visual stimuli on the grid may have different referential links (i.e., semantic, phonological, syntactic). A central question concerns when and under what conditions individuals fixate on these objects. Thus, fixations on these referents and the competition between them provide valuable information about the early stages of language processing, cognitive mechanisms, and the integrative decision between language and vision. There are also several variants of the VWP with a blank screen instead of a grid (e.g., Spivey et al., 2000) or with more complex visual information (e.g., Kamide,

Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Brown-Schmidt, Byron, & Tanenhaus, 2004).

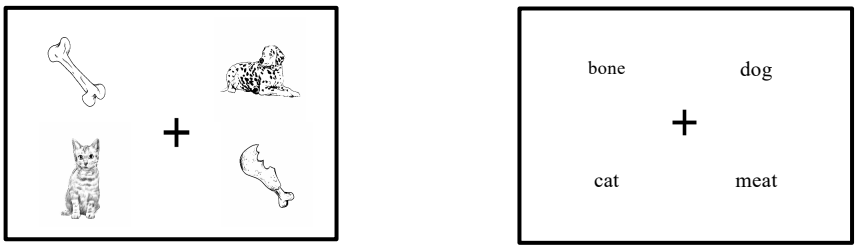

Figure 1. VWP paradigms include two displays: a four-object display without borders (left; "look-and-listen" or "look-listen-and-click-response") and a printed word (right; "click on the word you heard").

Irrespective of specific experimental tasks, VWP has significantly reshaped psycholinguistics over the last 30 years, encompassing all linguistic levels and developmental stages. The present study aims (i) to systematically review the most critical studies, (ii) to discuss the implications of these studies, (iii) to identify gaps in the previous literature, and (iv) to suggest directions for future work. The basic theoretical assumption is based on the claim that linguistic input directs visual attention dynamically and incrementally, reflecting the moment-by-moment integration of visual context and language. This allows us to extract basic stages of cognitive processes and mechanisms of language comprehension.

Cooper's (1974) study is the first to use the VWP as we know it today. The results showed that listeners would spontaneously direct their attention to pictures in a visual display that were mentioned in a story, even when they were not instructed to do so. This is the basic assumption of the VWP: participants naturally look at objects that have been mentioned in speech. The world is full of visual stimuli, making the VWP a highly appropriate method for addressing issues related to language processing, such as lexical access, ambiguity resolution, the influence of context, selective attention and sequencing.

Subsequent VWP studies building on Cooper (1974) adopted more controlled picture arrangements. The referents in a typical VWP experiment can be classified into three categories: the target (the intended referent), competitors (items that are phonologically, semantically, or visually similar to the target), and distractors (unrelated items that serve as visual/auditory controls with no phonological or semantic relation to the target items). For example, in Allopenna et al. (1998), a *beaker* was a target, *beetle* was a cohort competitor, *speaker* was a rhyme competitor, and *carriage* served as an unrelated distractor. Participants viewed four images while they listened to spoken instructions like "Pick up the

beaker; now put it below the diamond." This design allowed for a precise assessment of referential competition during listening. The strength of the VWP lies in its capacity to track such competition in online analysis through eye-tracking data. As a result, it can reveal automatic processing tendencies even when participants are unaware of them. Such tendencies refer to cognitive or attentional biases occurring during online language processing, generally displayed by gaze durations of fixations on visual referents. Notably, VWP studies have shown that fixation patterns are tightly time-locked to unfolding speech (Cooper, 1974; Tanenhaus et al., 1995). The following sections outline our systematic and cross-linguistic review approach (Section 2), evaluate VWP studies on online language processing in online spoken word recognition and comprehension (Section 3), and discuss the current contributions and limitations of the methodology in VWP studies (Section 4).

2 Methods

We followed the PRISMA (Page et al., 2021) guidelines to systematically address our hypotheses and research aims. We searched for previous literature across multiple databases (i.e., Web of Science, PubMed, Scopus, Google Scholar, and APA) with additional studies using VWP which are identified through citation tracking. Our investigation covered all records up to 29 April 2025. Eligible studies (including research articles, conference proceedings, presentations, and dissertations) focused on speech comprehension in healthy adults and encompassed a range of publication languages. Studies were screened using predefined inclusion and exclusion criteria. Figure 2 presents a detailed PRISMA flowchart, outlining each stage of study identification, screening, eligibility, and inclusion. Reasons for exclusion at each stage (e.g., scope mismatch, methodological issues) are fully documented for transparency. Study selection followed a three-step process: (1) initial title and abstract screening to remove irrelevant studies; (2) full-text review for relevance to criteria such as adult language processing, use of VWP, and methodological clarity; and (3) quality assessment based on methodological rigour and clarity of experimental design. The selected studies were classified according to linguistic units and language functions at the word and sentence level during speech comprehension (see Table A1 in the Appendix for details).

The inclusion criteria were as follows: (1) studies targeted healthy adult participants, as the review focuses on language processing in mature cognitive systems; (2) they employed speech comprehension tasks using auditory, visual, auditory-visual, or orthographic modalities; and (3) study findings were categorised by linguistic unit, level of processing (word or sentence), task type, and language. Studies focusing solely on general memory capacity, without addressing its interaction with linguistic processing, were excluded. Based on

these criteria, 128 studies were deemed eligible. We recorded the reference details, experimental procedures, tasks, designs, linguistic focus and diversity, statistical methods, techniques, analysis and language profiles of participants for each study. Figure 2 shows the systematic and cross-linguistic review process used to identify and extract relevant VWP studies. As the figure illustrates, the main aim of the present study is to provide a comprehensive analysis of key VWP studies from the past three decades, focusing on methodological improvements and cross-linguistic interactions. First, we outline our systematic study selection approach; then we evaluate findings related to word- and sentence-level processing; finally, we conclude with implications for further research. In line with its objectives, this review places particular emphasis on cross-linguistic diversity, systematically identifying and comparing VWP studies conducted in a wide range of languages beyond English. In this regard, this review synthesised notable studies and highlighted underexplored dimensions of online language processing in VWP studies, thereby revealing an underlying focus for future work.

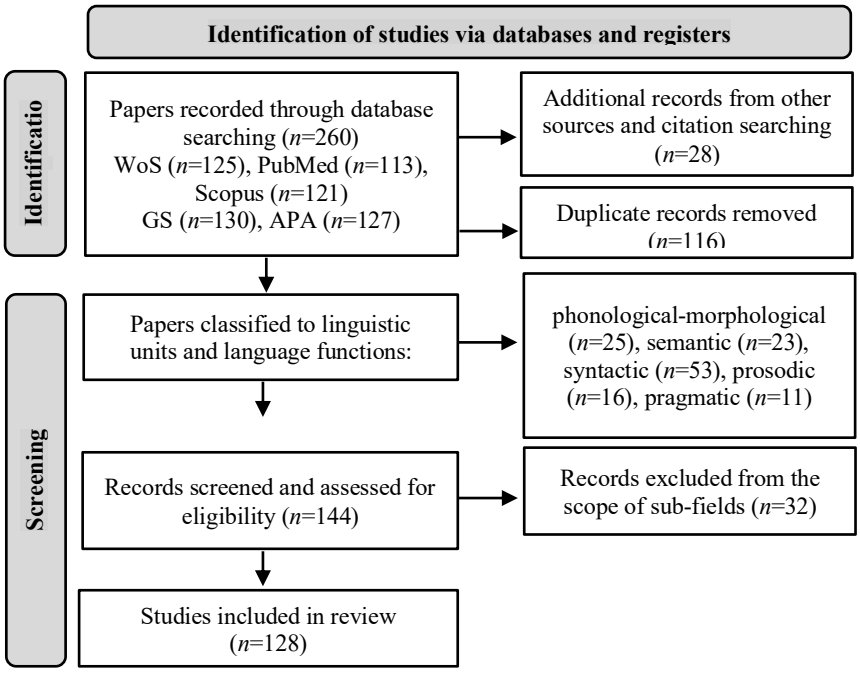

Figure 2. PRISMA flow diagram of literature search and selection

3 Language comprehension: Word and sentence-level processing

In reviewing previous studies using the VWP, we assessed experimental design, tasks, findings, and cross-linguistically based theoretical implications. To present a balanced perspective, we highlight both the strengths and, where relevant, the limitations of each study. The basic roles of language functions in word- and sentence-level processing are discussed separately, covering linguistic aspects (i.e., phonological, morphophonological, semantic, morphosyntactic, prosodic, and pragmatic) of speech comprehension. The first section focuses on word-level processing, including lexical access, phonemic variation, phonological awareness, lexical stress, and ambiguity, which are the core components of morphological and semantic processing in spoken word recognition. The following section addresses syntactic ambiguity, referential domains, prosodic focus, conversational elements, referents, interference, and connectives, thus examining how syntax, prosody, and pragmatics shape speech comprehension.

3.1 Word-level processing

3.1.1 Phonological and morphological processing: Lexical access and alternation

This section reviews studies that have used the VWP to investigate phonological and morphological processes, with an emphasis on methodological approaches and their implications. Speech comprehension studies on phonological processing has often focused on whether subphonetic variations affect lexical access. Notably, McMurray, Tanenhaus, and Aslin (2002, 2009) used the VWP to explore how voice onset time (VOT) along a continuum affects lexical activation during spoken word recognition in American English. They examined the gradient effects of within-category phonetic variation. In particular, McMurray et al. (2009) found that such variation, contrary to phoneme-level inhibition, might lead to lexical garden-path effects, especially when manipulating voiced and voiceless consonants. Their results suggest that subphonetic mismatches in initial word positions may influence both semantic priming and lexical processing dynamics.

The pioneering study by Allopenna, Magnusson, and Tanenhaus (1998) introduced an eye-movement paradigm to investigate the manifestation of competitor effects for objects whose names rhymed with the target, as predicted by continuous mapping models. The study clearly explained the correlation between activation levels modelled by the continuous mapping model, such as the TRACE model (McClelland & Elman, 1986), and fixation probabilities observed in eye-tracking data. By incorporating experimental conditions (e.g., referent – *cup*, cohort – *beetle*, rhyme – *speaker*, and unrelated – *carriage*) into

a VWP design, Allopenna et al. (1998) aimed to examine the basic nature of lexical competition during experimental tasks on visual referencing. This approach enabled researchers to connect theoretical modelling of empirical observations more directly to cognitive processes. In addition, the results showed a competition for lexical activation between rhymes and cohorts, with participants more likely to initiate an eye movement to a rhyme or cohort than to a non-competitor.

Morphophonological alternations (McQueen & Viebach, 2007; Salverda & Tanenhaus, 2010; Farris-Trimble & McMurray, 2018; Schreiber & McMurray, 2019) and phonological awareness (Dahan, Magnusson, & Tanenhaus, 2001a; Marquis et al., 2020; Gregg, Inhoff, & Li, 2023) have been investigated in many languages based on listeners' use of auditory-visual knowledge during online word recognition. Dahan et al. (2001a) reported striking results on word recognition in the phonological processing using multi-frequency tasks. In multi-frequency tasks, the presentation of targets and distractors in visual scenes varied according to their phonological roles. Dahan et al. (2001b) also conducted an eye-tracking study on how subcategorical phonetic inconsistencies influence lexical competition in spoken word recognition. Their results showed the sensitivity of lexical access to phonetic information. Later, McQueen and Viebach (2007) investigated spoken word recognition by manipulating the syllable structure of Dutch with monosyllabic and bisyllabic targets (e.g., *buffel*, *buffalo*) in VWP using a typical phonological competitor (e.g., *buffer*, *buffer*) and two unrelated distractors (e.g., *kind*, *child*). Overall, the results showed that participants' average fixation durations became longer when the competitor was phonologically violated compared to the distractors and the target.

A related study by Cutler, Weber, and Otake (2006) used a look-and-listen methodology to examine lexical access and phonetic mapping asymmetries between Japanese L1 and American English L2 listeners. Similar cross-linguistic findings were reported by Bruggeman and Cutler (2024), showing that L2 listeners performed better in phonetic mapping tasks, suggesting an acoustic-phonetic dominance of the L1. In two recent studies, Freeman and Marian (2022) investigated cross-linguistic activation in word recognition by comparing L1 and L2 English listeners with L1 Spanish speakers. Results showed strong activation effects tied to the initial position of words during visual target recognition. Desmeules-Trudel and Zamuner (2023) also explored bilingual word recognition, comparing Canadian English L2 listeners with native French L1 speakers. Across two eye-tracking experiments manipulating nasal combinations in word-final positions, early bilinguals outperformed both late bilinguals and L1 speakers in recognising phonetic details of French vowel nasalisation. Similarly, Dailey, Straboni, and Peperkamp (2023) examined allophonic nasal-oral vowel variation using an offline gating task and the VWP. Their eye-tracking results

showed that French listeners used recognition strategies comparable to those of English listeners (see also Soto & Schmid, 2024).

A recent study by Burness, McMullin, and Zamuner (2019) used artificial languages to investigate morphophonological learning, specifically manipulating sibilant harmony. They created two artificial languages with 36 noun bases, each appearing in three forms: a bare form and two prefixed forms (/ʃo/ and /fe/). One language included regressive sibilant harmony, while the other did not. Results suggested that long-range harmony processing in spoken word recognition is influenced by the visual position of the target and competitor. In another recent study, Bao, Arnhold, and Järvikivi (2024) examined novel word learning in Mandarin Chinese using different phonological contrasts (consonant, tone, or both), including homophones and nonhomophones. Native and non-native English speakers were assessed on their ability to discriminate novel word pairs and select target items while listening to auditory stimuli. Results highlighted the impact of phonological contrast type and language background on word learning performance.

In summary, the above studies suggest that morphophonological alternations, phonological awareness, competition, prediction, and lexical access input from L1 (Dahan et al., 2001a; McMurray et al., 2002, 2009; Farris-Trimble & McMurray, 2018; Apfelbaum, Klein-Packard, & McMurray, 2021; Kutlu et al., 2024; Xu et al., 2025; Zhao et al., 2024) and L2 (e.g., Cutler et al., 2006; Burness et al., 2019; Freeman & Marian, 2022; Bruggeman & Cutler, 2024; Berghoff & Bylund, 2024; Soto & Schmid, 2024; Bramlett & Wiener, 2025) reflect the direction of listeners' attention during spoken word recognition. These factors interact with the integration of auditory, visual, and orthographic cues, collectively guiding attention and influencing gaze patterns and fixation durations during word recognition.

3.1.2 Semantic processing: Spoken word recognition and meaning

Several works have investigated incremental semantic integration, particularly at the sentence level, focusing on speech interpretation and lexical ambiguity resolution. However, less is known about semantic processing during word recognition. Sedivy et al. (1999) examined how contextual contrasts in prenominal English adjectives affect semantic interpretation in the visual world. Participants processed target objects (e.g., a *tall glass*) under different visual contexts, including no-contrast conditions (e.g., a *single glass*) and contrast conditions (e.g., a *smaller glass*). Findings indicated that contextual information influences semantic processing early in comprehension. Similarly, a well-known study in Dutch by Dahan and Tanenhaus (2005) demonstrated a visual similarity effect in a word-matching task involving referents (e.g., a *snake*) and visually similar competitors (e.g., a *rope*). Gaze and fixation patterns over time supported

the linking hypothesis (Allopenna et al., 1998), showing a strong relationship between fixation probabilities and lexical activation levels.

Psycholinguistic approaches to the link between spoken word recognition and semantic processing vary depending on how lexical access and output representations are conceptualised, often drawing on different aspects of memory capacity. Duñabeitia et al. (2009) explored the role of visual representations by comparing the semantic processing of abstract and concrete words in Spanish. Their findings showed that participants focused more on objects associated with abstract words than with concrete ones in the VWP, suggesting a heightened reliance on visual context during the processing of abstract language.

Another study on semantic processing (Hadar et al., 2016) investigated the effect of listeners' working memory capacity under different listening conditions. They used a VWP task similar to that of McQueen and Viebach (2007), manipulating the competitor and target according to their phonological and semantic representations in the onset and offset positions of the syllables. Hadar et al. (2016) found that listeners with high working memory load showed delayed target discrimination between the spoken target and the phonological competitor. Similarly, Zamuner et al. (2016) highlighted the importance of participants' memory load during speech comprehension. They compared English (non)words with visual referents in two experiments. The results suggest the importance of auditory production, as recognition of spoken words was faster for nonwords. For example, when a participant is processing a mispronounced target (e.g., /*vup*/) and the visual scene contains two nonword images (e.g., /*tup*/ and /*pem*/), they explain that the gaze patterns of adult participants can be unclear in choosing the correct target word in the VWP.

Semantic information and memory load strongly affect spoken word recognition, comprehension, and retrieval. Two studies have examined how semantic information interacts with word memory using different tasks: Kumcu and Thompson (2020) investigated word retrieval in a blank VWP. Participants were presented with four words on a two-by-two grid, and their recognition memory was tested while looking at a blank screen. The results indicated that participants looked more at the previous, but now empty, locations of the probed word. They looked even more when retrieving less imageable words, suggesting a compensation for low imageability. Later, another study (Kumcu & Thompson, 2022) compared the effect of the actual locations of words in the visual world and their semantic locations (e.g. *bird* - top location) in a similar recognition memory task across two experiments. The results showed that the semantic locations of the words influenced memory retrieval even when spatial information was not important for accurate performance of the task. A recent study (Rehrig et al., 2022) investigated the link between meaning and visuospatial mapping by conducting five eye-tracking experiments using visual world tasks. The results suggest that visual attention to an object's affordance in

visual scenes interacts strongly with grasping, as it is influenced by semantic information in real scenes and memory load.

The studies reviewed above have examined semantic processing in word recognition and lexical ambiguity using various tasks within the VWP, including orthographic similarity (Shen, Qu, & Li, 2016), listening and reading judgments (Novick, Thompson-Schill, & Trueswell, 2008; Degen & Tanenhaus, 2016; Shkuropatskaya et al., 2020; Kaplan, Raju, & Arunachalam, 2021; Sarrett & McMurray, 2025), referent identification (Arias-Trejo & Plunkett, 2010), pronoun resolution (Kim & Grüter, 2021), lexical interference (Prystauka et al., 2024), and tasks involving categorisation, recognition, and prediction (Ahn & Song, 2023; Corps et al., 2022; Hintz et al., 2024; Li & Qu, 2024; Soroli, 2024; Chen et al., 2025). Collectively, these studies provide strong evidence for an interaction between semantic processing and memory load. Although some have found no such interaction, findings generally indicate that semantic representations, especially in the context of lexical ambiguity and meaning, are influenced by memory demands.

3.2 Sentence-level processing

3.2.1 Syntactic processing: Ambiguity resolution, referents and thematic roles

VWP studies have extensively explored syntactic processing during online processing of spoken language comprehension. The paradigm has been used to investigate a range of syntactic patterns: trace violations (Tanenhaus et al., 1995; Dahan et al., 2001b), syntactic ambiguity and resolution (Altmann & Kamide, 1999; Trueswell et al., 1999; Arnold et al., 2000, 2004; Sturt, 2003; Chambers, Tanenhaus, & Magnusson, 2004; Brown-Schmidt, Byron, & Tanenhaus, 2005; Brown-Schmidt & Konopka, 2008; Coco & Keller, 2015; Stern et al., 2019; Yao et al., 2024; Abdel-Latif et al., 2025; Alzahrani, 2025), pronoun resolution and referential domains (Spivey et al., 2002; Runner, Sussman, & Tanenhaus, 2006; Kaiser & Trueswell, 2008; Ellert, 2013; Arslan, Bastiaanse, & Felser, 2015; Cunnings et al., 2017; Fotiadou, Muñoz, & Tsimpli, 2020; Han et al., 2021; Degen & Pophristic, 2023), verb constraints (Bruggeman et al., 2025; Zhu & Grüter, 2025), object naming (Altmann, 2004; Altmann & Kamide, 2007; Belke & Meyer, 2007; Fallon & Pylkkänen, 2024), garden-path structures (Dahan & Tanenhaus, 2004; Knoeferle & Crocker, 2006; Farmer, Cargill, & Spivey, 2007; Hintz, Meyer, & Huettig, 2017), as well as case marking, aspect, argument structure, and thematic roles (Frenck-Mestre et al., 2019; Huang & Snedeker, 2020; Karaca et al., 2024; Özsoy et al., 2023; Gómez Vidal, 2024; Minor et al., 2023; Fuchs & Sekerina, 2025).

In a widely cited VWP study, Tanenhaus et al. (1995) demonstrated how visual context affects rapid online processing investigating various types of

syntactic cues. Results showed that participants prefer nonlinguistic cues, particularly those aligned with behavioural goals, to establish referents during the early stages of online language comprehension. Accordingly, their findings emphasized how fast the visual context can shape the online processing of language comprehension and interpretation. Overall, these studies highlight the importance of integrating grammatical constraints with contextual information. This suggests us a more comprehensive account of how syntactic and nonlinguistic cues interact in online language processing.

Syntactic complexity and ambiguity resolution are fundamental topics of considerable ongoing interest in psycholinguistics. The VWP allows researchers to investigate moment-to-moment processing via syntactic ambiguity resolution (i.e., garden-path structures, late closure effects, relative clause (RC) attachments, and working memory capacity). Previous work (Tanenhaus et al., 1995; Altmann & Kamide, 1999; Arnold et al., 2000) has shown how participants use nonlinguistic knowledge (i.e., world knowledge, visual perception that influences linguistic processing independent of language-specific cues, and contextual familiarity) from visual contexts to disambiguate syntactic structures. Experimental designs of such studies are usually designed according to the participants' tendency to look at the potential target or competitor in a visual context. For example, if a participant reads or hears a sentence such as "A girl is reading a book with a friend", the participant will tend to process a girl or a book due to the syntactic role of the prepositional phrase with a friend. On the other hand, studies show different linguistic patterns depending on the syntactic structure and the experimental layouts (i.e., auditory or visual displays) during online processing. For example, the canonical subject-verb-object (SVO) order in English and the canonical subject-object-verb (SOV) order in Turkish may indicate different interpretations due to the syntactic positions and constraints of the verb and object.

As is known, RC attachment is one of the most studied topics in syntactic disambiguation. In sentences with RC attachments, such as "The man saw the girl on the street who was wearing a red hat," the clause can refer to either the man or the girl, leading to two possible interpretations. RC attachments are a key focus in VWP research across different languages (e.g., Coco & Keller, 2015 for English; Pozniak, Hemforth, & Scheepers, 2018 for French; Stern et al., 2019 for Spanish; Scheepers, 2003 for German). These studies have generally relied on the online processing costs of cognitive load when the parser encounters syntactic complexity, distance effects or disambiguation. When using VWP to investigate RC attachment preferences, studies focus on how participants process the visual context cues to disambiguate syntactic structures in online processing.

Disambiguation studies investigate the comprehension of garden-path sentences, which are characterised by initial misleading syntactic cues. Traditional approaches to syntax-first models generally propose that the parser

can select a single structure and reanalyse the syntactic mechanisms using semantic and contextual information (Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Trueswell & Tanenhaus, 1994; Tanenhaus et al., 1995) during online processing of syntactic resolution. On the other hand, the constraint-based models suggest that both semantic, statistical, and contextual information contribute to the reanalysis and resolution of garden-path sentences (Elman, Hare, & McRae, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994). From this perspective, Farmer et al. (2007) conducted three experiments measuring visual context identification of magnitude and gradient effects in garden-path sentences and resolution of syntactic ambiguity in American English. The results support constraint-based models of syntactic reanalysis processing. Overall, these findings support the idea that individual differences may affect how difficult it is to process sentence reanalysis and garden-path structures (Christianson et al., 2001; Ferreira, Bailey, & Ferraro, 2002).

Syntactic processing has a direct influence pronoun resolution in relation to referential domains. Previous studies (Sturt, 2003; Arnold et al., 2004; Brown-Schmidt et al., 2005; Brown-Schmidt & Konopka, 2008; Kaiser & Trueswell, 2008; Runner et al., 2006; Ellert, 2013; Arslan et al., 2015; Cunnings et al., 2017; Fotiadou et al., 2020; Han et al., 2021; Hert, Järvikivi, & Arnhold, 2024; Kim & Xiang, 2024) show that parsers rapidly integrate memory and visual context to resolve referents. For example, Arnold et al. (2004) used eye-tracking experiments to investigate the use of grammatical gender in different visual contexts for pronoun resolution. The results suggest that grammatical gender can quickly interfere with cognitive language mechanisms and memory-based processes for retrieving linguistic information from a visual context. From a bilingual processing perspective using VWP during listening, Arslan et al. (2015) investigated the processing difficulties in evidentiality resolution of monolingual L1 and L2 Turkish speakers. The researchers compared the use of processing strategies between L1 and L2 Turkish speakers in terms of online interpretation of evidentiality markers (i.e., direct [-DI] and indirect [-mIş]). They addressed a gap between neurolinguistic and psycholinguistic research on evidentiality. Results indicated reduced sensitivity to both the semantic and pragmatic roles of direct evidentiality markers in both early (heritage) and late L2.

The online processing of grammatical gender morpheme agreement has received much attention in recent years. Studies conducted by examining different types of languages have generally found higher search rates for target items when this item is preceded by a gender-relevant adjective (Grüter, Lew-Williams, & Fernald, 2012; Dussias et al., 2013; Hopp & Lemmerth, 2016; Fuchs, 2021; Loerts, Wieling, & Schmid, 2013; Lundquist & Vangsnes, 2018; Lee, 2023; Johannessen et al., 2024). For instance, a gender agreement study in German articles (masculine *der*, feminine *die*, neuter *das*) (Hopp, 2013) found that slower looking percentages were obtained for the masculine condition when

it is compared to the neuter and feminine conditions. Hopp and Lemmerth (2016) further investigated the asymmetry of gender information in German articles. They found that participants looked more quickly at adjective-relevant target items than at other gender cues. Later, findings in Lundquist and Vangsnes (2018) for the anticipatory use of VWP in Norwegian on the language-internal asymmetries between three gender cues revealed differences between language dialects. In accordance with these results, speakers of two dialects of Norwegian could not use the masculine articles when the competitor was constructed with the feminine articles. Speakers of three Norwegian dialects also performed good in using all gendered articles. A recent study by Fuchs (2023) examined masculine, feminine, and neuter gender categories in Polish. A recent study by Fuchs (2023) examined masculine, feminine, and neuter gender categories in Polish. Using the VWP, the experiment presented auditory stimuli in carrier phrases paired with three colours (green, red, and blue). Monolingual Polish speakers successfully applied gender to adjectives during online speech comprehension across all gender categories.

Second, the interpretation of argument structure and thematic roles is one of the most widely studied topics that plays an important role in the interplay between visual context interpretation and language processing (Carreiras et al., 1997; Kuchinsky, Bock, & Irwin, 2011; Branigan, Pickering, & Cleland, 2000). Previous studies suggest that the interpretation processing of thematic roles and argument structure can be affected by visual contexts such as referential domains in a visual scene in online language comprehension (Gleitman et al., 2007; Huang & Snedeker, 2009). Further, studies show that parsers use verb information while analysing syntactic structure to predict anticipatory eye movements for objects in the visual scenes and/or contexts. Hintz et al. (2017) conducted three eye-tracking experiments in order to measure verb-mediated anticipatory gaze in Dutch via using listening and language comprehension tasks. Their results indicated that functional relations strongly predict anticipatory eye movements in long and short visual scenes. Accordingly, if the target item is predictable, participants hear a sentence like “The man peels the apple”; however, they hear “The man draws the apple”, if it is not predictable.

In summary, the studies on sentence-level processing provide us with a robust framework for syntactic complexity and ambiguity resolution. Previous research highlights the involvement of cognitive mechanisms in syntactic reanalysis, prediction, and comprehension processes.

3.2.2 Prosodic processing: Lexical stress and prosodic focus

VWP studies of prosodic processing have addressed prosodic violations in general. Researchers have focused on lexical stress (Dahan, Tanenhaus, & Chambers, 2002; Reinsch & Weber, 2012; Jesse, Poellmann, & Kong, 2017;

Kong & Jess, 2017; Zahner, Kutscheid, & Braun, 2019), prosodic focus (Brown et al., 2015; Henry, Hopp, & Jackson, 2017; Ge et al., 2021; Müller et al., 2021; Perdomo & Kaan, 2021, Caldas, 2025; Zhang & Zhang, 2025), the interaction between semantic, syntactic, and pragmatic processing (Mulders & Szendrői, 2016; Kırcalı, Uzun, Aydın, 2021), and speech rate (Huettig & Guerra, 2019; Fernandez et al., 2020). Online processing studies investigating prosodic processing at the level of lexical stress have led to the hypothesis that listeners show different patterns of interpretation based on the accentuation of noun phrases (NP). This hypothesis predicts that when listeners encounter a deaccented NP, they are more likely to integrate this sentence with the most prominent component in the visual context, as opposed to an accented NP.

Previous VWP research has aimed to elucidate the cognitive mechanisms underlying the interpretation of prosodic units. Dahan et al. (2002) investigated the prosodic roles of accent and referents using lexical competitors in American English, such as *candy* vs. *candle* (see also Bramlett & Wiener, 2025 for L2 stress perception). While their first experiment compared the placement of lexical accent of NPs according to referents in a discourse-based context, the next experiment revealed this interaction with (non-)focal referents. Findings indicated that when listeners looked at the competitor, they were more likely to fixate on a new entity if the target was accented rather than deaccented. Zahner et al. (2019) presented critical findings on the resolution of lexical stress in German across different fundamental frequency (f0) orientations in different pitch accents, namely medial-peak and early-peak accents. In the first experiment, they measured the effects of lexical stress judgment and identification tasks on the prosodic perception of different pitch accents. They then compared fixations on the target and competitor with initial and penultimate stress manipulations by presenting the experimental items using auditory-visual and orthographic techniques. Zahner et al. (2019) found that participants were more likely to fixate on an initially stressed competitor when the penultimate-stressed target was produced with an early-peak accent. This contrasted with conditions where the f0 peak aligned with the stressed syllable. As a result, the inclusion of high, unstressed syllables temporarily caused them to be perceived as stressed, directly affecting the process of lexical activation. Such findings contribute to a deeper understanding of how prosody shapes the cognitive mechanisms underlying speech perception and word recognition.

Prosodic focus effects in quantifier ambiguity resolution have been observed in a limited number of studies. Kırcalı et al. (2021) examined the interaction of prosodic focus in a semantic-syntactically oriented task using two VWP experiments in Turkish. While the first experiment focused on the SOV, the second experiment investigated the object-subject-verb (OSV) order. Findings from the first experiment showed the effect of prosody and syntax on the interpretation of sentences with *a(n)-every* versus *every-a(n)* order. Conversely,

the results of the second experiment showed less prosodic influence on syntactically-semantically oriented structures.

Recent VWP studies have also focused on the effect of speech rate on anticipatory gaze at different speech rates such as 3.5, 4.5, 5.5 and 6.0 syllables per second. Fernandez et al. (2020) provided evidence for the effect of speech rate on the prediction of filler gap dependency comprehension. Their first experiment examined this prediction difference across the lifespan with younger and older adults, while the second experiment compared L1 English speakers and L2 English speakers with L1 German. They found similar evidence for anticipatory eye movements from both experiments to the findings in Huettig and Guerra (2019), in which the effect of speech rate was examined in Dutch neuter gender cues. Results highlight the significant role of speech rate for anticipatory eye movements and group-level differences for filler gap dependency.

Overall, studies on the segmental and suprasegmental cues in visual context guide spoken word recognition during online processing. In general, listeners tend to visually discriminate between target and competitor lexical accent violations according to their primary and secondary stress assignments. The research contributes to a deeper understanding of the integrated processing of prosody with semantics and syntax that modulate the cognitive mechanisms and temporal dynamics of language comprehension.

3.2.3 Pragmatic processing: Conversation, referent, and interference

Current VWP accounts of pragmatic processing often centre on the Gricean maxim of quantity (Grice, 1975), which concerns the provision of an appropriate amount of information in context. Many VWP studies (e.g., Engelhardt, Bailey, & Ferreira, 2006; Huettig et al., 2006; Barr, 2008; Köhne-Fuetterer et al., 2021; Malarski, Jankowiak, & Dziubalska-Kołaczyk, 2025) suggest that online comprehension of spoken or written discourse involves the interpretation of pragmatic cues related to this conversational norm. Engelhardt et al. (2006) provided empirical support through three eye-tracking experiments. In their production task, speakers often overdescribed referents. However, in a subsequent listening judgment task, listeners did not consistently view overdescriptions as inferior to shorter ones. They used VWP in the final experiment and examined how listeners processed these over-described expressions. The findings indicate only moderate online adherence to Gricean principles, suggesting that pragmatic processing during comprehension is shaped by both linguistic input and broader cognitive mechanisms. These results underscore the complexity of communicative behaviour and the nuanced role of pragmatic norms in language processing.

A related body of research using VWP has examined the role of conversation, referent, filler, and interference contexts using a variety of methodologies. Barr

and Seyfeddinipur (2010) have measured how English listeners explore the visual context described by a speaker in a mouse-tracking experiment. The results showed that when listeners heard a speaker include the filler word *umm* in the utterance, listeners produced a significantly longer anticipatory gaze than when processing *umm* presented in a noise context. The study in question showed that the expectation induced by the presence of fillers depends on the specific characteristics of the listeners. Another study (Brown-Schmidt, Gunlogson & Tanenhaus, 2008) focused on speech and interference processing. They investigated the role of production and perception of wh-question structures such as "What's above the cow with shoes?" using a target language game in unscripted conversations. These findings support earlier evidence that shared knowledge between speakers significantly influences initial language processing. They also demonstrate that direction of common ground effects varies sentence type and the level of interaction in the conversation.

Finally, Coco, Keller, and Malcolm (2016) examined the role of visual context and memory by conducting two experiments with different anticipation effects in real-world scenes. Real-world scenes, which provided contextual cues in Experiment 1, facilitated anticipatory eye movements. When the target object was absent from the scene, participants tended to infer and direct their gaze to contextually relevant regions. Experiment 2 investigated whether contextual inference, as observed in Experiment 1, requires the simultaneous presence of the scene or whether participants can rely on memory representations instead. Participants were again presented with the same real-world scenes as in Experiment 1, except that the scene disappeared before participants heard the accompanying sentence. The main question was whether participants would still make anticipatory eye movements based on their memory of the scene and its contextual information, even in the absence of visual input at the time of hearing the sentence. Individuals were shown to rely on the use of global scene representations to make contextual inferences in anticipatory speech processing.

The question of how discourse and pragmatic interaction are induced by a mapping between language and the mental representation of images has been investigated in a limited number of VWP studies (e.g., Altmann & Kamide, 2009; Brown-Schmidt, 2009; Wei et al., 2019). These studies often used the auditory-visual channels to convey linguistic information. However, they differed in terms of language use and comprehension. Specifically, two studies focused on monolingual participants (Altmann & Kamide, 2009; Brown-Schmidt, 2009), while one study involved bilingual participants (Wei et al., 2019). Altmann and Kamide (2009) conducted two experiments to examine the effect of online discourse-pragmatic mapping in English within different visual scenes. In the first experiment, participants heard either the structure "The woman will put the glass on the table" or "The woman is too lazy to put the glass on the table". Participants then heard "The woman will pick up the bottle and carefully pour

the wine into the glass" in a similar context, with no change in the visual scene. In a second experiment, a similar context was used, but the visual scene was removed before the auditory stimuli began. They showed a strong interaction between the dynamic updating of the mental representation and the online mapping of language.

Another study in English (Brown-Schmidt, 2009) investigated the referential domains and interpretation of ambiguity resolution in in-and-out conversation using different dialogue settings. Results from three different eye-tracking experiments suggest that partner-specific interpretation may be more likely in interactive dialogue settings, with the number of trials and stimulus features potentially influencing this phenomenon. Finally, the VWP was used in Wei et al. (2019) for L1 Dutch and L1 Chinese participants to measure causal coherence relations in two eye-tracking experiments. In their first experiment, subject and object connectives were investigated using a listening judgment task in Dutch. For the next experiment, a reading task for underspecified connectives in Chinese was used. Their results showed that the degree of subjectivity embedded in connectives serves as a distinct processing cue for the construction of mental representations, even across two typologically different languages. This suggests that the influence of subjectivity on cognitive processing transcends linguistic boundaries and is applicable across different linguistic structures and systems.

4 Discussion and Conclusion

This systematic review investigates the current state of VWP research, based on studies selected through predefined inclusion and exclusion criteria (see Figure 2) and classified by language level (see Figure 3). Table A1 (see Appendix) summarises the 128 studies included in the review, detailing the research focus, experimental tasks, and languages used. Broadly speaking, these studies examine how healthy adults process various linguistic levels and functions, and how language processing interacts with visual context. They do not focus on clinical populations, such as individuals with dyslexia, aphasia or fluency disorders. While previous comprehensive reviews (e.g., Huettig, Rommers, & Meyer, 2011; Huettig, Olivers, & Hartsuiker, 2011; Salverda, Brown, & Tanenhaus, 2011; Altmann & Kamide, 2004) have addressed the VWP research program from a general perspective, the unique contribution of the present review lies in its cross-linguistic focus. By incorporating a bibliometric approach, we map the distribution of languages represented in VWP research and identify patterns across linguistic and methodological domains.

In line with our hypotheses, this systematic and cross-linguistic review identified various gaps, particularly with regard to the types of experimental tasks employed in research on linguistic diversity and online language processing. For example, previous studies have predominantly used click response tasks due to

the structural features of the languages examined (i.e., phonetics, morphology and syntax), whereas look-and-listen tasks are more likely to be used because of the interpretive language components involved (i.e., prosody, semantics and pragmatics). These linguistic patterns provide a remarkable basis for the variety of experimental tasks, designs, language components and linguistic diversity, setting concrete directions for future research. Although studies using online language processing continue to develop, many languages from non-Indo-European families remain underrepresented, despite the consistent increase in studies using VWP. This might be overcome to some extent by cross-linguistic studies, but findings from first/native languages remain limited. Similarly, task designs specific to experiments display a limited number of variations. While studies focusing on syntactic or morphosyntactic components and their intersections stand out, many other language components remain underrepresented due to complexity-related effects. In this context, these observations in the studies mostly provide important data on the need for more comprehensive experiments, including a wider typological range and methodological diversity, to improve the generalisability and ecological validity of VWP findings during online language processing. In the sections that follow, we discuss our findings by language level, method of analysis, and language, offering a critical assessment of the existing literature and a reference point for future psycholinguistic research.

The results of the preliminary analysis based on Table A1 (see Appendix) show that the number of studies gradually increased over the 30 years from 1994 to 2025. 43.75% of the studies (n = 56/128) were published in the first half of the period between 1994 and 2010, and 56.25% of them (n = 72/128) were published in the second half between 2010 and 2025. Although this result shows that the popularity of the paradigm remains, it is hardly surprising given the general increase in the number of studies in all fields. In terms of language level, the majority of studies, 62.50% (n = 80/128), focused on the sentence level, while 37.50% of them dealt with the word level (n = 48/128). Within the sentence level, the syntactic level was most frequently studied (66.25%, n = 42/64), followed by the prosodic (20%, n = 16/80) and pragmatic (13.75%, n = 11/80) levels. Within the word level, 52.08% (n = 25/48) were at the phonological, phonetic, and morphological level, while 47.92% (n = 23/48) were at the semantic level. The results show that the frequency of studies focusing on different levels of language does not vary much over time to form a clear pattern. In terms of languages, 78 out of 128 studies (60.94%) in the selected corpus included English as the main language, which, as expected, makes English the most dominant language in the literature. Of these, 19 studies included languages other than English (i.e. German, Italian, Spanish, Greek, French). Dutch follows English with 10 studies (7.81%), of which two studies included other languages (i.e. English and Chinese). The following languages, in decreasing order of frequency, are Spanish

(10 studies), Turkish (five studies), German (11 studies), French (five studies), Chinese with its varieties (six studies) and languages with one study each (Greek, Arabic, Norwegian, Polish, Hebrew, Hungarian, Japanese, Korean, Kazakh, Murrinhpatha, Brazilian Portuguese, Finnish). In terms of years, the diversification of languages in the second half, i.e. from 2010 to 2025, compared to the first half, i.e., from 1994 to 2010, is quite striking. While only 3.57% (*n* = 2/56) of the studies included languages other than and/or in addition to English, this figure rises to 70.83% (*n* = 51/72) after 2010. In terms of the tasks used, the click response task seems to be the most common, with 68.75% (*n* = 88/128), followed by the look and listen task with 27.34% (*n* = 35/128). The other tasks account for only 3.91% (*n* = 5/128) of the included studies. Although there is no clear pattern as with the diversification of languages by year, it is important to note that tasks other than click response and look-and-listen were introduced after 2002, suggesting the introduction of different approaches to the paradigm over time. In terms of analysis method, we see the dominance of linear mixed-effects models (LMMs) with 42.97% (*n* = 55/128) and ANOVAs with 32.81% (*n* = 42/128) over other methods. The change in the preferred choice of analysis over time is also very clear: while in the first half of the period up to 2010 ANOVA was used in 62.50% (*n* = 35/56) of the studies, in the second half between 2010 and 2025, LMMs were used in 63.89% (*n* = 46/72) of the studies. Notably, this period also marked the introduction of time series analyses such as generalised additive-mixed models (GAMMs) and growth curve analysis (GCA).

Overall, the studies reviewed support the notion that the VWP is a very powerful and flexible tool in psycholinguistics for investigating language processes, revealing how visual context influences spoken language comprehension and production. As discussed above, several different sets of variables have been manipulated to investigate their interactions and influences on processing. Since the original visual world study (Tanenhaus, 1995), experiments have presented creative variations of the paradigm both in terms of the procedure and tasks and also the visual array. In this vein, we suggest that future VWP studies adhere to open science practices in methodology and data analysis procedures (see Ataman, Çağlar, & Kırkıcı, 2021) for better replicability and comparable results. The development of standards for VWP and eye-tracking methodology could also improve the accuracy of research findings. The other strength lies in the fact that the paradigm allows for millisecond-by-millisecond analysis, paving the way for precise time-course investigations of the interplay between speech and vision, rather than collapsing the data to average fixation/dwell time durations on the critical pictures. In our review, although ANOVAs and LMMs are the most commonly used statistical methods to analyse data from the visual world, it is evident that time course analyses such as GCA and GAMMs are also gaining momentum as a relatively more sensitive analysis to investigate the effect of time on fixations to images in the visual world.

In this respect, the reported analysis, Table A1 (see Appendix) and Figure 3 highlight the gaps in the literature in terms of languages, language levels and tasks; and, in a sense, chart the future course of studies. We underline the fact that, in particular, languages in the Sino-Tibetan, Semitic, Japonic, Uralic, and Turkic language families (see Figure 3) are still critically understudied concerning the VWP, which calls for further research in these languages at all language levels. Our results also suggest the fact that the choice of task and visual context in such studies appears to be quite critical, as it can alter the results of seemingly similar studies, something that has been highlighted in other selective reviews (see Apfelbaum et al., 2021). It is also noteworthy that a growing number of reviews address methodological (Cavicchio, Melcher, & Poesio, 2014; Salverda & Tanenhaus, 2017; Sekerina, Campanelli, & Van Dyke, 2016) and statistical issues in the analysis of VWP time course data, including the introduction of a new approach that combines statistical and computational modelling (Mirman et al., 2008), a critique of the interpretation of time course data with proposals for new analytical models (Magnuson, 2019; McMurray, 2023) and the comparison and interpretation of multiple statistical models (e.g., LMM, GCA, GAMM, cluster-based permutation analysis, and divergence point analysis) in the analysis of time course data (Ito & Knoeferle, 2023). Nevertheless, the visual dimension of the VWP is as significant as the linguistic dimension. Our findings revealed a significant gap, specifically for linguistic diversity. Future work should therefore prioritise the inclusion of understudied languages and methodological transparency.

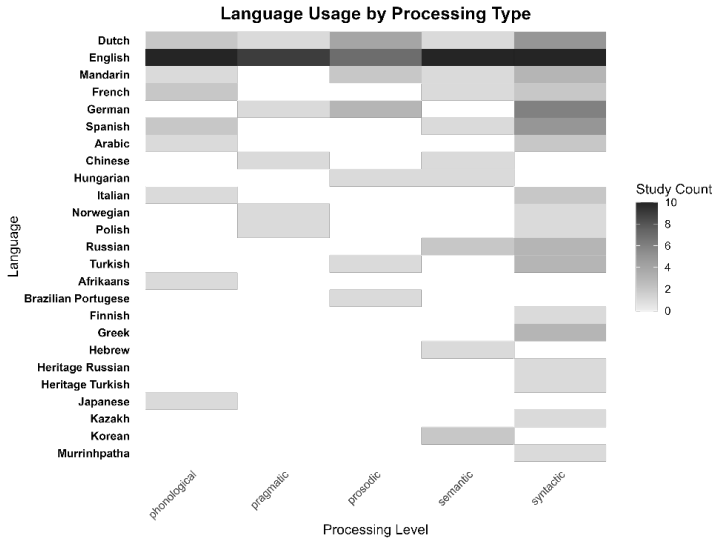

Figure 3. Distribution of processing types across language domains in WVP studies

In conclusion, this review offers a cross-linguistic perspective on VWP research across a range of languages. One limitation is the inclusion of only English-language publications, which may have excluded relevant studies published in other languages. As with any systematic review, our selection criteria may have led to the omission of studies from the past 30 years that nonetheless meet the general scope. Although we did not conduct a meta-analysis due to the diverse methodologies and limited comparability of language functions, future research could address specific questions through a combination of systematic and meta-analytic approaches. This would help to deepen our understanding of spoken language comprehension and production in the VWP. Further studies should also aim to include understudied languages and apply advanced statistical techniques to better capture the complexity and variability of online language processing. Lastly, since VWP is gaining attention as a potential biomarker for clinical populations, such as individuals with dyslexia, aphasia, and fluency disorders, future studies could include clinical populations, which is a limitation of the present review.

Author Contributions: **İpek Pınar Uzun:** Writing – original draft, Conceptualisation, Investigation, Methodology, Data Analysis. **Alper Kumcu:** Writing – review and editing, Data Analysis.
Submission statement and verification: This study has not been previously published elsewhere. It is not under review in another journal.
Conflict of Interest Statement: The authors declare that there are no financial or academic conflicts of interest between themselves or with other institutions, organisations or individuals that may affect this study.
Data Use: The review data and materials are publicly available at the Open Science Framework: https://osf.io/mp637/?view_only=094a69e7789e45f2b1f175f7c11c52f1
Ethical Approval/Participant Consent: There is no need for ethical approval in the study.
Financial Support: This study was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), 1002-A Rapid Support Program (#122K664).

References

Abdel-Latif, K. H. A., Koelewijn, T., Başkent, D., & Meister, H. (2025). Assessment of speech processing and listening effort associated with speech-on-speech masking using the visual world paradigm and pupillometry. *Trends in Hearing, 29*, 1–13. https://doi.org/10.1177/23312165241306091

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World

Paradigm. *Journal of Memory and Language, 121*, Article 104279. https://doi.org/10.1016/j.jml.2021.104279

Allopenna, P. D., Magnusson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2004). Now You See It, Now You Don't: Mediating the Mapping between Language and the Visual World. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye movements and mental representation. *Cognition*, *111*(1), 55–71. https://doi.org/10.1016/j.cognition.2008.12.005

Alzahrani, A. (2025). The predictive processing of number information in L1 and L2 Arabic speakers. *Languages, 10*(2), Article 25. https://doi.org/10.3390/languages10020025

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1-2), 63–80. https://doi.org/10.1016/j.jecp.2009.10.002

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye-tracking. *Cognition*, *76*(1), B13–B26. https://doi.org/10.1016/s0010-0277(00)00073-1

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and the, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578-582. https://doi.org/10.1111/j.0956-7976.2004.00723.x

Arslan, S., Bastiaanse, R., & Felser, C. (2015). Looking at the evidence in visual world: Eye-movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality. *Frontiers in Psychology*, *6*, 1387. https://doi.org/10.3389/fpsyg.2015.01387

Ataman, E., Çağlar, O. C., & Kırkıcı, B. (2021). Dilbilim araştırmalarında açık bilim. *Dilbilim Araştırmaları Dergisi, 32*(2), 149–175. https://doi.org/10.18492/dad.936072

Bao, W., Arnhold, A., & Järvikivi, J. (2024). Phonology, homophony, and eyes-closed rest in Mandarin novel word learning: An eye-tracking study in adult native and non-native speakers. *Applied Psycholinguistics*, *45*, 213–242. https://doi.org/10.1017/S0142716424000031

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455. https://doi.org/10.1080/01690960903047122

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes*, *22*(8), 1178–1211. https://doi.org/10.1080/01690960701461541

Berghoff, R., & Bylund, E. (2024). L2 activation during L1 processing is increased by exposure but decreased by proficiency. *International Journal of Bilingualism, 28*(3), 555–569. https://doi.org/10.1177/13670069231175270

Bramlett, A. A., & Wiener, S. (2025). Individual differences modulate prediction of Italian words based on lexical stress: a close replication and LASSO extension of Sulpizio and McQueen (2012). *Journal of Cultural Cognitive Science, 9*, 55–81. https://doi.org/10.1007/s41809-024-00162-6

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5

Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience*, *30*(1-2), 149–166. https://doi.org/10.1080/01690965.2013.862285

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2004). That's not it and it's not that: The role of conceptual composites in in-line reference resolution. In M. Carreiras & C. Clifton Jr. (Eds.), *On-line sentence processing: ERPS, eye movements and beyond* (pp. 2009−2228). Psychology Press.

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*(2), 292–313. https://doi.org/10.1016/j.jml.2005.03.003

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Bruggeman, L., & Cutler, A. (2024). Lexical recognition processes in L2-dominant bilingualism. *Frontiers in Language Sciences*, 3. https://doi.org/10.3389/flang.2024.1275435

Bruggeman, L., Kidd, E., Nordlinger, R., & Cutler, A. (2025). Incremental processing in a polysynthetic language (Murrinhpatha). *Cognition, 257*, Article 106075. https://doi.org/10.1016/j.cognition.2025.106075

Burness, P. A., McMullin, K., & Zamnuer, T. (2019). The role of long-distance phonological processes in spoken word recognition: A preliminary investigation. *Toronto Working Papers in Linguistics, 41*(1), 1–19. https://doi.org/10.33137/twpl.v41i1.32756

Caldas, V. G. (2025). Prosody and sentence processing in Brazilian Portuguese: A visual world paradigm study. *Revista da Abralin, 23*(2), 161–194. https://doi.org/10.25189/rabralin.v23i2.2230

Cavicchio, F., Melcher, D., & Poesio, M. (2014). The effect of linguistic and visual salience in visual world studies. *Frontiers in Psychology*, *5*, 176. https://doi.org/10.3389/fpsyg.2014.00176

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, *25*(4), 438–446. https://doi.org/10.3758/BF03201120

Chambers, C. G., Tanenhaus, M. K., & Magnusson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696. https://doi.org/10.1037/0278-7393.30.3.687

Chen, K. K., Johansen, I. M., & Chow, W.-Y. (2025). Revising noun predictions based on English measure phrases: evidence from visual-world eye-tracking. *Language, Cognition and Neuroscience, 40*(1), 87–102. https://doi.org/10.1080/23273798.2024.2399111

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Coco, M. I., & Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing*, *16*(2), 131–150. https://doi.org/10.1007/s10339-014-0642-0

Coco, M. I., Keller, F., & Malcolm, G. L. (2016). Anticipation in real-world scenes: The role of visual context and visual memory. *Cognitive Science*, *40*(8), 1995–2024. https://doi.org/10.1111/cogs.12313

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language, 122*, Article 104298. https://doi.org/10.1016/j.jml.2021.104298

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the Visual World Paradigm. *Studies in Second Language Acquisition*, *39*(4), 621–652. https://www.jstor.org/stable/26331046

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening, *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498. https://doi.org/10.1037/0278-7393.30.2.498

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459. https://doi.org/10.3758/bf03193787

Dahan, D., Magnusson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dahan, D., Magnusson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. L*anguage and Cognitive Processes*, *16*(5-6), 507–534. https://doi.org/10.1080/01690960143000074

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

Dailey, M. M., Straboni, C., & Peperkamp, S. (2023). Using allophonic variation in L2 word recognition: French listeners' processing of English vowel nasalization. *Second Language Research*. https://doi.org/10.1177/02676583231181472

Degen, J., & Pophristic, S. (2023). The role of production expectations in visual world paradigm linking hypotheses. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*. Retrieved from https://escholarship.org/uc/item/7dv5142s

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201. https://doi.org/10.1111/cogs.12227

Desmeules-Trudel, F., & Zamuner, T. S. (2023). Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research*, *39*(2), 333–362. https://doi.org/10.1177/02676583211030604

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*(2), 284–292. https://doi.org/10.1016/j.cognition.2008.11.012

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When Gender and Looking Go Hand in Hand: Grammatical Gender Processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Ellert, M. (2013). Resolving ambiguous pronouns in a second language: A visual-world eye-tracking study with Dutch learners of German. *International Review of Applied Linguistics in Language Teaching*, *51*(2), 171–197. https://doi.org/10.1515/iral-2013-0008

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573.

Farmer, T. A., Cargill, S. A., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*(4), 570–595. https://doi.org/10.1016/j.jml.2007.04.003

Fallon, J., & Pylkkänen, L. (2024). Language at a glance: How our brains grasp linguistic structure from parallel visual input. *Science Advances, 10*(42), Article eadr9951. https://doi.org/10.1126/sciadv.adr9951

Farris-Trimble, A., & McMurray, B. (2018). Morpho-phonological regularities influence the dynamics of real-time word recognition: Evidence from artificial language learning. *Laboratory Phonology, 9*(1), 1–34. https://doi.org/10.5334/labphon.41

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Ferreira, F., & Clifton, C. (1986). The Independence of Syntactic Processing. *Journal of Memory and Language*, *25*, 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. http://www.jstor.org/stable/20182754

Fotiadou, G., Muñoz, A., & Tsimpli, I. (2020). Anaphora resolution and word-order across adulthood: Ageing effects on online listening comprehension. *Glossa: A journal of general linguistics, 5*(1), 71. https://doi.org/10.5334/gjgl.997

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Freeman, M. R., & Marian, V. (2022). Visual word recognition in bilinguals: Eye-tracking evidence that L2 proficiency impacts access of L1 phonotactics. *Studies in Second Language Acquisition*, *44*(3), 759–787. https://doi.org/10.1017/S027226312100053X

Frenck-Mestre, C., Kim, S. K., Choo, H., Ghio, A., Herschensohn, J., & Koh, S. (2019). Look and listen! The online processing of Korean case by native and non-native speakers, *Language, Cognition, and Neuroscience*, *34*(3), 385–404, https://doi.org/10.1080/23273798.2018.1549332

Fuchs, Z. (2021). Facilitative use of grammatical gender in Heritage Spanish. *Linguistic Approaches to Bilingualism, 12*(6), 845–871. https://doi.org/10.1075/lab.20024.fuc

Fuchs, Z. (2023). Processing of grammatical gender agreement morphemes in Polish: evidence from the Visual World Paradigm. *Morphology*, *34*, 1–32. https://doi.org/10.1007/s11525-023-09418-3

Fuchs, Z., & Sekerina, I. A. (2025). New evidence for the role of morphological markedness of gender agreement cues in monolingual and heritage-bilingual facilitative processing. In *Essays in Honor of Maria Polinsky*. University of California. https://doi.org/10.7280/S97W697F

Ge, H., Mulders, I., Kang, X., Chen, A., & Yip, V. (2021). Processing focus in native and non-native speakers of English: An eye-tracking study in the visual world paradigm. *Applied Psycholinguistics*, *42*(4), 1057–1088. https://doi.org/10.1017/S0142716421000230

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Gómez Vidal, B. (2024). *Processing argument structure: Eye-tracking evidence from Spanish native speakers in the visual world paradigm* [Unpublished doctoral dissertation]. University of the Basque Country (UPV/EHU).

Gregg, J., Inhoff, A. W., & Li, X. (2023). Lexical competition influences correct and incorrect visual word recognition. *Quarterly Journal of Experimental Psychology*, *76*(5), 1011–1025. https://doi.org/10.1177/17470218221102878

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41–58). Academic Press.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Lang. Res.*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Hadar, B., Skrzypek, J. E., Wingfield, A., & Ben-David, B. M. (2016). Working memory load affects processing time in spoken word recognition: Evidence from eye-movements. *Frontiers in Neuroscience*, *10*, 221. https://doi.org/10.3389/fnins.2016.00221

Han, C. H., Moulton, K., Block, T., Gendron, H., & Nederveen, S. (2021). Pronouns are as sensitive to structural constraints as reflexives in early processing: Evidence from visual world paradigm eye-tracking. *Frontiers in Psychology*, *12*, 611466. https://doi.org/10.3389/fpsyg.2021.611466

Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience*, *32*(10), 1229–1249. https://doi.org/10.1080/23273798.2017.1327080

Hert, R., Järvikivi, J., & Arnhold, A. (2024). The importance of linguistic factors: He likes subject referents. *Cognitive Science, 48*(4), Article e13436. https://doi.org/10.1111/cogs.13436

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *43*(9), 1352–1374. https://doi.org/10.1037/xlm0000388

Hintz, F., Voeten, C. C., Dobó, D., Lukics, K. S., & Lukács, Á. (2024). The role of general cognitive skills in integrating visual and linguistic information during sentence comprehension: individual differences across the lifespan. *Scientific Reports, 14*, Article 17797. https://doi.org/10.1038/s41598-024-68674-3

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Lang. Res.*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H., & Lemmerth, N. (2016). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Huang, Y. T., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, *200*, 104251. https://doi.org/10.1016/j.cognition.2020.104251

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*(1), 65–80. https://doi.org/10.1016/j.actpsy.2005.06.002

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 55(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Jesse, A., Poellmann, K., & Kong, Y. Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of speech, language, and hearing research: JSLHR*, *60*(1), 190–198. https://doi.org/10.1044/2016_JSLHR-H-15-0340

Johannessen, J. B. et al. (2024). Cross-linguistic effects in grammatical gender assignment and predictive processing in L1 Greek, L1 Russian, and L1 Turkish speakers of Norwegian as a second language. *Second Language Research, 41*(1), 217–259. https://doi.org/10.1177/02676583241227709

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709–748. https://doi.org/10.1080/01690960701771220

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133−156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37−55. https://doi.org/10.1023/a:1021933015362

Kaplan, M. J., Raju, A., & Arunachalam, S. (2021). Real-time processing of event descriptions for partially- and fully-completed events: Evidence from the visual world paradigm. *Proceedings of the Linguistic Society of America, 6*(1),118–132. https://doi.org/10.3765/plsa.v6i1.4954

Karaca, F., Brouwer, S., Unsworth, S., & Huettig, F. (2024). Morphosyntactic predictive processing in adult heritage speakers: effects of cue availability and spoken and written language experience. *Language, Cognition and Neuroscience*, *39*(1), 118–135. https://doi.org/10.1080/23273798.2023.2254424

Kırcalı, C. E., Uzun, İ. P., & Aydın, Ö. (2021). Syntactic and prosodic processing of quantifier ambiguity in Turkish. *Proceedings of the Sixth Workshop on Turkic and Languages in Contact with Turkic, 6,* 5066. Linguistic Society of America (LSA). https://doi.org/10.3765/ptu.v6i1.5066

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition, 43*(1), 133–154. https://doi.org/10.1017/S0272263120000443

Kim, S. J., & Xiang, M. (2024). Encoding discourse structure information during language comprehension: Evidence from web-based visual world paradigm experiments. In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. University of California. https://escholarship.org/uc/item/3m93r04n

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Kong, Y. Y., & Jesse, A. (2017). Low-frequency fine-structure cues allow for the online use of lexical stress during spoken-word recognition in spectrally degraded speech. *The Journal of the Acoustical Society of America*, *141*(1), 373. https://doi.org/10.1121/1.4972569

Köhne-Fuetterer, J., Drenhaus, H., Delogu, F., & Demberg, V. (2021). The online processing of causal and concessive discourse connectives. *Linguistics, 59*(2), 417–448. https://doi.org/10.1515/ling-2021-0011

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 748–756. https://doi.org/10.1037/a0022637

Kumcu, A., & Thompson, R. L. (2020). Less imageable words lead to more looks to blank locations during memory retrieval. *Psychological Research*, *84*(3), 667–684. https://doi.org/10.1007/s00426-018-1084-6

Kumcu, A., & Thompson, R. L. (2022). Remembering spatial words: Sensorimotor simulation affects verbal recognition memory. *Quarterly Journal of Experimental Psychology*, *75*(9), 1694–1710. https://doi.org/10.1177/17470218211059011

Kutlu, E., Kin-Packard, J., Jeppsen, C., Tomblin, J. B., & McMurray, B. (2024). The development of real-time spoken and written word recognition derives from changes in ability, not maturation. *Cognition, 251*, Article 105899. https://doi.org/10.1016/j.cognition.2024.105899

Li, X., & Qu, Q. (2024). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review, 31*(1), 249–258. https://doi.org/10.3758/s13423-023-02348-5

Loerts, H., Wieling, M., & Schmid, M. S. (2013). Neuter is not common in Dutch: Eye movements reveal asymmetrical gender processing. *Journal of Psycholinguistic Research*, *42*, 551–570. https://doi.org/10.1007/s10936-012-9234-2

Lundquist, B., & Vangsnes, Ø. (2018). Language separation in bidialectal speakers: Evidence from eye tracking. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01394

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

Mirman, D., Magnuson, J.S., Estes, K.G., Dixon, J.A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271–80. https://doi.org/10.1016/j.cognition.2008.02.003

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(3),1–27. https://doi.org/10.1007/s41809-019-00035-3

Malarski, K., Jankowiak, K., & Dziubalska-Kołaczyk, K. (2025). Dialect use and style-shifting in the speech of Polish migrants in Norway: Acquisition of Norwegian (L3). In K. Jankowiak, M. Połczynska-Bletsos, & K. Dziubalska-Kołaczyk (Eds.), *From sub-lexical to discourse-level effects in bi- and multilingual language processing* [Editorial]. *Frontiers in Psychology, 16*, Article 1558057. https://doi.org/10.3389/fpsyg.2025.1558057

Marquis, A., Al Kaabi, M., Leung, T., & Boush, F. (2020). What the eyes hear: An eye-tracking study on phonological awareness in Emirati Arabic. *Frontiers in Psychology*, *11*, 1–10. https://doi.org/10.3389/fpsyg.2020.00452

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *10*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

McMurray, B. (2023). I'm not sure that curve means what you think it means: Toward a [more] realistic understanding of the role of eye-movement generation in the Visual World Paradigm. *Psychon Bull Rev*, *30*, 102–146. https://doi.org/10.3758/s13423-022-02143-8

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661–671. https://doi.org/10.1080/17470210601183890

Minor, S., Mitrofanova, N., Guajardo, G., Vos, M., & Ramchand, G. (2023). Aspect processing across languages: A visual world eye-tracking study. *Front. Lang. Sci.*, *1*(1052205). https://doi.org/10.3389/flang.2022.1052205

Mulders, I., & Szendrői, K. (2016). Early association of prosodic focus with alleen 'only': Evidence from eye movements in the visual-world paradigm. *Frontiers in Psychology*, *7*, 150. https://doi.org/10.3389/fpsyg.2016.00150

Müller, G., Bodnár, E., Skopeteas, S., & Kröger, J. M. (2021). On the impact of case and prosody on thematic role disambiguation: An eye-tracking study on Hungarian. *Language and Speech*, *64*(4), 930–961. https://doi.org/10.1177/0023830920974709

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, *107*(3), 850–903. https://doi.org/10.1016/j.cognition.2007.12.011

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, 71. https://doi.org/10.1136/bmj.n71

Perdomo, M., & Kaan, E. (2021). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research, 37*(2), 349–375. https://doi.org/10.1177/0267658319879196

Pozniak, C., Hemforth, B., & Scheepers, C. (2018). Cross-Domain Priming from Mathematics to Relative-Clause Attachment: A Visual-World Study in French. *Frontiers in psychology*, *9*, 2056. https://doi.org/10.3389/fpsyg.2018.02056

Prystauka, Y., Altmann, G. T. M., & Rothman, J. (2024). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods, 56*(6), 3504–3522. https://doi.org/10.3758/s13428-023-02176-4

Rehrig, G., Barker, M., Peacock, C.E. et al. (2022). Look at what I can do: Object affordances guide visual attention while speakers describe potential actions. *Atten Percept Psychophys, 84*, 1583–1610. https://doi.org/10.3758/s13414-022-02467-6

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrase. *Cognitive Science*, *30*(2), 193–241. https://doi.org/10.1207/s15516709cog0000_58

Özsoy, O., Çiçek, B., Özal, Z., Gagarina, N., & Sekerina, I. A. (2023). Turkish-German heritage speakers' predictive use of case: webcam-based vs. in-lab eye-tracking. Frontiers in Psychology, 14, Article 1155585. https://doi.org/10.3389/fpsyg.2023.1155585

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1108–1117. https://doi.org/10.1037/a0019901

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta psychologica*, *137*(2), 172–180. https://doi.org/10.1016/j.actpsy.2010.09.010

Salverda, A. P., & Tanenhaus, M. K. (2017). The Visual World Paradigm. In Annette M. B. de Groot and Peter Hagoort (Eds). *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 89–110). Wiley-Blackwell.

Sarrett, M. E., & McMurray, B. (2025). Timecourse of bottom-up and top-down language processing during a picture-based semantic priming task. *Language, Cognition and Neuroscience, 40*(1), 122–144. https://doi.org/10.1080/23273798.2024.2409136

Scheepers, C. (2003). Syntactic priming of relative clause attachment: Persistence of structural configuration in sentence production. *Cognition*, *89,* 179-205. https://doi.org/10.1016/S0010-0277(03)00119-7

Schreiber, K. E., & McMurray, B. (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics*, *81*(4), 1147–1166. https://doi.org/10.3758/s13414-019-01712-9

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Sekerina, I. A., Campanelli, L., & van Dyke, J. A. (2016). Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology*, *14*(7), 873. https://doi.org/10.3389/fpsyg.2016.00873

Shen, W., Qu, Q., & Li, X. (2016). Semantic information mediates visual attention during spoken word recognition in Chinese: Evidence from the printed-word version of the visual-world paradigm. *Attention, Perception & Psychophysics*, *78*(5), 1267–1284. https://doi.org/10.3758/s13414-016-1081-z

Shkuropatskaya, M. G., Vlasov, M. Zhukova, T. V., & Isaeva, I. P. (2020). *Semantic and associative factors in lexical processing of polysemous bionyms in Russian: Evidence from an eye-tracking study in the visual world paradigm*. PrePrint. https://www.researchgate.net/publication/342702028

Soto, C., & Schmid, M. S. (2024). Carrot or parrot? An eye-tracking study on spoken word recognition in a language attrition context. *The Language Learning Journal, 52*(2), 199–217. https://doi.org/10.1080/09571736.2023.2294058

Soroli, E. (2024). How language influences spatial thinking, categorization of motion events, and gaze behavior: A cross-linguistic comparison. *Language and Cognition, 16*(4), 924–968. https://doi.org/10.1017/langcog.2023.66

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

Spivey, M. J., Tyler, M. J., Richardson, D. C., & Young, E. E. (2000). Eye movements during comprehension of spoken scene descriptions. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, 487–492.

Stern, M.C., Madsen, C.N., Stover, L.M. et al. (2019). Language history attenuates syntactic prediction in L1 processing. *Journal of Cultural Cognitive Science, 3*, 235–255. https://doi.org/10.1007/s41809-019-00048-y

Sturt, P. (2003). The time course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.777786

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Lawrence Erlbaum Associates, Inc.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/s0010-0277(99)00032-3

Wei, Y., Mak, W. M., Evers-Vermeul, J., & Sanders, T. J. M. (2019). Causal connectives as indicators of source information: Evidence from the visual world paradigm. *Acta Psychologica*, *198*, 102866. https://doi.org/10.1016/j.actpsy.2019.102866

Xu, X., Ji, C., Li, T., & Pickering, M. J. (2025). The prediction of segmental and tonal information in Mandarin Chinese: An eye-tracking investigation. *Language, Cognition and Neuroscience, 40*(1), 56–70. https://doi.org/10.1080/23273798.2024.2395549

Zahner, K., Kutscheid, S., & Braun, B. (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics*, *74*, 75–95. https://doi.org/10.1016/j.wocn.2019.02.004

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67. https://doi.org/10.1016/j.jml.2015.10.003

Zhang, Z., & Zhang, Q. (2025). Linear incrementality in focus and accentuation processing during sentence production: Evidence from eye movements. *Frontiers in Human Neuroscience, 18*, Article 1523629. https://doi.org/10.3389/fnhum.2024.1523629

Zhao, Z., Ding, J., Wang, J., Chen, Y., & Li, X. (2024). The flexibility and representational nature of phonological prediction in listening comprehension: Evidence from the visual world paradigm. *Language and Cognition, 16*(2), 481–504. https://doi.org/10.1017/langcog.2023.38

Zhu, Y., & Grüter, T. (2025). Native speakers and learners of Mandarin predict upcoming arguments in dative constructions based on categorical and gradient verb constraints. *Bilingualism: Language and Cognition, 28*(4), 601–612. https://doi.org/10.1017/S1366728924000749

Yao, P., Hall, D., Borer, H., & Stockall, L. (2024). Dutch–Mandarin learners' online use of syntactic cues to anticipate mass vs. count interpretations. *Second Language Research, 40*(4), 803–831. https://doi.org/10.1177/02676583231175106

Appendix

Table A1. An overview of the studies included in the present review

Research	Task	Language
Phonological and morphological processing		
Allopenna et al., 1998	TRACE prediction and referent gating	L1 English
Dahan et al., 2001a	word recognition	
McMurray et al., 2002, 2009	lexical access	
Cutler et al., 2006	phonetic mapping	L1 Japanese, L2 English
McQueen & Viebach, 2007	word judgment	L1 Dutch
Salverda & Tanenhaus, 2010	phonological-orthographic processing	L1 English
Farris-Trimble & McMurray, 2018	morphophonological alternations	
Burness et al., 2019	long-distance phonological learning	L1 Arabic
Schreiber & McMurray, 2019	phonemic recognition	L1 English
Marquis et al., 2020	phonological awareness	L1 English
Apfelbaum et al., 2021	phonological competition	L1-L2 English, L1 Spanish
Freeman & Marian, 2022	phonemic recognition	L1 English, L2 French
Dailey et al., 2023	phonemic gating	L1 Dutch, L2 English
Desmeules-Trudel & Zamuner, 2023	word recognition	Native & non-native English
Bruggeman & Cutler, 2024		L1 English, L2 Afrikaans
Bao et al., 2024	novel word learning	L1 English
Berghoff & Bylund, 2024	cross-language activation	L1 Spanish, L1 English
Kutlu et al., 2024	lexical competition	L1 Mandarin
Soto & Schmid, 2024		L1 Italian, L2 English
Zhao et al., 2024	homophone and tonal competitor	L1 Mandarin
Bramlett & Wiener, 2025	lexical stress	
Xu et al., 2025	phonological prediction	
Semantic processing		
Sedivy et al., 1999	contextual interpretation	L1 English
Dahan & Tanenhaus, 2005	spoken-word recognition and matching	L1 Dutch
Novick et al., 2008	visual context and lexical ambiguity	L1 English
Duñabeitia et al., 2009	abstract vs. concrete words	L1 Spanish
Arias-Trejo & Plunkett, 2012	word referent identification	L1 English
Degen & Tanenhaus, 2016	listening judgment	
Hadar et al., 2016	spoken word recognition and memory	L1 Hebrew
Shen et al., 2016	orthographic similarity	L1 Chinese
Zamuner et al., 2016	nonwords/pronunciation	L1 English
Shkuropatskaya et al., 2020	polysemous bionyms judgment	L1 Russian

Kumcu & Thompson, 2020, 2022	spatial memory & retrieval	L1 English
Kaplan et al., 2021	event processing	
Rehrig et al., 2022	meaning-grasp mapping	
Kim & Grüter, 2021	pronoun resolution	L1-L2 English, L1 Korean
Corps et al., 2022	gender-stereotype prediction	
Ahn & Song, 2023	definiteness & prediction	L1 English
Hintz et al., 2024	semantic prediction	L1 Hungarian
Li & Qu, 2024	semantic-phonological prediction	L1 Mandarin
Soroli, 2024	motion event categorisation	L1 English, L1 French
Prystauka et al., 2024	semantic verb constraint and lexical interference	L1 Russian
Chen et al., 2025	noun prediction revision	L1 English
Sarrett & McMurray, 2025	semantic priming	
Syntactic processing		
Tanenhaus et al., 1995	ambiguity resolution & judgment	L1 English
Altmann & Kamide, 1999		
Trueswell et al., 1999	visual-linguistic saliency	
Dahan et al., 2001b	TRACE prediction	
Arnold et al., 2000, 2004	pronoun resolution & reference	
Spivey et al., 2002	ambiguous PP and prosodic judgment	
Sturt, 2003	gender resolution	
Chambers et al., 2004	ambiguous PP and instrument task	
Altmann, 2004	tense interpretation and visual context	
Dahan & Tanenhaus, 2004	case marking	L1 Dutch
Brown-Schmidt et al., 2005	pronoun resolution	L1 English
Runner et al., 2006	Binding Theory	
Knoeferle & Crocker, 2006	scene comprehension	
Altmann & Kamide, 2007	semantic/phonological blocking	
Belke & Meyer, 2007	sentence description and judgment	
Farmer et al., 2007	garden-path sentences	
Brown-Schmidt & Konopka, 2008	referential context & communication	L1 Spanish, L2 English
Kaiser & Trueswell, 2008	pronoun/reflexive interpretation	L1 Finnish
Grüter et al., 2012	gender recognition tasks	L1 English, L1 Italian, & L2 Spanish
Dussias et al., 2013	picture naming and plausibility	
Loerts et al., 2013	colour/gender matching	L1 Dutch
Ellert, 2013	pronoun resolution & referent assignment	L1 German, L2 Dutch
Hopp, 2013	gender production/comprehension	L1 English, L2 German
Hopp & Lemmerth, 2016		L1 Russian, L2 German
Arslan et al., 2015	evidentiality interpretation	L1, early-late L1 Turkish, L2 German
Coco & Keller, 2015	RCs ambiguity	L1 English

Cunnings et al., 2017	overt pronoun resolution	L1 English, L2 Greek
Hintz et al., 2017	verb-mediated prediction	L1 Dutch
Pozniak et al., 2018	RCs attachment & prediction	L1 French
Stern et al., 2019		L1 Spanish, L2 English
Lundquist & Vangsnes, 2018	dialectal gender processing	Norwegian, Heritage
Fuchs, 2021, 2023	grammatical gender agreement	Spanish, Polish
Frenck-Mestre et al., 2019	thematic roles & structural ambiguity	L1 French, L1 Kazakh, L2 Korean
Fotiadou et al., 2020	anaphora resolution and memory	L1 Greek
Huang & Snedeker, 2020	thematic constraints	
Han et al., 2021	pronouns/demonstratives interpretation	L1 English
Degen & Pophristic, 2023	incremental reference production	
Özsoy et al., 2023	agent-patient assignment	L1 Turkish, Heritage Turkish-German
Karaca et al., 2024	case marking	Heritage Turkish & L1 Turkish
Minor et al., 2024	grammatical aspect	L1 Russian, L1 Spanish, L1 English
Gómez Vidal, 2024	thematic role and argument structure	L1 Spanish
Johannessen et al., 2024	gender assignment and prediction	L1 Greek, L1 Russian, L1 Turkish, L2 Norwegian
Hert et., 2024	pronoun resolution	L1 German
Kim & Xiang, 2024		L1 English
Fallon & Pylkkänen, 2024	recognition and grammatical judgment	
Yao et al., 2024	anticipation of mass/count syntactic cues	L1 Dutch, L2 Mandarin
Abdel-Latif et al., 2025	speech-on-speech masking	L1 German
Alzahrani, 2025	prediction of grammatical number	L1 & L2 Arabic
Bruggeman et al., 2025	polysynthetic verb comprehension	L1 Murrinhpatha
Fuchs & Sekerina, 2025	gender agreement	L1 Russian, Heritage Russian-English
Zhu & Grüter, 2025	verb constraints and argument prediction	L1, L2 & Heritage Mandarin
Prosodic processing		
Dahan et al., 2002	accent placement and referent	L1 English

Reinsch & Weber, 2012	lexical stress	L1 Dutch
Brown et al., 2015	segmental lengthening	L1 English
Mulders & Szendrői, 2016	prosodic focus	L1 Dutch
Jesse et al., 2017 Kong & Jesse, 2017	prosodic prominence	L1 English
Henry et al., 2017 Zahner et al., 2019	thematic prediction pitch alignment and stress perception	L1 German
Huettig & Guerra, 2019		L1 Dutch
Fernandez et al., 2020	speech rate and anticipatory gaze	L1 English, L1 German
Ge et al., 2021	prosodic focus	L1 Cantonese, L1 Dutch, L2 English
Kırcalı et al., 2021	quantifier ambiguity	L1 Turkish
Müller et al., 2021	prosody and case ambiguity	L1 Hungarian
Perdomo & Kaan, 2021	contrastive pitch accent	L1 English, L1 Mandarin, L2 English
Caldas, 2025	prosodic phrasing in garden-path	L1 Brazilian Portuguese
Zhang & Zhang, 2025	focus and accentuation	L1 Mandarin
Pragmatic processing		
Engelhardt et al., 2006	Gricean Maxim interpretation	L1 English
Huettig et al., 2006	contextual appropriateness	
Barr, 2008	semantic-pragmatic constraints	
Brown-Schmidt et al., 2008	referential domains and conversation	
Barr & Seyfeddinipur, 2010	disfluency interpretation	
Altmann & Kamide, 2009	discourse mapping	
Brown-Schmidt, 2009	referential interpretation	
Coco et al., 2016	contextual guidance and memory	
Wei et al., 2019	causal connectives	L1 Dutch, L1 Chinese
Köhne-Fuetterer et al., 2021	semantic/pragmatic anticipation	L1 German, L1 English
Malarski et al., 2025	dialect use and style-shifting	L1 Polish, L3 Norwegian