Revealing the Temporal Dynamics of Spoken Language Processing Using the Visual World Paradigm

Authors: Wei Yipu, Wei Yipu

Date: 2023-07-12T00:00:00+00:00

Abstract

The visual world paradigm is an eye-tracking experimental paradigm that investigates real-time spoken language processing by tracking and measuring eye fixation trajectories on visual objects. The theoretical foundation for applying this paradigm to language comprehension research is the eye-mind linking hypothesis (e.g., cooperative interaction theory, goal-based linking hypothesis theory, etc.), which establishes meaningful associations between eye movement patterns and spoken language processing. Data obtained using the visual world paradigm can provide precise temporal information about spoken language processing, with commonly employed data analysis methods including analysis of mean fixation proportion within time regions of interest, divergence point analysis, growth curve analysis, etc. This paradigm provides crucial evidence for studying issues such as lexical speech recognition, syntactic ambiguity resolution, semantic comprehension, and discourse-pragmatic information processing.

Full Text

Visual World Paradigm Reveals the Time Course of Spoken Language Processing

Yipu Wei
School of Chinese as a Second Language, Peking University

Abstract

The visual world paradigm (VWP) is an eye-tracking experimental method that investigates real-time spoken language processing by tracking and measuring eye movements toward visual objects. The theoretical foundation of this paradigm

in language comprehension research rests on linking hypotheses (e.g., the coordinated interplay account, goal-based linking hypothesis) that establish meaningful connections between eye movement patterns and speech processing dynamics. Data obtained through the VWP provide precise temporal information about spoken language processing, with commonly employed analytical methods including analysis of mean fixation proportions within time windows, divergence point analysis, and growth-curve analysis. This paradigm has yielded critical evidence for issues concerning lexical and phonological recognition, syntactic ambiguity resolution, semantic comprehension, and discourse-pragmatic information processing.

Keywords: visual world paradigm; eye-tracking; spoken language processing

Introduction

The time course of language processing has long been a central issue in psycholinguistics. Investigating this question holds significance at three levels. First, understanding when different types of linguistic information (phonological, semantic, syntactic, discourse, pragmatic) and information from various sources (linguistic input, visual context, world knowledge) are processed is crucial for constructing models of language comprehension. For instance, McRae et al.'s (1998) constraint-based model of language processing was proposed based on evidence regarding the time course of ambiguous sentence comprehension. Second, examining how factors influencing language comprehension (e.g., word frequency, language proficiency, cognitive abilities) exert their effects requires temporal information about language processing. Magnuson et al. (2003), for example, demonstrated that word frequency affects lexical recognition by examining how quickly listeners identify referents from speech input. Third, the timeline of linguistic element processing can serve as an important indicator of language comprehension ability, with applications in child language acquisition, second language processing, and assessment of older adults' linguistic capacities (Saryazdi & Chambers, 2021). As a vital tool for investigating the time course of spoken language processing, the visual world paradigm provides precise temporal information that illuminates processing at various linguistic levels.

The visual world paradigm is an experimental method that studies real-time spoken language comprehension by tracking and measuring eye movements in visual contexts (Allopenna et al., 1998; Salverda & Tanenhaus, 2018). With the integration of eye-tracking equipment and computer interfaces in the late 1960s, real-time recording of eye movements and automated data processing became feasible. By the mid-1970s, substantial progress had been made in using eye-tracking technology to study written text reading (see review: Rayner, 1978). Concurrently, Cooper (1974) first attempted to measure spoken language comprehension using eye-tracking technology, establishing an initial link between listeners' fixations on visual objects and language processing. However, it was not until Tanenhaus et al. (1995) published their seminal paper in *Science* demonstrating how eye-tracking could reveal the processing of ambiguous

sentences that the visual world paradigm (named by Allopenna et al., 1998) became widely adopted in spoken language research, emerging as one of the most important methodological tools in psycholinguistics and cognitive psychology (Qiu et al., 2009; Lin & Wang, 2018).

This paper elucidates how the eye-tracking visual world paradigm can be used to investigate the time course of spoken language processing. To address this issue, the paper first introduces the linking hypotheses that connect eye movements to language comprehension processes, thoroughly explaining the temporal characteristics of the paradigm's tasks and data, and how these features can be leveraged for data analysis. Subsequently, focusing on the time course of spoken language processing, the paper reviews empirical findings from the past two decades using this paradigm in phonological, semantic, syntactic, discourse, and pragmatic processing, further illustrating the contributions of this highly time-sensitive method to research on the temporal dynamics of spoken language processing.

1. Linking Hypotheses Between Eye Movements and Language Processing

The theoretical foundation of the visual world paradigm rests on linking hypotheses that connect eye movement patterns to the cognitive processes underlying spoken language comprehension (Allopenna et al., 1998; Tanenhaus et al., 2000). Specifically, as listeners process spoken information, they construct dynamic mental representations of the described situations; their attention to specific entities within these representations shifts with incoming linguistic information, and correspondingly, their fixation locations in visual space change (Altmann & Kamide, 2007). These fixations and their movements, accompanied by changes in pupil position, can be effectively measured through eye-tracking, thereby revealing the time course of spoken language processing. Over the past two decades, numerous specific linking hypotheses have been proposed to explain how visual attention is allocated to referent objects (see review: Magnuson, 2019). This paper summarizes three influential linking hypothesis theories to further clarify the theoretical basis for applying the visual world paradigm to spoken language processing. Although these hypotheses do not directly define the time course of specific linguistic element processing, they encompass several stages of spoken language processing that serve as the prerequisite foundation for investigating processing timelines.

Knoeferle and Crocker (2006, 2007) proposed the **coordinated interplay account**, which divides visual-world-based spoken language comprehension into three main stages: (1) integrating newly input words into the existing sentence structure to form new sentence interpretations, and using this new information together with prior linguistic information and relevant world knowledge to generate predictions about upcoming linguistic material; (2) searching working memory (which includes the previously viewed visual scene) for objects referred to by words or objects predictable from stage one information; and (3) map-

ping linguistic input (nouns, verbs, etc.) onto objects and actions in the visual scene, using visual scene information to revise previously formed sentence interpretations and generate new predictions (Knoeferle & Crocker, 2006, 2007; Pyykkönen-Klauck & Crocker, 2016). Notably, although these three processes are presented sequentially in the coordinated interplay account, the theory does not preclude the possibility that they may overlap or occur simultaneously in time. This account highlights the importance of visual scene information for spoken language comprehension; moreover, although these scene representations gradually fade from working memory after the visual scene disappears, memory for the scene continues to exert significant influence on subsequent sentence processing (Knoeferle & Crocker, 2007).

Altmann and Mirković (2009) proposed an alternative linking hypothesis that similarly acknowledges the joint influence of linguistic information (e.g., real-time language input, contextual information) and non-linguistic information (e.g., visual scenes, world knowledge) on sentence processing. However, unlike Knoeferle and Crocker's (2006, 2007) coordinated interplay account, Altmann and Mirković (2009) argue that the processes of visual scene interpretation and linguistic comprehension are inseparable both in mental representation and processing time—because linguistic and non-linguistic information are stored in the same system, jointly constituting a dynamic representation of the situation. When listeners receive a piece of information, representations of objects (including experiences and knowledge associated with those objects) become activated. As listeners continuously receive information from different sources (linguistic input, visual scenes, world knowledge), these object representations continuously change. When information from different sources overlaps, activation of the object representation strengthens. Different states of this representational system manifest at the mental representation level as attention allocation, which in turn influences eye movement patterns. In other words, the temporal trajectory of fixations on visual objects accompanying sentence input is influenced and driven by a common representational system that includes linguistic information, contextual information, visual scenes, and world knowledge.

Both of the above linking hypotheses adopt a language comprehension perspective, viewing changes in eye fixation during spoken language processing as the result of joint action between linguistic input and visual information. These hypotheses treat language processing as an independent task, unrelated to the behavioral task goals in the experimental procedure. However, such language comprehension-based linking hypotheses do not address how the actions required to complete the task itself affect referential processing (Chambers et al., 2004), nor do they consider that in visual search, eye movements are inherently linked to behavioral task goals—that is, participants fixate more on objects relevant to their task objectives. To better explain the relationship between language processing and eye movements, Salverda et al. (2011) proposed the **goal-based linking hypothesis**, incorporating the new dimension of "task goals" into linking hypotheses.

Unlike language comprehension-based linking hypotheses, the goal-based linking hypothesis posits that not only context and linguistic input can constrain language processing, but task goals themselves can also serve as constraints—visual objects directly relevant to task execution attract more eye fixations, whereas objects irrelevant to goal execution do not. This hypothesis suggests that spoken language processing in visual contexts first involves a fundamental task of mapping linguistic input onto selectable objects in the visual scene, with eye movements serving this task goal by locking onto potential referent objects; objects that do not afford the required action receive few fixations. For example, when hearing the instruction "put the cube into the can," only cans with sizes that can accommodate the cube become target containers that attract fixations (Chambers et al., 2004). Salverda et al. (2011) argue that additional tasks such as clicking or moving objects jointly constitute the task goal structure in spoken language processing tasks and influence eye movements. For instance, when participants listen to sentences with a truth-value judgment task, they show earlier and more significant predictive fixations compared to listening without such a task (Altmann & Kamide, 1999), locking onto referent targets more quickly in the time course. The goal-based linking hypothesis introduces new requirements for refining and hierarchically structuring task goals during language processing.

Spoken language processing studies using the eye-tracking visual world paradigm are based on these linking hypotheses and can be divided into two main research directions according to how visual information is utilized. The first type of study uses the visual scene as a backdrop for presenting objects, where attention to specific referents in the mental representation is projected onto the visual scene, and listeners form fixations on referent objects accordingly; the eye movement trajectory formed by fixating on backdrop objects reveals how different linguistic components are processed in real time (e.g., Cooper, 1974; Cozijn et al., 2011; Kaiser, 2016). The second type treats visual information as a contextual constraint, primarily exploring how information in the visual environment (e.g., number of candidate objects, object size comparisons, depicted event actions) itself influences language processing (e.g., Chambers et al., 2002; Knoeferle et al., 2005; Tanenhaus et al., 1995). These two types of studies employ similar tasks, but at the theoretical level of linking hypotheses, the first emphasizes the simultaneity and inseparability of visual scene interpretation and spoken input comprehension, while the second treats visual scene processing as a relatively independent process, highlighting the role of the visual scene itself in spoken language processing. The latest trend in eye movement research has begun to focus on the potential role of task goals in language processing. Although linking hypotheses incorporating the goal dimension have completed initial theoretical construction, research comparing processing effects under different task goals remains a gap in the literature.

2.1 Paradigm and Tasks

A typical visual world paradigm experiment involves spoken language instructions presented auditorily and objects displayed as visual stimuli (in the real world or on a computer screen). While comprehending the spoken instructions, participants' fixation locations on visual objects are recorded in real time by an eye tracker for subsequent analysis (see Figure 1 [Figure 1: see original paper]). Visual stimulus images generally appear before the language instructions with a certain preview duration; language instructions are presented at a relatively fixed rate. Previous research has found that factors such as image complexity, preview duration, speech rate, and task instruction type (whether participants are explicitly told to predict target objects) can all influence eye movement results (Huettig & Guerra, 2019; Ferreira et al., 2013).

The visual world paradigm includes two main types of experimental tasks: **active tasks** (action-based tasks), which require participants to respond behaviorally to language instructions (e.g., grasping, moving, or clicking objects; see Hanna & Tanenhaus, 2004; Tanenhaus et al., 1995), and **passive tasks** (listen-and-look tasks), where participants simply listen to language instructions and view pictures or scenes without making behavioral responses (Altmann & Kamide, 1999; Knoeferle et al., 2005). Regarding differences between these tasks, Salverda et al. (2011) noted that in active-task visual world paradigm experiments, participants direct substantial fixations toward target objects before grasping, moving, or clicking them; passive-task experiments do not exhibit this fixation pattern—a factor that may lead to differences in eye movement patterns between the two task types. Pyykkönen-Klauck and Crocker (2016) reviewed and compared eye movement results from experiments using both task types, concluding that some linguistic effects (e.g., word frequency effects) are more sensitive in active tasks, with participants locking onto target objects faster, demonstrating more rapid real-time language comprehension. Passive listen-and-look visual world paradigm experiments, which do not require participants to perform additional tasks, have relatively better ecological validity (Huettig et al., 2011a) and can be used to examine which spoken language processing effects are universal in language-visual interactions and which exist only under specific experimental task conditions (Huettig et al., 2011b).

The visual world paradigm has two main variants: the **printed-word paradigm** (Huettig & McQueen, 2007) and the **blank screen paradigm** (Altmann, 2004). In the printed-word paradigm, visual stimulus images are replaced with words appearing on the screen. Participants hear speech input related to these words while their fixation trajectories on each letter are recorded for analysis. This variant can be used to examine phonological recognition processes and how orthographic information is processed in real time. The blank screen paradigm is primarily used to reveal the role of short-term memory in real-time language processing. After the visual stimulus image is presented for several seconds, a blank screen appears (typically for 1 second) before the spoken instruction is played. Experiments using this paradigm

demonstrate that even after objects from the visual stimulus image disappear, participants still look toward the locations where relevant objects previously appeared when hearing language instructions (Knoeferle & Crocker, 2007). The blank screen paradigm provides evidence for mental representations: once formed, these representations can be temporarily stored in short-term memory and participate in subsequent language processing without relying on visual stimuli.

2.2 Data and Variables

Common dependent variables in visual world paradigm data analysis are fixations and saccades. The most frequently used fixation measure is **fixation proportion**, defined as the proportion of fixations falling within a specific region of interest across all trials during a designated time window. Common saccade measures include saccade proportion (the proportion of saccades directed to the target region of interest across all trials) and saccadic reaction time (the time required to make a saccade to the target region of interest after target word onset). Independent variables in the data can be within-subject factors (e.g., experimental vs. control conditions, ambiguous vs. unambiguous sentences) or between-subject factors (e.g., different language background groups, age groups).

The advantage of the visual world paradigm lies in the high temporal precision of its data. Current research-grade eye trackers can achieve a sampling rate of 1000 Hz, capturing eye position every millisecond and providing accurate time course information. Taking fixation proportion as an example, researchers can not only compare mean fixation proportions across different conditions within a time window to identify specific effects in spoken language processing, but more importantly, they can investigate when effects emerge (i.e., when fixation proportions begin to differ significantly between conditions) and how effects change over time.

2.3 Utilizing Temporal Information for Data Analysis

Temporal precision is the most important characteristic of visual world paradigm data, and leveraging temporal information effectively is key to data analysis. Based on how temporal information is utilized, existing analytical methods can be categorized into three types: (1) comparison of mean fixation proportions within designated time windows; (2) analysis of effect onset and duration time courses; and (3) analysis of curve patterns showing how effects change over time. To better illustrate the application scenarios and analytical logic of these three methods, this paper uses experimental object schematics (Figure 2 [Figure 2: see original paper]) and fixation proportion data plots (Figure 3 [Figure 3: see original paper]) from Allopenna et al. (1998) as examples (detailed discussion of this study appears in Section 3.1).

The first method is the most commonly used and intuitive approach for analyz-

ing visual world paradigm data—comparing mean fixation proportions within designated time windows, such as comparing listeners' fixation proportions on several objects in Figure 2 during the approximately 375 ms from target word "beaker" onset to offset. This analytical approach uses fixation proportion, duration, or saccade measures as dependent variables and within-subject and between-subject factors as independent variables, employing statistical methods such as t-tests, ANOVA, and linear mixed-effects models to compare differences in fixation proportions between objects or between conditions (application examples: Gardner et al., 2021; Grüter et al., 2020). Compared to t-tests and ANOVA, linear mixed-effects models are currently the most widely used analytical method, as they can incorporate variance between subjects and between trials as random variables, enabling more accurate modeling and testing of effects. It should be noted that such statistical methods typically require normally distributed data, whereas fixation proportions are bounded between 0 and 1 and generally require prior log or logit transformation (Ito & Knoeferle, 2022). Analyzing mean fixation proportions within designated time windows is the simplest visual world paradigm data analysis method, applicable to most experimental designs. Its main disadvantage is that artificially set time windows reduce the temporal precision of the data and cannot capture trends in how fixation proportions change over time; a compensatory approach is to include different time windows as independent variables in the analysis model to test whether the time window variable itself significantly affects fixation proportions.

The second method involves analyzing the time course of effect onset and duration. This approach fully utilizes the precise temporal information of the visual world paradigm to investigate the exact time when a spoken language processing effect emerges. **Divergence point analysis** subdivides the potential effect time period into small windows (e.g., 20 ms), comparing fixation proportions between two conditions within each small window to identify the earliest time point at which the two fixation proportion curves begin to diverge significantly. For example, in Figure 3, the divergence point between the fixation proportion curve for the target referent "beaker" and the phonological cohort competitor "beetle" occurs around 400 ms, later than the divergence point between the target referent and the rhyme competitor "speaker." Divergence point analysis can statistically calculate the specific time points at which different curves begin to diverge significantly.

Simple divergence point analysis can only identify the onset time point of an effect (the divergence point of curves under two conditions) but cannot test the temporal interval of the divergence point or statistically compare whether two divergence points differ significantly across conditions. Advanced divergence point analysis based on bootstrapping can provide confidence intervals for each divergence time point, enabling cross-condition comparisons (Stone et al., 2021; application example: Corps et al., 2021). This advanced method provides an effective analytical tool for comparing the time courses of real-time language processing across different populations. For example, first-language (L1) and second-language (L2) speakers may not differ in effect magnitude for a partic-

ular language processing effect (e.g., predictive processing), but may differ in when the effect begins (Kaan & Grüter, 2021). This analytical approach can effectively test whether L2 speakers' predictive processing begins significantly later than that of L1 speakers. In addition to divergence point analysis, cluster-based permutation analysis (Barr et al., 2014) and bootstrapped differences of timeseries (Seedorff et al., 2018) can also be used to identify when data from two conditions begin to differ significantly (see comprehensive review of eye-tracking data analysis methods: Ito & Knoeferle, 2022). However, these methods cannot analyze how effects change over time across conditions, requiring the third type of method for curve analysis.

The third method primarily analyzes curve patterns of effects over time in visual world paradigm data. **Growth-curve analysis** models and analyzes how fixation proportion curves in key regions of interest change over time across different conditions, testing whether curve patterns differ between conditions and thereby verifying whether effects change over time (Mirman, 2014; Mirman et al., 2008). Unlike the first analytical approach, growth-curve analysis includes not only linear models with time as a variable but can also incorporate quadratic and cubic terms of time variables to model curvilinear patterns of fixation proportions over time. For example, in Figure 3, the fixation proportion for the phonological cohort competitor "beetle" shows a parabolic pattern of first increasing then decreasing, with a different slope than the rhyme competitor "speaker"—a pattern that can be analyzed using growth-curve models including quadratic time variables. In spoken language processing, fixation patterns over time are often not simply linear increases or decreases, and modeling and comparing these curves enables more precise analysis of the temporal development of language comprehension (application examples: Henry et al., 2022; Koring et al., 2012). However, growth-curve analysis has a notable limitation: time-series data exhibit autocorrelation, where adjacent time windows show high correlation in fixation positions, increasing the likelihood of Type I statistical errors (false positives) (Huang & Snedeker, 2020). Therefore, it often needs to be combined with the first and second types of methods to jointly verify effects. Generalized additive mixed models can also be used to model nonlinear data curves, using thin plate regression splines to more flexibly model changing curves while reducing statistical autocorrelation, partially compensating for the limitations of growth-curve analysis (Porretta et al., 2018).

3. Visual World Paradigm and the Time Course of Spoken Language Processing

Early debates about the time course of language processing primarily focused on the immediacy of processing. Early experiments mainly employed lexical recognition, cued recall, and self-paced reading tasks, yielding evidence that tended to support delayed-integration interpretation (e.g., Garnham et al., 1996; Stewart et al., 2000), suggesting that language users process language by delaying integration until sentence end (delayed-integration interpretation; Millis & Just,

1994). However, with the adoption of methods such as eye-tracking and event-related potentials (ERP) that enable precise measurement of reading times and brain signals, increasing evidence supports incremental interpretation, where language users process information immediately as it is encountered (incremental interpretation; Traxler et al., 1997; Cozijn et al., 2011; Koornneef & Van Berkum, 2006). For eye-tracking measures in visual contexts, although approximately 200 ms is required from auditory signal reception to eye movement response (Matin et al., 1993; Saslow, 1967), numerous spoken language experiments using the visual world paradigm have found effects of eye fixations on target objects after test word onset but before the next word begins, demonstrating that processing of information in spoken language occurs immediately (see Sections 3.1–3.5).

Building on the widespread acceptance of incremental processing, recent discussions of language processing time courses have focused on when language users utilize contextual information to understand language. Language users might combine semantic information of test words with prior context immediately as test words appear, or they might engage in predictive processing of test words' phonological, semantic, and even syntactic structures during context processing, before test word onset (expectation-based account; Levy, 2008). The visual world paradigm offers clear advantages over reading paradigms and ERP measures in detecting predictive effects (Huettig & Guerra, 2019). While most studies using the latter methods can only capture effects at test word positions arising from consistency between test word semantics and contextual information, the visual world paradigm can examine how context influences participants' fixation patterns in visual scenes even before keyword onset, providing crucial evidence for predictive processing in spoken language. The following sections analyze how the visual world paradigm addresses time course questions at different linguistic levels—phonological, semantic, syntactic, discourse, and pragmatic. It should be noted that information at different levels is not independent in spoken language processing but rather mutually influential (see review: Kuperberg & Jaeger, 2016); this paper separates these levels for organizational clarity.

3.1 Lexical Recognition and Phonological Prediction

In the visual world paradigm, listeners hearing a word search for its referent within the visual domain. This characteristic enables the paradigm to test lexical recognition processes and investigate how listeners use available information to predict phonological forms. Allopenna et al. (1998) used this paradigm to test whether the matching process between phonological input and lexical representations in spoken word recognition occurs incrementally over time. If this matching process is temporally incremental, it would predict that the phonological cohort competitor "beetle" for the target referent "beaker" would produce stronger interference than the rhyme competitor "speaker" (see Figure 2), because "beetle" overlaps phonologically with "beaker" at word onset, while

"speaker" overlaps later in the word. Allopenna et al.'s visual world paradigm eye-tracking results confirmed this hypothesis: fixation proportions on both the target object "beaker" and the competitor "beetle" increased during early stages of phonological processing (see Figure 3), whereas fixation proportion on the object "speaker" increased only at later stages of word processing and with relatively smaller magnitude. The eye-tracking fixation proportion data from the visual world paradigm effectively revealed the matching process between phonological input and lexical representations in word recognition.

Regarding whether language users can predict upcoming words' phonological information through contextual information, existing ERP studies have yielded conflicting results and have not obtained stable, replicable phonological prediction effects (DeLong et al., 2005; Nieuwland et al., 2018). The visual world paradigm provides strong evidence for investigating phonological prediction. Ito et al. (2018) employed a visual world paradigm eye-tracking experiment and found that under highly predictable contexts (e.g., "The tourists expected rain when the sun went behind the…"), listeners not only fixated predictively on the target object ("cloud") but also fixated more on the target's phonological competitor ("clown," which shares the onset syllable with "cloud"). This finding confirms the existence of phonological form prediction. More importantly, this predictive effect emerged 500 ms before target word onset in the visual world paradigm, providing strong evidence that phonological form prediction in language processing is proactive. Compared to results from other paradigms that only find integration effects at target word positions, the visual world paradigm offers more direct evidence for language prediction. Furthermore, the visual world paradigm provides empirical evidence for the mechanism of phonological prediction: like semantic prediction, phonological prediction is based on association—by processing context, language users activate corresponding semantic and phonological forms in their mental lexicon, thereby forming expectations about upcoming words (Kukona, 2020; comparison of phonological and semantic prediction: Karimi et al., 2019). Notably, phonological prediction research using Western languages faces an unavoidable problem: target words (e.g., "cloud") and their phonological competitors (e.g., "clown") overlap not only phonologically but also orthographically. Li et al. (2022) used Mandarin Chinese, where phonological and orthographic information are relatively separate, and similarly found phonological form prediction in a visual world paradigm experiment, validating the universality of phonological prediction.

3.2 Syntactic Processing and Ambiguity Resolution

The visual world paradigm contributes to research on the time course of syntactic processing in two main ways. First, the paradigm can be used to analyze ambiguity resolution processes in ambiguous sentences, such as garden-path sentences. Tanenhaus et al. (1995) first used the visual world paradigm to investigate processing of structurally ambiguous English sentences and the influence of visual scenes on sentence disambiguation. In "Put the apple on the

towel in the box," the phrase "on the towel" is structurally ambiguous before "in the box" appears: it can be interpreted as either the direction of the action "put" or as a location modifier for "the apple." Using eye-tracking in the visual world paradigm, Tanenhaus et al. found that when only one apple was present in the visual scene, listeners were more likely to interpret "on the towel" as the direction of action (eye movements moved directly from the apple to the towel); when two apples were present, listeners were more likely to interpret it as a location modifier for "the apple" rather than the action direction (after fixating on the apple on the towel, they looked directly at the true target location—the box).

Second, the visual world paradigm provides new evidence for when different types of information are processed during syntactic analysis. Early two-stage accounts held that syntactic structure analysis precedes processing of other non-structural information (including lexical semantics, world knowledge, discourse) in sentence comprehension (initial syntactic analysis; Frazier, 1987). Constraint-based accounts argued that sentence processing involves joint constraints from multiple levels of information (Trueswell et al., 1994), which influence syntactic structure analysis early in sentence processing. Visual world paradigm studies support the latter hypothesis. For example, Snedeker and Trueswell (2004) studied ambiguous prepositional phrase structures ("Choose the cow with the stick" vs. "Tickle the pig with the fan"), where "with the stick/fan" could be either an object modifier or an instrument for completing the action. They found that information from the visual scene (number of objects) and verb bias (verbs biased toward modifier interpretation like "choose" vs. verbs biased toward instrument interpretation like "tickle") both influenced syntactic structure analysis of ambiguous sentences early in processing, manifested as listeners looking at different target objects depending on object number and verb bias. Additionally, Chambers et al. (2002, 2004) found that world knowledge related to object shape, size, and properties also influenced syntactic structure analysis, and these effects occurred at the earliest stages of sentence processing, refuting the theoretical assumption that syntactic structure analysis occurs first.

3.3 Semantic Predictive Processing

A major contribution of the visual world paradigm to semantic processing research is revealing that semantic processing is not only immediate but often predictive (Altmann & Kamide, 1999; Kamide et al., 2003; theoretical review: Pickering & Gambi, 2018). Altmann and Kamide (1999) first used the visual world paradigm to study the time course of verb-argument integration: compared to an irrelevant verb like "move," listeners hearing the verb "eat" in "the boy will eat…" fixated on the cake in the visual scene earlier. This demonstrates that verb semantic information (i.e., that "eat" requires edible arguments) helps listeners predict argument referents. Kamide et al. (2003) summarized key characteristics of semantic processing: (1) the combination of verb and subject jointly facilitates semantic prediction—for example, the combination of subject "the man"

and verb "ride" predicts a high-probability object like "motorbike"; (2) besides verbs, case markers attached to arguments also activate predictive processing, as in verb-final Japanese where listeners can predict upcoming argument referents through case markers even before the verb appears.

Visual world paradigm research on semantic processing extends beyond verb-argument structures. Chow and Chen (2020) used this paradigm to study the integration of Chinese classifier information with world knowledge in context, finding that Chinese speakers can form expectations about upcoming nouns early in processing based on world knowledge in the context, and these expectations are influenced by classifiers and further revised later in processing. Additionally, Grüter et al. (2020) investigated classifier processing in L1 and L2 Chinese speakers, finding that both groups are sensitive to grammatical collocational information contained in classifiers and use this information for predictive processing. However, L2 speakers rely more on semantic information in processing (e.g., the classifier "tiáo" collocates with long, thin objects), showing increased fixations on distractors that violate classifier collocations but match semantic features of long, thin objects.

3.4 Discourse-Level Processing

The visual world paradigm can be used to investigate two important issues in discourse comprehension: referential relations and connective relations. First, eye-tracking in the visual world paradigm can effectively test the establishment process between pronouns and their antecedents. It is generally believed that when listeners hear a pronoun that corefers with prior discourse and fixate on a relevant object, this indicates that the object is considered a potential target referent (Runner et al., 2003). Based on this mechanism, researchers have used the visual world paradigm to investigate numerous time course issues in referential processing. For example, Arnold et al. (2000) first found that gender cues and the order in which referents are mentioned both have immediate effects on referent resolution: listeners can use gender-marked forms (e.g., English singular third-person "he" or "she") to lock onto referents early in processing; simultaneously, the first-mentioned person in a sentence (e.g., the subject in SVO sentences) is more likely to be interpreted as the referent. In research on how implicit causality affects pronoun resolution, Pyykkönen and Järvikivi (2010) found that implicit causality effects appear immediately after the verb: listeners fixate more on the person biased by the verb after hearing it—for example, in "John frightened Bill because…," the verb "frighten" is biased toward the first person, so listeners fixate more on John when hearing "frightened"; in "John feared Bill because…," the verb "feared" is biased toward the second person, so listeners fixate more on Bill when the verb appears. This finding demonstrates that referential processing occurs immediately and even predictively, rather than through delayed integration (see also: Cozijn et al., 2011).

The visual world paradigm also provides rich empirical evidence for the real-time establishment of connective relations in language comprehension. Wei et

al. (2019) used the visual world paradigm to investigate the processing of subjective causal relations (claim-evidence) and objective causal relations (cause-effect) and the role of Chinese connectives. The study found that when listeners heard the subjective causal connective "kějiàn" ("it can be seen that") compared to the objective causal connective "yīn'ér" ("therefore"), they fixated more on the speaker in the visual scene. This suggests that processing of subjective and objective causal relations may differ in terms of confirming and tracking the speaker, and that speaker tracking occurs immediately with the input of subjective causal connectives, providing experimental evidence for the immediacy of discourse processing. Mak et al. (2017) investigated the role of two Russian connectives in establishing connective relations by providing two alternative referents in the visual scene and tracking listeners' fixation patterns. The study found that connective "i" ("and," marking continuation with consistent subjects across clauses) and connective "a" ("and/but," marking shift with different subjects across clauses) helped both monolingual and bilingual children predict whether the subject of the second clause would shift, confirming predictive processing in spoken discourse comprehension.

3.5 Extraction and Processing of Pragmatic Information

When pragmatic implicatures are processed and whether this process precedes semantic analysis are important issues in pragmatics. The literal-first hypothesis (Huang & Snedeker, 2009, 2011) holds that processing of literal semantic meanings of scalar terms (e.g., "some" should be interpreted as: some—and possibly all) precedes processing of their pragmatic implicatures (some—but not all). Levinson (2000) argues that pragmatic implicatures are processed automatically by default; constraint-based processing theory suggests whether pragmatic implicatures are activated preferentially depends on sufficient contextual support (Degen & Tanenhaus, 2015, 2016).

The visual world paradigm is an important experimental tool for comparing the time courses of semantic and pragmatic information processing. Huang and Snedeker (2011) found in a visual world paradigm eye-tracking experiment that when processing "some," listeners first fixated on objects compatible with the semantic interpretation of "some" (some—and possibly all), while using the pragmatic implicature of "some" (some—but not all) to disambiguate and exclude the referent "all" occurred later than semantic processing of "some" (approximately 800 ms later). Degen and Tanenhaus (2016) found that this delayed pragmatic implicature processing only occurred when number words also appeared in the instructions; when number words were absent, pragmatic implicature processing of "some" was not later than literal semantic processing. Gardner et al. (2021) improved the visual object set from Huang and Snedeker (2011) to better match the concept of "some," finding that when sufficient contextual support exists, pragmatic implicature processing is rapid and immediate—listeners can use the pragmatic implicature of "some" to quickly lock onto target objects. Additionally, language users' processing of pragmatic

information is greatly influenced by speaker credibility—when facing highly credible speakers, participants can use pragmatic meanings of scalar adjectives to lock onto target objects earlier, whereas no early pragmatic processing effects appear when facing less credible speakers (Gardner et al., 2021).

4. Main Contributions, Limitations, and Future Directions of the Visual World Paradigm

The eye-tracking visual world paradigm provides two important types of information for language comprehension research: visual dimension fixation measures and precise temporal measurement. The former offers rich possibilities for experimental design in psycholinguistics and cognitive psychology; the latter provides accurate time course information for spoken language processing at all levels—phonological, lexical, syntactic, semantic, discourse, and pragmatic—greatly expanding relevant theories of language comprehension. Combining these two types of information can effectively reflect how listeners' fixation locations in visual scenes change over time upon receiving spoken language input, thereby providing direct evidence for the important issue of time course in language comprehension. Visual world paradigm experiments analyzing highly time-sensitive eye-tracking data have found that processing at all linguistic levels shows immediate and even predictive characteristics, differing from delayed-integration findings in some early studies and demonstrating that research results on language processing time courses are closely tied to the methods employed. Furthermore, because the visual world paradigm primarily relies on listening tasks and does not require participants to have full literacy skills, it can be used to investigate language processing in young children, L2 learners, and populations with specific language impairments (research examples: Canseco-Gonzalez et al., 2010; McMurray et al., 2010; Weber & Cutler, 2004).

One major limitation of the visual world paradigm is its inability to provide processing duration data, thus preventing investigation of issues related to processing difficulty in language comprehension (Salverda & Tanenhaus, 2018). Additionally, visual world paradigm experiments can only present a limited number of static objects in visual space, which differs from the complex visual environments of everyday language comprehension. Real language comprehension environments may include more objects and dynamic actions and events, limiting the generalizability of findings from this paradigm (Huettig et al., 2011). Moreover, in experimental environments presenting only a limited number of objects, listeners may form expectations about linguistic input in advance and strategically fixate on certain objects, meaning that eye movement trajectories may not fully reflect language processing processes (Henderson & Ferreira, 2004). In response to this concern, Dahan and Tanenhaus (2004), based on their lexical recognition research, argued that word frequency effects on lexical recognition are not influenced by the presence or number of competitors in visual space, suggesting that presenting a limited number of objects in visual space does not affect the validity of the visual world paradigm.

Eye-tracking research using the visual world paradigm still has considerable room for development. First, although the assumptions about visual and linguistic information comprehension processes proposed in linking hypothesis theories have been confirmed by substantial empirical evidence, the important role of task goals in language processing remains to be further explored. Comparing how language processing unfolds over time under different task goals will be one direction for future visual world paradigm eye-tracking research. In recent years, eye-tracking research has begun using three-dimensional virtual reality (VR) technology, which can highly approximate natural language communication scenarios while maintaining precise experimental control. Some visual world paradigm eye-tracking experiments using VR technology have successfully replicated classic findings in language processing, such as predictive language processing (Eichert et al., 2018; Heyselaar et al., 2020). Such technological improvements not only enhance the ecological validity of the visual world paradigm but can also test factors influencing language processing processes in near-authentic language use environments. Both theoretical and technological innovations provide new opportunities and greater possibilities for more accurate and effective collection and interpretation of eye-tracking data and for exploring language processing.

References

林桐, 王娟. (2018). 基于视觉情境范式的口语词汇理解研究进展. 心理技术与应用, 6(09), 570-576.

邱丽景, 王穗苹, 关心. (2009). 口语理解的视觉－情境范式研究. 华南师范大学学报, 1, 115-122.

Allopenna, P. D., Magnuson, 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, 22*(1), 1–12. https://doi.org/10.1016/j.cub.2009.12.014

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition, 93*(2), 79–87. 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. (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., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4). https://doi.org/10.1111/j.1551-6709.2009.01022.x

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 eyetracking. *Cognition, 76*(1), B13–B26. https://doi.org/10.1016/s0010-0277(00)00073-1

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General, 143*(1), 404–413. https://doi.org/10.1037/a0031813

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or cárcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes, 25*(5), 669–705. https://doi.org/10.1080/01690960903474912

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*(1), 30–49. https://doi.org/10.1006/jmla.2001.2832

Chambers, C. G., Tanenhaus, M. K., & Magnuson, 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

Chow, W. Y., & Chen, D. (2020). Predicting (in)correctly: Listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience, 35*(9), 1149–1161. https://doi.org/10.1080/23273798.2020.1733627

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken Language. *Cognitive Psychology, 107*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-x

Corps, R. E., Brooke, C., & Pickering, M. J. (2021). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language, 122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. . (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language, 64*(4), 381–403. https://doi.org/10.1016/j.jml.2011.01.001

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), 498–513. https://doi.org/10.1037/0278-7393.30.2.498

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science, 39*(4), 667–710. https://doi.org/10.1111/cogs.12171

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

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods, 50*(3), 1102–1115. https://doi.org/10.3758/s13428-017-0929-z

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*(3), 165–182. https://doi.org/10.1016/j.jml.2013.06.001

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Lawrence Erlbaum Associates.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 35*(4), 517–543. https://doi.org/doi.org/10.1006/jmla.1996.0028

Gardner, B., Dix, S., Lawrence, R., Morgan, C., Sullivan, A., & Kurumada, C. (2021). Online pragmatic interpretations of scalar adjectives are affected by perceived speaker reliability. *PLoS ONE, 16*(2), e0245130. https://doi.org/10.1371/journal.pone.0245130

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience, 35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*(1), 105–115. https://doi.org/10.1016/j.cogsci.2003.10.002

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 1–58). Psychology Press. https://doi.org/10.4324/9780203488430

Henry, N., Jackson, C. N., & Hopp, H. (2022). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research, 38*(3), 397–422. https://doi.org/10.1177/0267658320963151

Heyselaar, E., Peeters, D., & Hagoort, P. (2020). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience, 36*(4), 440–461. https://doi.org/10.1080/23273798.2020.1859568

Huang, Y., & 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

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45*(6), 1723–1739. https://doi.org/10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes, 26*(8), 1161–1172. https://doi.org/10.1080/01690965.2010.508641

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

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011a). 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. (2011b). 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

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins.

Kaiser, E. (2016). Discourse-level Processing. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 151–184). John Benjamins Publishing.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive procesing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37–55. https://doi.org/10.1023/a:1021933015362

Karimi, H., Brothers, T., & Ferreira, F. (2019). Phonological versus semantic prediction in focus and repair constructions: No evidence for differential predictions. *Cognitive Psychology, 112*, 25–47. https://doi.org/10.1016/j.cogpsych.2019.04.001

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}

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54*, 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Koring, L., Mak, P., & Reuland, E. (2012). The time course of argument reactivation revealed: Using the visual world paradigm. *Cognition, 123*(3), 361–379. https://doi.org/10.1016/j.cognition.2012.02.011

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition, 46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance, 48*(5), 531–547. https://doi.org/10.1037/xhp0000999

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science, 3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Mak, W. M., Tribushinina, E., Lomako, J., Gagarina, N., Abrosova, E., & Sanders, T. (2017). Connective processing by bilingual children and monolinguals with specific language impairment: Distinct profiles. *Journal of Child Language, 44*(2), 329–345. https://doi.org/10.1017/s0305000915000860

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380. https://doi.org/10.3758/bf03206780

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*(3), 283–312. https://doi.org/10.1006/jmla.1997.2543

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language, 33*(1), 128–147. https://doi.org/10.1006/jmla.1994.1007

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R.* CRC Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7*, 1–24. https://doi.org/10.7554/eLife.33468

Porretta, V., Kyröläinen, A. J., Rij, J. Van, & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. Howlett, & L. Jain (Eds.), *Intelligent Decision Technologies 2017. Smart Innovation, Systems and Technologies* (Vol. 73, pp. 268–277). Springer.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pyykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually Situated Language Comprehension* (pp. 67–82). John Benjamins Publishing.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology, 57*(1), 5–16. https://doi.org/10.1027/1618-3169/a000002

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660. https://doi.org/10.1037/0033-2909.85.3.618

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements. *Cognition, 89*, B1–B13. https://doi.org/10.1016/S0010-0277(03)00065-9

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. (2018). The visual world paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide* (pp. 89–110). Wiley-Blackwell.

Saryazdi, R., & Chambers, C. G. (2021). Real-time communicative perspective taking in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(3), 439–454.

Saslow, M. G. (1967). Latency of saccadic eye movement. *Journal of the Optical Society of America, 57*(8), 1030–1033. https://doi.org/10.2466/pms.2003.96.1.173

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the bootstrapped differences of timeseries (BDOTS) to analyze visual world paradigm data (and more). *Journal of Memory and Language, 102*, 55–67. https://doi.org/10.1016/j.jml.2018.05.004

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language, 42*(3), 423–443. https://doi.org/10.1006/jmla.1999.2691

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition, 24*(5), 833–841. https://doi.org/10.1017/s1366728920000607

Tanenhaus, M. K., Magnuson, J. S., & Dahan, D. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580. https://doi.org/10.1023/a:1026464108329

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.7777863

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology, 50A*(3), 481–497. https://doi.org/10.1080/027249897391982

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25. https://doi.org/10.1016/S0749-596x(03)00105-0

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

Note: Figure translations are in progress. See original paper for figures.