encyclopedia

MDPI

Entry

Tracking Eye Movements as a Window on Language Processing: The Visual World Paradigm

Marta Tagliani * and Michela Redolfi

Department of Cultures and Civilizations, University of Verona, 37129 Verona, Italy
* Correspondence: marta.tagliani@univr.it

Definition: This entry overviews the pioneering experimental studies exploiting eye movement data to investigate language processing in real time. After examining how vision and language were found to be closely related, herein focus the discussion on the evolution of eye-tracking methodologies to investigate children's language development. To conclude, herein provide some insights about the use of eye-tracking technology for research purposes, focusing on data collection and data analysis.

Keywords: visual world; eye-tracking; language processing; language acquisition

check for updates

Citation: Tagliani, M.; Redolfi, M. Tracking Eye Movements as a Window on Language Processing: The Visual World Paradigm. *Encyclopedia* **2023**, *3*, 245–266. https://doi.org/10.3390/encyclopedia3010016

Academic Editors: Nikos A. Salingaros and Raffaele Barretta

Received: 22 December 2022
Revised: 13 February 2023
Accepted: 23 February 2023
Published: 25 February 2023

1. Introduction

Until the 1970s, experimental studies on linguistic competence and processing have exclusively relied on offline measures of comprehension. In classical psycholinguistic paradigms such as lexical decision [1] or sentence–picture verification tasks [2,3], participants are asked to evaluate the truthfulness of the linguistic input provided, either against pictures or their word knowledge. In these paradigms, sentence comprehension is assessed by measuring participants' response latencies and accuracy in expressing metalinguistic evaluations after being presented with the linguistic stimulus. However, while response choices and reaction times are behavioral measures that provide information on linguistic comprehension, such tasks do not tap into real-time processing of spoken language, and, as a consequence, reveal less about the speaker's efficiency and knowledge.

Another paradigm is the visual world paradigm, an experimental methodology which employs the recording of participants' eye movements during listening tasks. Unlike long-established psycholinguistic paradigms, eye movement data provide exhaustive information on the time course of language comprehension as well as relevant insights on how visual and linguistic sources of information interact in real-time. In a typical visual world set-up, participants are instructed to listen to sentences carefully and look wherever they want on the screen or interact with objects or screen-based pictures (e.g., by moving them). The simplicity of such set-up makes the task execution extremely effortless, as it relies on the human tendency to look at relevant parts of the visual scenario as critical words are mentioned. In fact, participants are not asked to do anything different from what they do in their everyday life, when they automatically integrate information from visual or written and spoken sources of information (e.g., while listening to the news on TV). The unchallenging nature of visual world studies makes this experimental paradigm extremely suitable to investigate language comprehension in populations with language disorders as aphasia [4,5] or developmental dyslexia [6–10], as well as in infants and young children [11–13].

This entry offers a detailed overview of how the visual world paradigm can be used efficiently to assess linguistic comprehension in children. In Section 2.1, the entry will review the pioneering eye-tracking studies, which have led to the affirmation of the visual world paradigm in psycholinguistic research. In Section 2.2, the entry will illustrate the main experimental procedures typically used in the visual world paradigm with both adult and child participants. The remainder of the entry will be devoted to the discussion of the different eye-tracking methodologies exploiting the relation between language and vision to study online language processing by infants and children, namely the Preferential-Looking Paradigm (Section 3.1) and the Looking-While-Listening Task (Section 3.2). Specific limitations and advantages of these different tasks will also be discussed. In conclusion, the entry will give some details about the eye-tracking technology, focusing on data collection and data analysis, and discussing some methodological limitations (Section 3.3).

2. Using the Visual World Paradigm to Study Language Processing

2.1. First Studies Tracking Eye Movements

The first observation that eye movements follow a pattern that is strictly related to a cognitive goal came from the seminal study conducted by Yarbus (1967) [14], who showed that subjects tend to look for visual referents that can provide useful information in a specific visual context. The most influential work in the exploration of the relationship between language and vision, however, was developed a few years later by Cooper (1974) [15]. He asked a group of adults to listen to a short text while looking at a display showing common objects, some of which were named in the spoken narrative. Participants were simultaneously presented with short stories and a visual display containing black and white drawings of common concrete objects (e.g., a lion, a dog, a zebra, a snake, and a camera). The visual scenario was manipulated so that the pictures on the screen either depicted objects that were directly mentioned, or were semantically related to target words presented in the spoken text (in italics in (1)). Consider, for instance, the short narrative about a safari in Africa in (1): while words such as *lion* and *zebra* have a direct visual referent on the screen, the word *Africa* is only semantically related to the pictures of animals such as a lion, a zebra, and a snake, which are known to be part of the African wildlife.

(1) While on a *photographic safari* in Africa, I managed to get a number of breath-taking shots of the wild terrain. (. . .) When I noticed a hungry *lion* slowly moving through the tall glass toward a herd of grazing *zebra*.

During the task, participants' eye movements were recorded using an eye movement camera system [16]. Despite no explicit instructions being given, Cooper found that participants focused their gaze more toward the objects that were mentioned in the text (e.g., the lion), than toward those that were not (e.g., the snake). Similarly, upon hearing *Africa,* their visual attention was drawn by pictures of African animals (e.g., the zebra and the lion) rather than by the picture of an unrelated animal (e.g., the dog). In addition, Cooper observed that the looks at the objects in the visual scenario were closely time-locked to the presentation of the linguistic input (within 200 ms after word offset). This indicates that listeners are able to actively exploit anticipatory cues from the speech stream, such as word initial phonemes or syllables, to make predictions concerning the upcoming linguistic information. This finding represents the first experimental evidence that spoken language guides visual attention: language-oriented eye movements are often fast and unconscious, as they reflect the online incremental activation of word semantics during the unfolding of the linguistic input.

In line with these findings, Just and Carpenter (1980) accounted for eye movements and fixations during written language comprehension [17]. College students were presented with technical texts about unfamiliar topics (e.g., the properties of flywheels). They were asked to read these passages as naturally as possible, and to recall their content after reading. During the experimental session, participants' eye movements were recorded by a television camera. Quite surprisingly, results showed that the duration of the fixations significantly differed from word to word within each passage. This evidence has led to the formulation of the influential *eye-mind hypothesis*, which hypothesizes that eye movements reflect the cognitive processes involved in the comprehension of written language. During reading, the parser fixates a word while processing it, and the duration of this fixation reflects the processing load required during comprehension. Hence, the readers would make longer fixations when the processing is more effortful, such as when encountering infrequent or complex words in the text, or when integrating contextual information provided either by previous linguistic information or by their world knowledge.

These studies represent the first pieces of experimental evidence of a real-time interaction between visual attention and language comprehension, as they showed that the parser immediately integrates written and spoken linguistic information with the visual context. Cooper (1974) recognized the advantages of eye movement analysis for investigating how perceptual and cognitive processes jointly determine the online understanding of linguistic sentences, fostering eye-tracking as a useful research methodology. He believed to have found a "practical new research tool for the real-time investigation of perceptual and cognitive processes and, in particular, for the detailed study of speech perception, memory, and language processing" [15] (p. 84). Nevertheless, his work has been largely ignored for more than twenty years, until the early 90s, when the psycholinguistic community began to exploit the systematic relationship between eye movements and speech processing on a large scale with the rise of the so-called *visual world paradigm* [18,19].

The study by Tanenhaus and colleagues (1995) [19] had a key role in the development of this research field. Thanks to a head-mounted video-based eye tracker, they investigated the effects of the visual context on language comprehension and how visual contextual information can affect syntactic processing. Participants were presented with sentences either containing temporary syntactic ambiguities or not, along the lines of the examples reported in (2) and (3). In (2), the prepositional phrase *on the napkin* is ambiguous between being a modifier of the Determiner Phrase (henceforth, DP) *the pear* (i.e., indicating the location of the pear to be picked up) and indicating the destination of the action (i.e., the place where the pear has to be put).

(2) Put the pear on the napkin on the table
(3) Put the pear that is on the napkin on the table

While listening to these instructions, participants were presented with two types of visual scenarios that supported one of the possible interpretations of the ambiguous Prepositional Phrase (henceforth, PP). The one-referent context (Figure 1A) contained four sets of objects: a football, a table, a napkin, and a pear placed on a napkin. Such a visual scenario suggested an interpretation of the PP *on the napkin* as destination place. When hearing the DP *the pear*, the listener was immediately able to identify the object to be moved because there was only one pear in the context. Hence, they were likely to assume that *on the napkin* referred to the destination of the action of *putting* rather than to another peculiar property of the pear itself. Conversely, in the two-referent context, a second pear—placed on a napkin—was presented instead of a football (Figure 1B). Here, the DP *the pear* did not have a univocal visual referent, and, hence, it was more likely for the listener to interpret the PP *on the napkin* as a modifier providing specific information about which pear had to be moved.

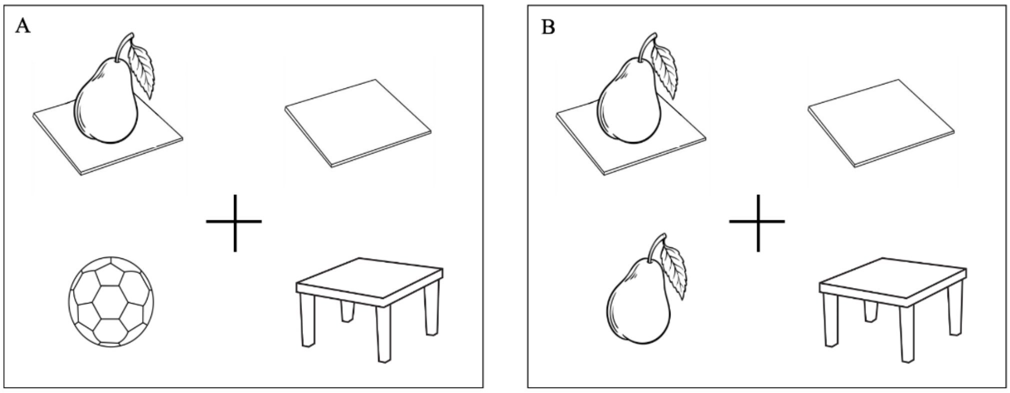

Figure 1. Examples of visual scenarios for the one-referent (**A**) and the two-referent (**B**) contexts modeled after [19] (Made by the authors).

If this type of syntactic ambiguity is resolved independently from the communicative context in which it is presented, the parser would always show a clear preference for the interpretation of the PP *on the napkin* as destination of the action for reasons of syntactic requirements of the ditransitive verb *to put* (attachment preferences are not necessarily driven by verbs' semantic given the findings concerning parsing principles (a.o., [20–22]). Instead, if visual contextual cues are integrated in the syntactic processing as soon as the linguistic input unfolds, a relevant experimental context might influence the parsing and the resolution of the syntactic ambiguity, resulting in a different interpretation of the same PP *on the napkin* in the two referent-conditions. The eye movement analysis showed indeed different patterns of fixations in the two visual contexts. When an ambiguous sentence such as (2) was presented in a one-referent context, participants initially focused their gaze on the empty napkin and switched toward the table only after sentence offset, indicating that they had initially interpreted the PP *on the napkin* as destination place. This was further confirmed by participants' looking pattern while listening to the unambiguous sentence such as (3), during which, after the individuation of the pear placed on the napkin, they immediately focused on the table without paying attention to the incorrect destination (e.g., the empty napkin). This demonstrates that, in the two-referent context, the PP *on the napkin* was immediately interpreted as modifier of the object and not as destination, regardless of the syntactic configuration of the test sentence. By monitoring eye movements, Tanenhaus et al. (1995) clearly demonstrated that visual context can significantly affect spoken language comprehension from the beginning of syntactic processing [19].

2.2. Procedures and Variants of the Visual World Paradigm

A few years later, Allopenna et al. (1998) [18] developed the first study using a screen-based presentation to investigate the time course of spoken word recognition in continuous speech and contextually coined the term *visual world paradigm* (see [23] for a detailed overview).

Setting up a visual world experiment involves making predictions about the distribution of fixations to a target object relative to other elements in the visual display at some critical points in the speech stimulus [24]. On a typical trial using this paradigm, participants hear an utterance while looking at an experimental display. During each trial, eye movements are recorded with an eye tracker. There are two common variants of visual world experiments: *look-and-listen* studies and *task-* or *action-based* studies.

Look-and-listen studies (sometimes misleadingly called *passive listening studies*) do not require participants to perform any explicit task. In a popular version of this study, introduced by Altmann and Kamide (1999) [25], visual stimuli consist of line drawings of semi-realistic scenes (Figure 2), while the auditory stimuli are utterances describing or commenting upon (some) pictures on the screen (e.g., *The boy will eat the cake*). The screen usually shows the objects mentioned in the sentences (e.g., a boy and a cake) and distractors which are never mentioned (e.g., a ball). As the interpretation of the language is co-determined by information in the visual scene, the listener's attention is drawn to referents, including pictures that the participant anticipates will be mentioned as the input unfolds, or pictures associated with specific implied events.

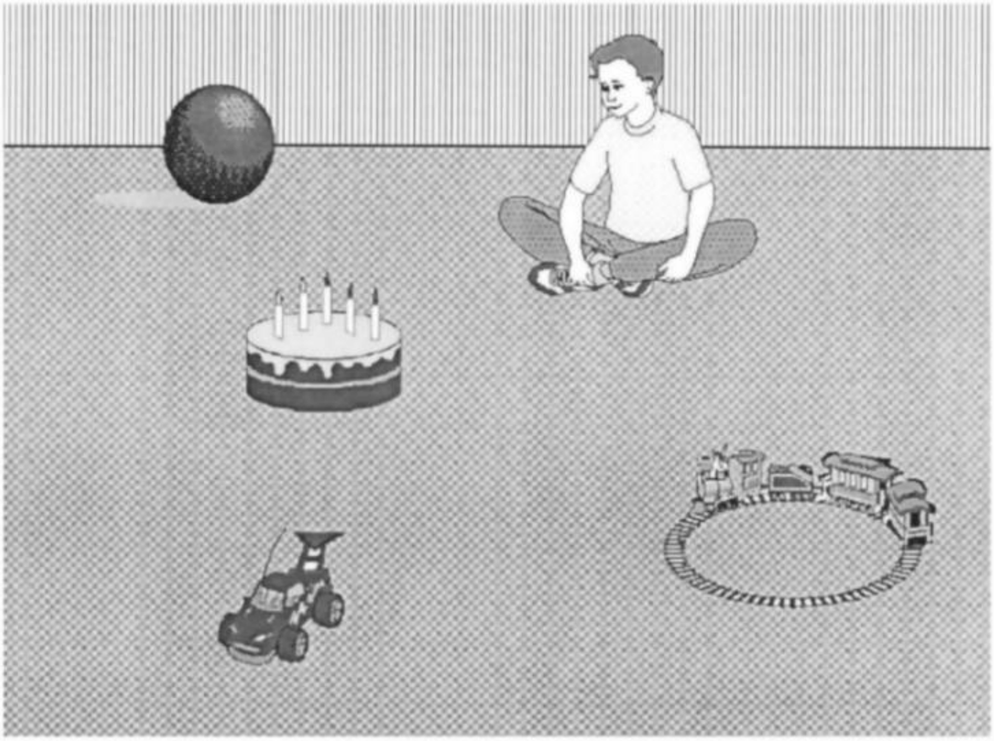

Figure 2. Example of a semi-realistic scene for the sentence *The boy will eat the cake.* Reprinted with permission from [25]. Copyright 1999 Elsevier.

Task or *action-based* studies display sets of objects, either laid out on a workspace (e.g., [19,25] or shown as drawings on a computer screen (e.g., [18]). In such tasks, participants interact with real-world objects or screen-based pictures to perform a motor task such as clicking and dragging pictures to follow explicit instructions (e.g., *Put the clown above the star*), clicking on a picture when its name is mentioned, or manipulating real objects (e.g., *Pick up the apple. Now put it in the box*). An example is given in Figure 3.

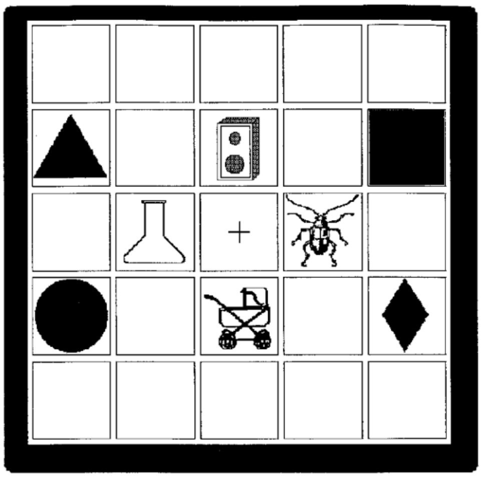

Figure 3. Example of a typical display in task- or action-based studies taken reprinted with permission from [18]. Copyright 1998 Elsevier. Participants were asked to, e.g., *Pick up the beaker and put it above the triangle*.

Visual world studies vary as for the complexity of the visual scene presented to participants. In the simplest case, the visual scene displays the target object and one unrelated distractor object. However, to test specific hypotheses about linguistic variables, either the linguistic stimulus or the pictures on the display can themselves be systematically manipulated, creating local or temporary ambiguities. For example, in order to investigate whether 5–7-year-old children would use their knowledge of the speaker's visual perspective to constrain reference, Nadig and Sedivy (2002) [26] created three different visual conditions, as shown in Figure 4 below.

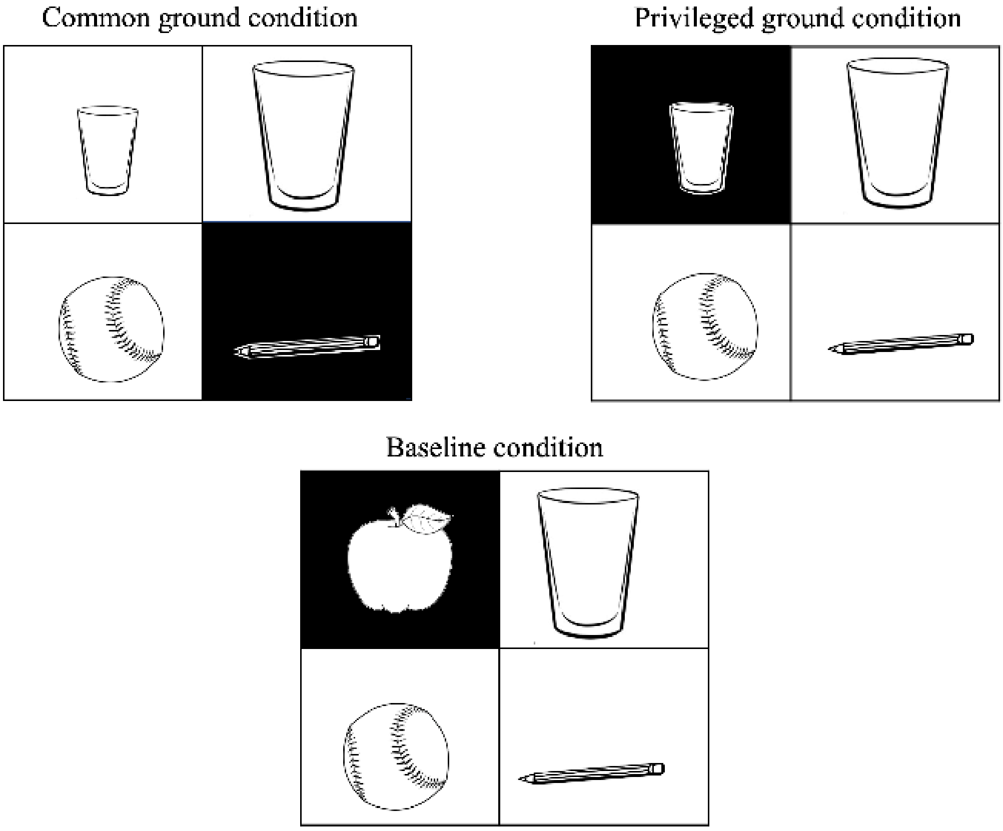

Figure 4. Sample displays for the three conditions tested in [26] (Permission pre-granted according to STM guideline). Shading denotes the wooden partition blocking the object from the speaker's view.

In this experiment, the child participant sat across the table from an adult speaker, each looking the opposite side of a vertical display case. One of the objects was blocked from the speaker's view, but not the child's (as indicated by the shaded background in Figure 4). For all three displays the speaker uttered *Pick up the glass*. While in the bottom display only one glass was present and target identification should occur right after the acoustic information, in the first two conditions a second glass was displayed, either visible or not by the speaker. This competitor objects created potential ambiguity for both the target reference—allowing for targeted predictions to be made—and for the child's taking into account of the speaker's knowledge. Thus, if differences in perspective are considered through the trials, eye movements in the privileged-ground condition should pattern similarly to the baseline. In such a design, the presence of the competitor in the top left display resulted in a globally ambiguous reference.

To avoid infelicitous stimuli (as in the common-ground condition above), many eye-tracking studies introduce *temporary* referential ambiguities. For example, in a study by Sedivy et al. (1999) [27], the display contained two tall objects (e.g., a glass and a pitcher, see Figure 5) with a target instruction such as *Pick up the tall glass*. Here, the instruction as a whole refers to a single object (i.e., the tall glass) and is perfectly felicitous. However, a temporary ambiguity is created as *tall* can refer to two items, thus influencing eye movements during the *tall* window.

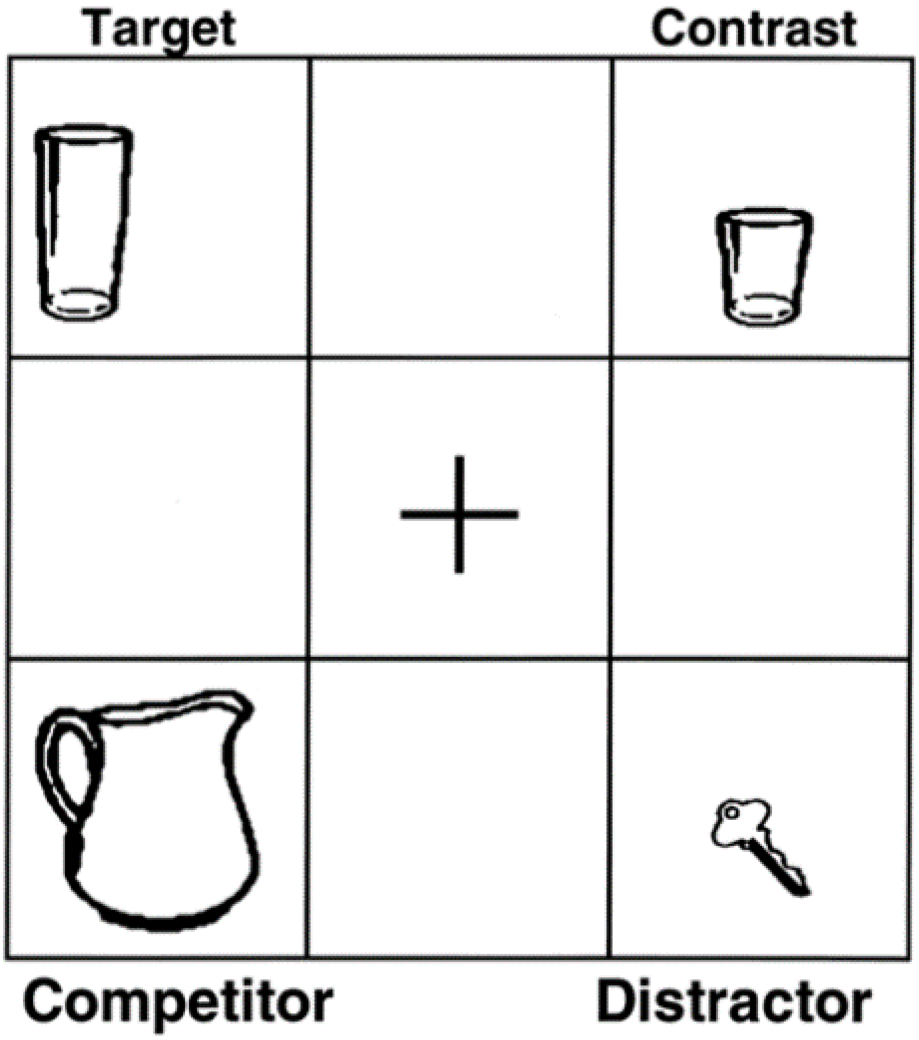

Figure 5. Example of the visual display reprinted with permission from [27]. Copyright 1999 Elsevier.

In conclusion, observing eye movements opens a window on the automatic interpretation of a sentence given a specific visual context, allowing to go further than behavioral response. Indeed, one great advantage of the visual world paradigm, compared to other psycholinguistic techniques, is that listeners do not need to perform any metalinguistic judgments, which might be difficult or impossible to elicit from some groups of listeners, including young children. Since this paradigm solely relies on the listeners' tendency to direct their gaze toward relevant parts of the display as they are mentioned, in previous decades, this experimental methodology has become a fundamental tool to study language development and competence in children, including pre-verbal infants. The next session will be dedicated to the evolution of the visual world paradigm to study young children's linguistic development.

3. Tracking Children's Eye Movements

Months before they speak their first words, young children reveal their developing language knowledge by responding meaningfully to the speech they hear. However, since the comprehension of the linguistic input can only be inferred through children's behavior in a specific context, receptive language competence has been less accessible than their speech production skills. In the last four decades, many valuable experimental techniques have been employed to investigate the emergence of language comprehension. Research on early cognitive abilities has examined how infants become attuned to sound patterns in the environment language over the first year [28] and how they attend to speech patterns relevant to their native language structure [29]. These studies made it possible to explore how first year infants become skilled learners, how they are able to make distributional analyses of phonetic features of their language and how they form acoustic–phonetic representations based on frequently heard sound patterns [30]. Quite interestingly, other studies have demonstrated that newborns are also able to perceive categorial phonemic distinctions [31], and they are able to do so even for distinctions not pertaining to the language of exposure [32].

As opposed to sound perception, learning words is said to come later, between 6 and 15 months of age, when infants start to understand other people's intention and reveal this progress though increasingly differentiated verbal and behavioral responses to speech. Early scientific studies on developmental language comprehension made use of methodologies such as: (i) diary studies providing observational data on early comprehension abilities, e.g., [33]; (ii) studies of vocabulary growth, e.g., [34]; (iii) naturalistic experiments on comprehension analyzing the understanding of familiar words, e.g., [35]; and (iv) novel word learning, e.g., [36]. However, "Language production reflected the observable half of children's language ability; comprehension was the other, inaccessible half of what children knew about language. Just as astronomers were not satisfied to study only the light side of the moon, researchers in language acquisition recognized that the dark side—language comprehension—held secrets to a process that had to be unlocked" [37] (p. 317).

To overcome these hurdles, different techniques have been developed to study language comprehension in real time. Among the most common, herein focuses on those that derived from Cooper's simple assumption that the probability of looking at an object increases when the object is mentioned [15]. This section offers an overview of the different eye-tracking methodologies exploiting the relation between language and vision to study online language processing by infants and older children, and how they have improved over time.

3.1. The Preferential-Looking Paradigm

In 1963, the developmental psychologist Robert Fantz published the first study using the preferential-looking method with young children, showing that newborns looked selectively at some visual stimuli (patterned images) over others (uniform images) [38]. In 1974, Horowitz asked whether visual fixations to images could be used as a window onto language development, see [36,38]. Then, Spelke (1976, 1979), in a dynamic version of

Fantz's (1958, 1964) paired-comparisons method [39,40], developed the preferential-looking paradigm to study intermodal perception in infants [41,42]. She presented four-month-old infants with two visual stimuli (e.g., a person clapping hands and a donkey falling onto a table) while presenting an auditory stimulus (e.g., the sound of hands clapping) and found that children looked more at the screen in which the event matched the picture than at the screen in which it did not. Although Spelke's study did not test children's language knowledge, this inspirational study led to the adaptation of this auditory–visual matching procedure to investigate the development of language comprehension in the early years of life.

At the end of the 70s, the first experimental procedures for testing infants' knowledge of object words were introduced. Benedict (1979) found that 12-month-old children would reliably direct to a familiar object when it was named, even when nonverbal behaviors of the speaker (including pointing and gaze) were eliminated, which often allowed toddlers to appear more linguistically capable than they really were [35]. A second innovative study was conducted by Thomas and colleagues (1981), who used eye movements as an index of word recognition comparing the ability of 11- and 13-month-old infants to identify a familiar named object among a series of competitors [43]. This initial finding was fundamental to enable the assessment of word recognition more objectively by exploiting, as a dependent measure, the time spent looking at a particular item over the other (see [44]). This technique allowed for the standardization of stimulus presentation, for a careful definition of what counted as a correct response, and for the elimination of nonverbal cues. In addition, for questions that focus on comprehension of verbs and events requiring motion, the advent of videotaped stimulus displays opened a new window into the exploration of language knowledge (e.g., [45]; see [46] for a detailed review).

The groundwork studies of Spelke (1976) and Thomas et al. (1981) provided a starting point for later research in which preferential-looking measures were further adapted to assess early language comprehension (e.g., [47,48]). More specifically, the method of Golinkoff et al. (1987) [49], known as the *intermodal preferential-looking paradigm* (IPLP), was revolutionary in that it combined use of visual and auditory stimuli simultaneously, soon becoming one of the most used experimental designs using this technique. In the procedures of these studies, infants sit on a parent's lap in the middle of two television monitors. A concealed audio speaker midway between the two monitors plays a linguistic stimulus that matches only one of the displays shown on the screens. Mounted atop the speaker is a light that comes on during each intertrial interval to ensure that the infant makes a new choice about which screen to look at on each trial. A hidden camera records the child's visual behavior. Researchers typically used the total looking time to the target picture and the duration of the longest look to the target picture as indexes of comprehension. With these measures, the IPLP allowed for the investigation of several different questions about burgeoning knowledge in the areas of phonology, semantics, syntax, and morphology in infants who are not yet speaking.

In the first paper using the IPLP, Golinkoff et al. (1987) conducted a series of experiments. In the first two, they examined whether 16-month-old infants could understand nouns (e.g., *dog, shoe*) and verbs (e.g., *drink, wave*). For example, in the noun experiment infants saw two static objects (e.g., a shoe and a boat) and heard, *Where's the boat? Find the boat!* In the verb experiment, infants saw two dynamic actions carried out by the same person (e.g., a woman drinking from a coffee cup and the same woman blowing on a sheet of paper) and heard *One is drinking and one is blowing. Which one is drinking?* Gaze patterns were coded in real-time through the use of a button box recording fixations on the target vs. the distractor and shifts between the two. Both the noun and the verb experiment showed that 16-month-old children looked significantly longer at, and oriented faster to, objects or events matching the linguistic stimulus they heard. Interestingly, although these participants had not begun to produce any verbs, they appeared to comprehend the verbs in the experiment.

In the third experiment, researchers also found that 28-month-old children who already produced multi-word sentences could use word order in a sentence to find which member of a pair of dynamic actions matched the language they were hearing. Visual events were constructed to differ only by who was performing an action and who was acted upon, as only verbs expressing reversable actions were included (e.g., *tickle*, *feed*). Hence, the task was very difficult for children who had to first analyze the visual stimuli to determine which character was the agent and which was the patient, and then use the language to find the particular event described. For example, on one monitor, toddlers saw Cookie Monster tickling Big Bird while Big Bird held a box of toys; on the other monitor, toddlers saw Big Bird tickling Cookie Monster. In the test trial, children heard, *Where's Cookie Monster tickling Big Bird?* Note that since both characters were moving, children could not just look to the event where the named character was in motion to solve the task. They found that 28-month-old children looked longer at the event that matched the sentence they heard over the event that contained the same participants and same action but depicted a reversed relationship between the participants. Moreover, using a similar paradigm, Hirsh-Pasek and Golinkoff (1996) found that 17-month-old infants were able to use word order in processing the described event [50]. These studies were the first reliable tests providing evidence of word order comprehension. Prior to these studies, researchers could only speculate about whether young children were sensitive to the grammar of their language before they actually started talking.

Since then, this technique has been adapted to investigate many other facets of language development. Golinkoff et al. (1995) investigated the ability of 3-year-olds to use Principles A and B of the Binding Theory [47]. As was the case for other syntactic and lexical phenomena studies, the IPLP found evidence for comprehension of these principles earlier than most other assessments. Naigles (1990) and Naigles and Kako (1993) [45,51] tested 2-year-old children's knowledge of verb meaning and sensitivity to meaning implications of transitive and intransitive sentence frames. Finally, the IPLP has been used to investigate lexical comprehension and production [52,53].

Advantages and Disadvantages

The IPLP is capable of revealing linguistic knowledge in young children for two reasons. First, unlike many other tasks used to explore language comprehension, this paradigm does not require children to point, select objects, answer questions, or act out commands. Children need merely employ fixations in order to fulfill the task requirements. Second, the paradigm usually does not set natural cues for understanding in conflict with each other, and it does not omit the contribution of these sources. In other words, in this paradigm, infants have access to syntactic, semantic, prosodic, and contextual information: when all these cues are provided, children may take advantage of what Hirsh-Pasek and Golinkoff (1996) called the "coalition" of cues, normally used in language comprehension to demonstrate the upper limits of their knowledge. Furthermore, this paradigm also made the general point that language development occurred more rapidly than previously thought. Language comprehension is ahead of language production and can be used as a vehicle to study emerging language knowledge [37]. The use of this methodology allowed researchers to find experimental evidence that infants analyze sentences they hear to find specific events in the world [46,50], that they are sensitive to the grammar found in sentences [48,52], and they even use sentence structure to glean something of the meaning of novel words [45,54–56].

Although it undoubtedly is a powerful laboratory tool, like all methods of investigation, the IPLP has its weaknesses. First, this paradigm can overestimate children's knowledge. This is because it always presents two alternatives, thus children could solve the task through elimination of alternatives or mutual exclusivity. Specifically, children might be tested on vocabulary or sentence structure and use their knowledge to discard one alternative to find the correct alternative [57–59]. Some of this depends on children's age, however, as infants are less likely to use this strategy [60].

A second limit is linked to the limited number of items researchers can study, given children's short attention spans, and the inability of the method to investigate individual differences in grammatical development [61]. "Although this method works well for group studies, it has proved impossible (at least so far) to adapt the preferential-looking technique for use with individual children. In the experiments they have conducted to date, Golinkoff and Hirsh-Pasek can obtain no more than four to six crucial target trials for any linguistic contrast. Although the results are quite reliable at the group level, the predicted pattern (i.e., preferential-looking at the pictures that match the language input) is typically displayed by only two thirds of the children with looking biases that average 66% for individual subjects. It should be clear why this kind of hit rate would be unacceptable for individual case studies." [61] (p. 228). Note, however, that one reason children do not look exclusively at the screen depicting the event matching the picture is that the tapes are specifically designed to be equally salient and to encourage active looking. Hence, the IPLP may not be suited to study individual differences by its very design [49].

In short, the potential for use of the IPLP is great. Although the paradigm shows some evident weaknesses, the advantages of the paradigm seem to outweigh the disadvantages. Indeed, it has been used to test children of various ages aiming to investigate a wide range of linguistic phenomena. In the next section, the entry will overview an evolved version of this paradigm, the *looking-while-listening task*, designed to overcome the weaknesses of the IPLP thanks to more refined technologies.

3.2. The Looking-While-Listening Task

Based on the studies by Thomas et al. (1981), Golinkoff et al. (1987), and Reznick (1990), in the 90s, Anne Fernald's research group developed a modified version of the preferential-looking method, the *looking-while-listening procedure*, to investigate whether particular features of child-directed speech might facilitate the identification of familiar words in fluent speech. Their initial goal in modifying the paradigm was "to increase the sensitivity, reliability, and validity of the measures, by making minor modifications to the procedure that served to eliminate confounding variables" [62]. According to them, earlier preferential-looking studies might potentially confound object salience with target status, since they used different stimuli as target and distractor objects. Moreover, some studies failed to counterbalance the side of target object presentation, which made it difficult to interpret infants' selective looking behavior unambiguously.

The first potentially influential change they undertook was to make sure that all target objects were also presented as distractors, to reduce the influence of object preference. Second, a major change concerned the measures used to capture infants gaze pattern in response to linguistic stimuli. Rather than coding eye movements in real-time using a button box, they began to code eye movement from videotapes, frame by frame, in slow motion. This change was introduced to eliminate the noise in the measurements due to the ca. 300-millisecond latency of the observer to press the button, a first step to achieve a greater precision, though requiring several hours of coding.

In Golinkoff et al. (1987)'s paradigm, word recognition was operationalized as a tendency to look longer at the target picture vs. the distractor, averaged over a 6 s measurement window following the offset of the linguistic stimulus. More recent psycholinguistic research with adults showed that experienced listeners can process language incrementally, generating hypotheses about the meaning on the basis of what they heard up to that moment [63]. This led Fernald's group to assume that children are simply considerably slower than adults and that a 6 s time window after the offset of the speech stimulus was necessary to give infants the time to process the language they were hearing.

In the early studies investigating the influence of prosodic features on their ability to recognize familiar words, using this modified preferential-looking method, Fernald et al. (1992) [63] obtained counterintuitive results. When using a percent-correct measure averaged over a 6 s measurement window, English-learning 24-month-old children apparently performed less well than the 18-month-old children. When reduced to a 2 s measurement window, the predicted improvement in word recognition became clear, being around 60% for 18-month-old children and 80% for 24-month-old children. Adopting the 6 s window, which was a standard for the IPLP, accuracy was greatly underestimated. Indeed, 24-month-old children had oriented quickly to the target picture upon hearing the speech stimulus, had looked at it for 2 to 3 s, but then tended to look at the other picture or to look away, as they were losing interest. This behavior has to be interpreted as a sign of rapid processing, since look-backs and look aways followed a correct response in most cases. Thus, this post-response "noise", when averaged into the percent-looking-to-target over a 6 s window, made 24-month-old children appear less accurate than 18-month-old children [62].

Two important procedural changes were made in the looking-while-listening procedure. First, they started to measure eye movements from the onset of the target word and not from the offset. Second, they coded eye movements at the possible finest level of resolution instead of coding eye movements based on average looking time over an arbitrary time window (which can be suitable for one age but not another). Thus, Fernald et al. (1998) were able to code eye movements with a 100-millisecond resolution [64]. In subsequent studies, resolution increased to 33 ms, the duration of a single video frame, enabling a more precise measure of reaction times, being able to capture child's latency to shift from one to the other picture [65]. Thanks to this improvement, the looking-while-listening procedure has become an increasingly powerful method for monitoring real-time language processing, enabling to measure both accuracy and reaction time in word recognition.

Procedure and Limits

The looking-while-listening procedure is superficially similar to the preferential-looking procedure in that infants are presented with two pictures on each trial and hear a linguistic stimulus naming one of them, as gaze patterns are recorded and manually coded frame by frame. What is new in this procedure is that the interest is not on a single preference score based on total looking at the target object over a time window, but rather on the time course of looking to the referent as the sentence unfolds. As Fernald and colleagues put it, "the static notion of 'preference' is irrelevant for our purposes. Rather than construing infants' looking behavior in response to spoken language as motivated by *preference*, we are interested in how children establish *reference* by making sense of spoken language from moment to moment" [62] (p. 190).

Experiments using this paradigm take place in a testing room with dimmed lights. The standard procedure consists of presenting pairs of images horizontally on two different screens. The target and distractor pictures are shown for 2 s prior to the onset of the speech stimulus, as shown in Figure 6B. A trial was thus divided into a pre-naming and a post-naming phase. Trials lasted 6–8 s on average. The entire experiment lasted about 5–6 min.

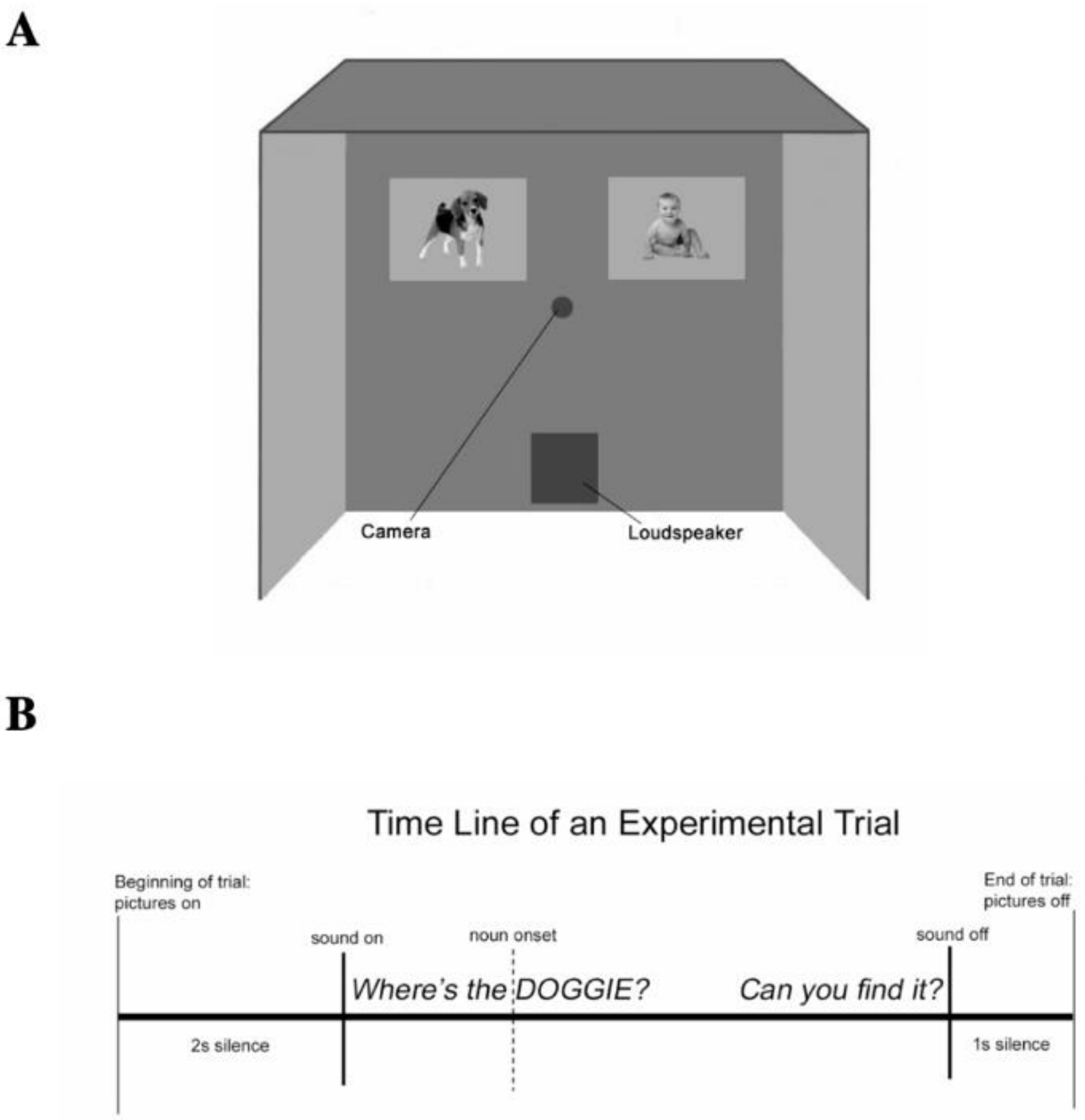

Figure 6. Configuration of test booth with rear-projection screen used in the looking-while-listening procedure (**A**). Schematic timeline for a typical trial (**B**) reprinted with permission from [62]. Copyright 2008 Benjamins.

In their comprehensive review of this paradigm, Fernald and colleagues (2008) listed a series of factors that need to be controlled when developing a study using this paradigm. First, both images need to be matched for size and salience. However, as demonstrated by Arias-Trejo and Plunkett (2010) [66], choosing a distracter image which is perceptually close to the target image (e.g., a balloon paired with an egg) can result in uninvited interference effects so that 18- to 24-month-old children failed to identify the target image. A second recommendation is that across all participants both objects in a given trial should be used as target and as distractor, in order to avoid any preference for one stimulus over another. Although desirable, such a control is not always possible given the restricted choice of items in young children and the need for a sufficient number of trials per participant. In the literature, experiments using the looking-while-listening procedure have controlled for this possible preference effect [65–67], presenting the same visual stimuli at least twice, while others have not [68–70] but still found comparable results.

As said, the looking-while-listening methodology differs critically from the preferential-looking paradigm in terms of the quantitative methods used for data reduction and analysis, yielding measures of speech processing with higher resolution. However, it is only with the advent of eye-tracking technology that also the looking-while-listening paradigm has been abandoned (see [71] for a detailed overview). Indeed, eye-tracking technology brought out three noteworthy limits of the procedure described above. First, the looking-while-listening procedure typically uses visual display with only two alternatives, rather than more complex scenes involving three or more displays, which are possible with eye-tracking technology. Second, no automated eye tracker is used, and eye movements are manually coded, thus resulting in lower accuracy and reliability of data analysis. Lastly, this procedure, as well as the preferential-looking paradigm, was developed for experiments with infants and young children. This does not allow for a comparison with an adult control group, which would establish a target baseline in speech processing.

3.3. The Eye-Tracking Technology

The eye tracker is a device using high-definition cameras in combination with near infrared lights. For the purposes of research in the linguistic field, it records the position of the pupil and the corneal reflection, i.e., the reflection of a point light projected by the near infrared light onto the participant's eyeball (e.g., [72], see Figure 7). As the participant's eye moves, the pupil moves, but the corneal reflection remains constant. Measuring the distance between these two points makes it possible to calculate the position of the eyeball within the head [73].

The camera records eye movements in real time either mounted on a headband and placed near the eye (head-mounted eye tracker), remotely from a desktop camera, or embedded within a computer screen (screen-based eye tracker).

Screen-based eye tracking involves presenting stimuli on a computer screen while measuring where the participant is looking. Within screen-based eye tracking, two approaches are widely used, i.e., head fixed and head free. The head-fixed modality requires the participant to use a chin-rest during tracking to keep a constant distance between the eye and the screen. Although this technique offers the best temporal and spatial resolution, it is not widely used in developmental research, due to young children's troubles in maintaining stillness on a chin-rest.

In contrast, head-free eye trackers allow participants to freely move their head in 3D space during the recordings. By setting this remote modality, the eye tracker tracks the position of the head during the experimental session including information about the distance of the head from the screen while calculating the location of the pupil and the corneal reflection. To do so, a target sticker is usually placed in the middle of the child's forehead, allowing the eye tracker to have a third reference during calculations.

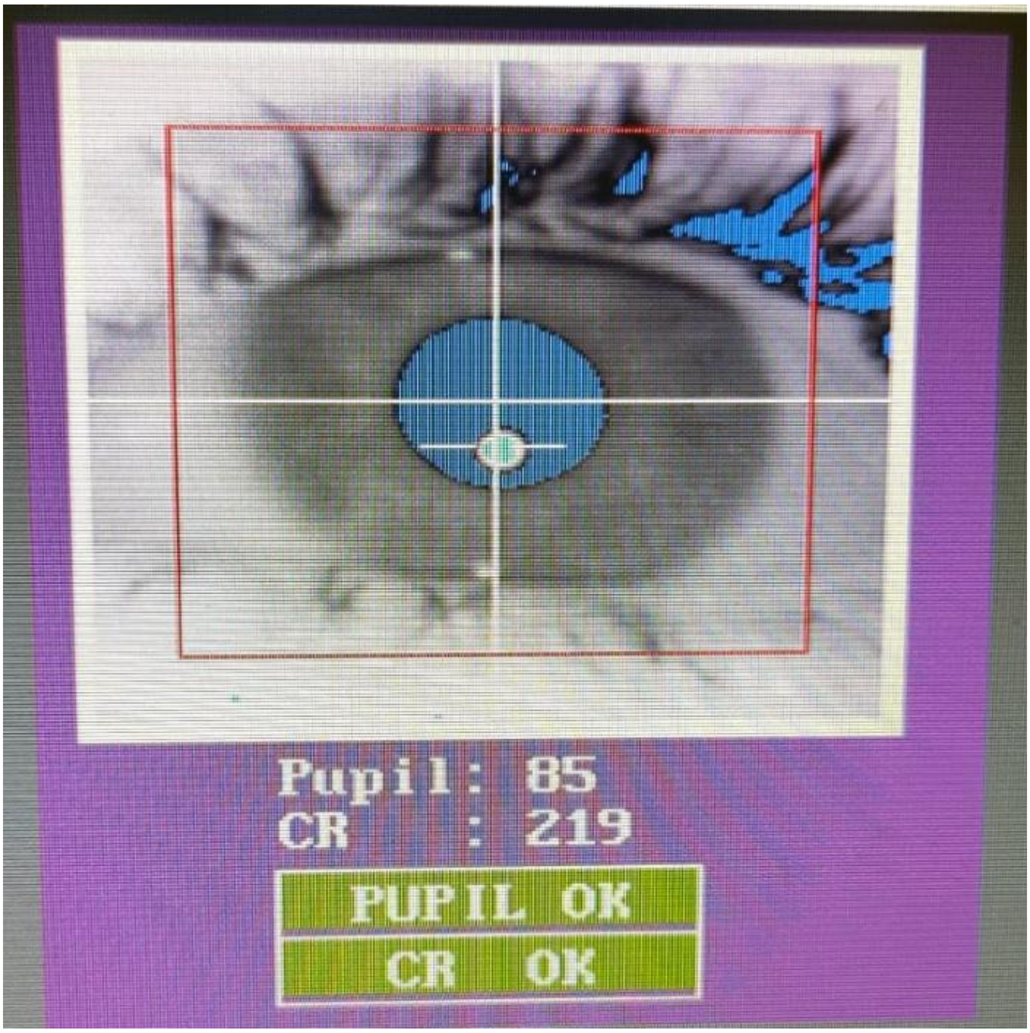

Figure 7. Example of the user view from EyeLink 1000 Plus eye-tracking system. The pupil is highlighted in blue. The little light-blue dot underneath is the corneal reflection, projected from the eye tracker into the participant's eye.

Whenever we perform everyday tasks concerning vision, shifts of attention are accompanied by shifts in gaze. These shifts are accomplished by ballistic moments, called saccades, which bring the attended region into the central area of the fovea where visual acuity is greatest, resulting in fixations [74]. The eye tracker output provides data about both saccades and fixations. Figure 8 shows an example of an experimental setting during an eye-tracking experiment with children.

All eye-tracking experiments start with a calibration phase to establish a mapping of screen coordinates and measurements. This is achieved by asking the participant to fixate a sequence of calibration points. In each of these points, the position of the pupil is recorded, thus that the eye tracker "learns" that when a participant is looking at, say, the middle of the screen ("point X"), the distance between the pupil and the corneal reflection is "vector Y". This allows the system to pair information about the relative position of the pupil and corneal reflections in the eye with specific spatial locations. If, during the experiment, the distance between the two points is "vector Y", it can be concluded that the participants were looking at "point X". At any point after the calibration phase, it is thus possible to estimate where the participant is looking within a given scene, and gaze position is available in screen coordinates. Usually, the calibration phase is followed by a validation run, aimed at determining whether the estimated eye position is close to the known coordinates. In addition, it is a recommended common practice, especially with young children, to present additional validation points in between the trials (the so-called drift corrections) and to recalibrate in case of failure.

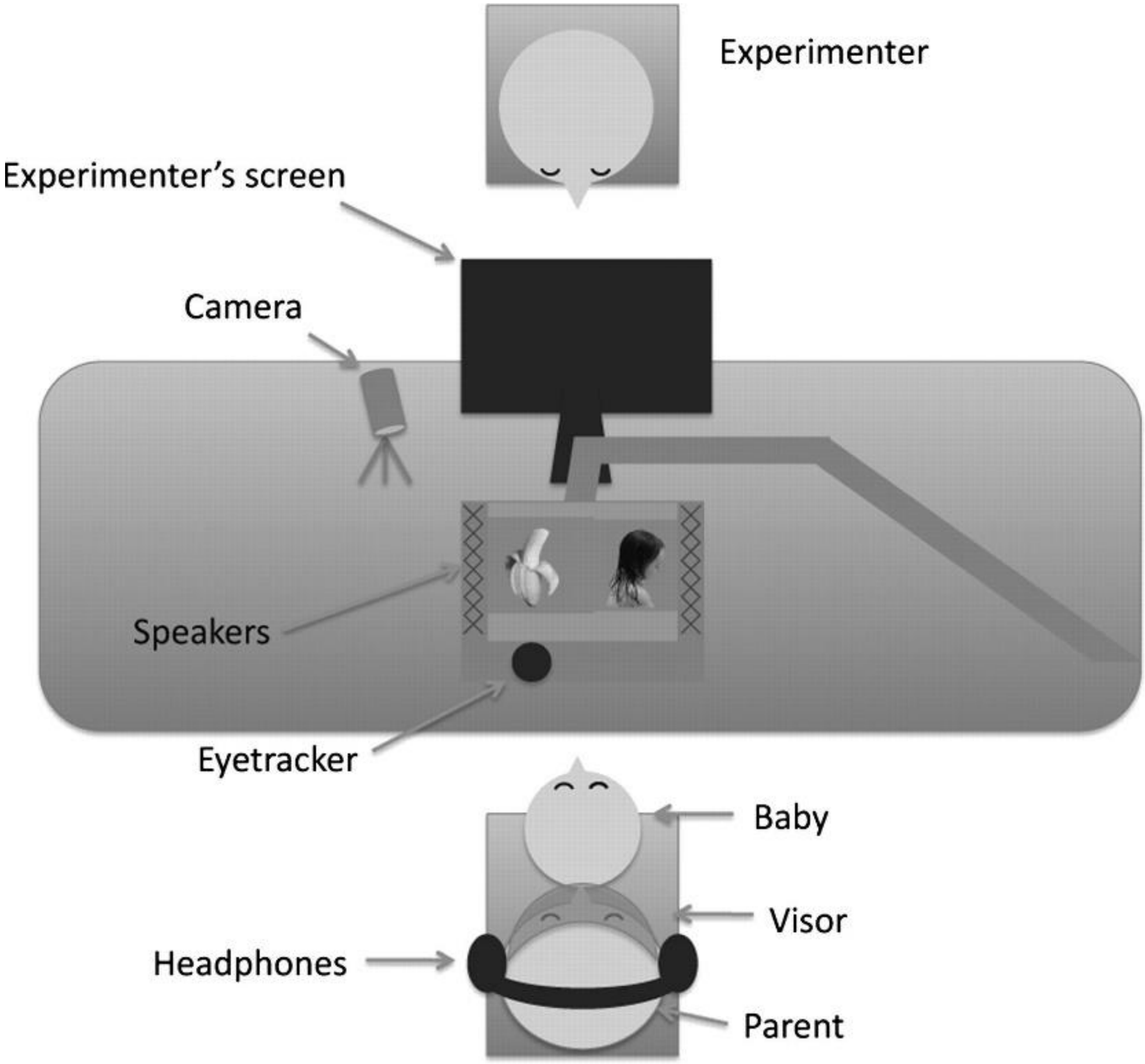

Figure 8. Example of the experimental setup used in Bergelson and Swingley (2012) (Adapted from [75]). The child sat on her parent's lap and was presented with images on a screen and sounds from a computer equipped with an eye tracker and speakers. The researcher sat behind a screen and was not visible to the child. The experimenter controlled the presentation of stimuli and monitored the child on a live-feed camera.

To assess what the participant was looking at throughout the trial, the experimenter defines interest areas (or regions of interest) in the visual world, each associated with one of more objects. An automated coding procedure then scores each fixation as directed at one area of interest or not directed at any. As for saccades, since they are triggered by a shift in visual attention to a new location, that location can be considered the locus of attention during a saccade [76].

3.3.1. The Analysis of Eye-Tracking Data

The eye tracker records eye movement data continuously from the very beginning of the speech stimulus, although it is sometimes useful to start recording even before the linguistic stimulus starts, in order to be able to determine whether there are any biases in gaze prior to stimulus presentation. The eye-tracking system automatically generates an output data file, logging the precise timing and location of the relevant events.

Regardless of all the measures that can possibly be reported and analyzed (e.g., total number of fixations to the target object or latencies for first fixations to the target object, see [76] for a detailed description of all measures), it is usually useful to acquire a temporally detailed, qualitative picture of the continuous distribution of eye movements over the objects in the visual display as the speech stimulus unfolds [24]. These are generally plotted by identifying very fine time bins (e.g., 50 or 100 ms time "slices" containing averaged eye movement data) and, for each bin, by computing an average across all subjects and items separately for each area of interest in the screen. In this way, it is possible to acquire the proportion of time in that bin spent fixating each object in the visual array. A proportion-of-fixations plot represents, at each moment in time, the proportion of trials with a look to each picture, averaged across participants (or items). Proportion-of-fixation plots usually present data aligned to a relevant linguistic event, which typically requires a further alignment across trials. Moreover, in evaluating the data, it is important to take into account that information in the speech signal influences eye movements with a delay of approximately 200–250 ms [77]. Figure 9 shows an example of the fixation proportions computed for the four interest areas.

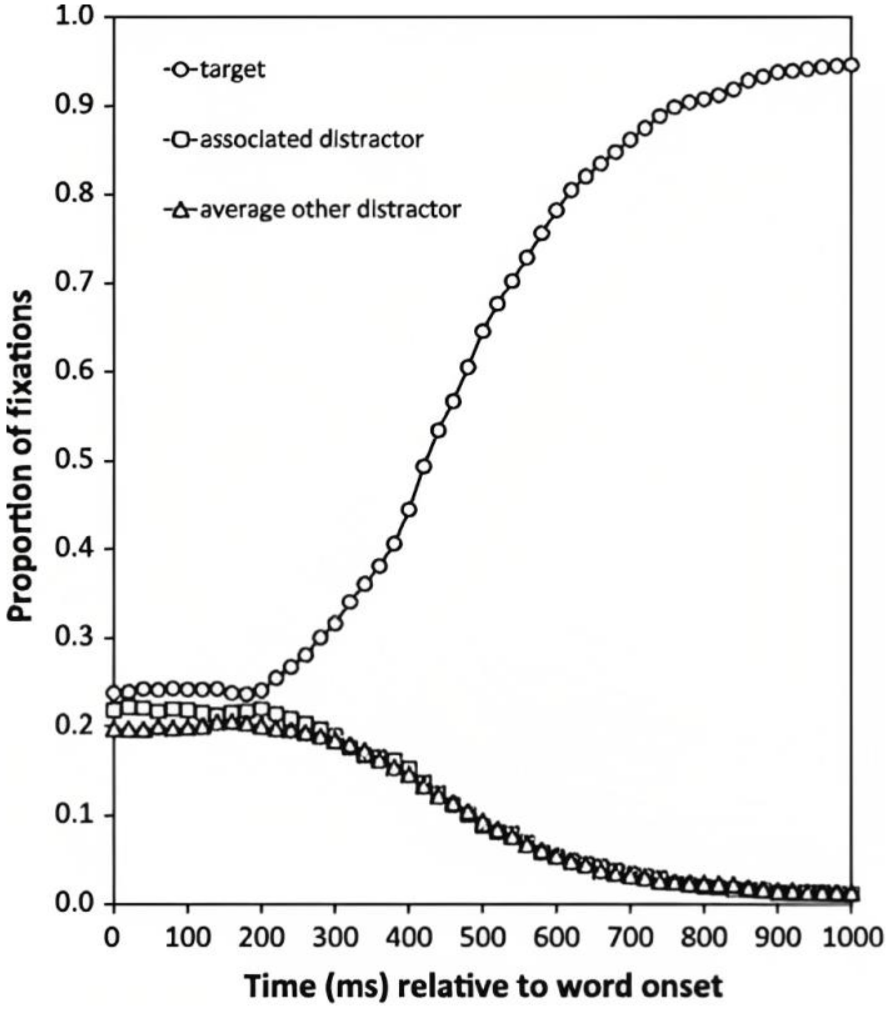

Figure 9. Proportion of fixations to the target, the associated distractor, and the averaged other distractors reprinted with permission from [76]. Copyright 2017 John Wiley and Sons. In this experiment, the participant saw a display with a target picture and three distractors and followed a simple spoken instruction to click on the target.

Presenting data in this qualitative manner often brings to light relevant patterns in the data that may be missed by the mere observation of measures such as the total number of fixations. The resulting graphs allow for a detailed picture of the participant's shifts of attention as the speech unfolded in time.

Visual word eye movement data can be analyzed with a range of statistical analyses. Rather than reporting separate analyses for each time bin, which could result in an increased potential for spuriously significant results, researchers often choose to create time-windows over which the eye movement data are averaged and then submitted for statistical analysis. Traditionally, separate analyses of variance were conducted for each bin or time-window (i.e., ANOVAs), using subject and items as random factors, with uncorrelated independent variables. Only measures on the target were specified as dependent variables. However, during the last decade, linear mixed models [78] reduced the number of analyses to be run, making it possible to specify subjects and items as crossed random effects in one single analysis. These models returned considerable responsibility to the data analyst, concerning hypotheses and fixed and random effects specification, as well as within-subject and within-items effects [79]. In addition, linear mixed models handled the problem of missing-at-random eye-movement data records and broke down the distinction between experimental and "correlational" analysis, allowing for the analysis of interactions [77]. Nonetheless, for a deep understanding, the implementation of formal computational models would be needed for a more refined linguistic analysis of eye-movement data, to avoid being misled by the researcher's intuition and schematics. The development of such models is the next step to take in this field [80,81].

In conclusion, it is worth noticing that all these advances in statistical inference are possible because eye-tracking yields a very high density of behavioral observations. The brief description of this technology, if compared with video cameras used for the tasks described in Sections 3.1 and 3.2, makes it clear that eye tracking provides temporal and spatial sensitivity to a much higher level of precision.

3.3.2. Limitations and Conclusions

Over the last decade, this technology has been proven to be the most suitable methodology to investigate the relationship between language and vision, allowing for theories of representation to be integrated with theories of processing and providing a clearer picture of how language is processed in real time.

This methodology made it possible to explore an exceedingly large number of topics from a new perspective. The entry mentions, among many others, studies with adult participants, such as, for instance, the research conducted on predictive processing (e.g., [82–85]) and on negation processing (e.g., [23,86–88]). A lot of innovative research has been conducted with children, such as studies on incremental processing (e.g., [13,89]), on early lexical development (e.g., [71,90,91]), and on predictive processing (e.g., [92,93]). Finally, eye-tracking technology has also allowed for research with atypical populations, which are more difficult to test with traditional offline methodologies. Recent studies have been conducted with bilingual SLI children (e.g., [94]), with aphasic patients (e.g., [95,96]), and with participants diagnosed with developmental dyslexia (e.g., [23,97,98]).

Nonetheless, the interpretation of eye movements data is hardly straightforward due to some intrinsic limitations of the visual world paradigm. The interpretation of fixation patterns and attentional shifts occurring during online language processing must take into account the possibility that language-mediated eye movements might reflect the lexical activation process occurring as the linguistic input unfolds. Hearing a word automatically elicits the mental activation of its semantic and perceptual features, and the listener's overt visual attention is drawn towards those objects in the visual scene which share some of these features with the mentioned word.

A related issue concerns the pre-activation of word candidates. As a result of a prolonged preview of the visual scene before the onset of the spoken input, participants might be faster in directing their gaze towards those visual objects sharing semantic, perceptual, and also phonological features with the mentioned word (e.g., [83,99,100]). In addition, the concurrent presence of different pictures in the visual scene might encourage the listeners to make inferences that would not normally be drawn during natural language processing. In a visual world set-up, listeners might tend to look at the corresponding picture only because they have it at their disposal, and this might bias the comprehension process.

These challenges in results interpretation are related to another key feature of the visual world paradigm, i.e., the absence of metalinguistic feedback on language comprehension. This makes, in fact, task execution effortless and extremely suitable for children and impaired populations, but it does not provide any hints on whether participants have reached a final understanding, which is not that trivial when it comes to children and atypical subjects.

A further concern is related to the domain of applicability of the visual world paradigm, namely the fact that it can hardly be used for the study of language that is not about concrete co-present referents. Consequently, one might argue that results from visual world studies cannot be generalized to less constrained situations. To our knowledge, however, there is no evidence that this is the case. Rather, conclusions drawn from visual world studies seem to scale up also to language that is not about a restricted visual context (see [74] for a full discussion).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meyer, D.E.; Schvaneveldt, R.W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *J. Exp. Psychol.* **1971**, *90*, 227–234. [CrossRef] [PubMed]
2. Wason, P.C. Response to Affirmative and Negative Binary Statements. *Br. J. Psychol.* **1961**, *52*, 133–142. [CrossRef] [PubMed]
3. Carpenter, P.A.; Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychol. Rev.* **1975**, *82*, 45–73. [CrossRef]
4. Dickey, M.W.; Choy, J.J.; Thompson, C.K. Real-time comprehension of wh- movement in aphasia: Evidence from eyetracking while listening. *Brain Lang.* **2007**, *100*, 1–22. [CrossRef] [PubMed]
5. Yee, E.; Blumstein, S.E.; Sedivy, J.C. Lexical-Semantic Activation in Broca's and Wernicke's Aphasia: Evidence from Eye Movements. *J. Cogn. Neurosci.* **2008**, *20*, 592–612. [CrossRef]
6. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **1998**, *124*, 372–422. [CrossRef] [PubMed]
7. De Luca, M.; Di Pace, E.; Judica, A.; Spinelli, D.; Zoccolotti, P. Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia* **1999**, *37*, 1407–1420. [CrossRef]
8. Desroches, A.S.; Joanisse, M.F.; Robertson, E.K. Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition* **2006**, *100*, B32–B42. [CrossRef]
9. Huettig, F.; Brouwer, S. Delayed Anticipatory Spoken Language Processing in Adults with Dyslexia—Evidence from Eye-tracking. *Dyslexia* **2015**, *21*, 97–122. [CrossRef]
10. Benfatto, M.N.; Seimyr, G.Ö.; Ygge, J.; Pansell, T.; Rydberg, A.; Jacobson, C. Screening for Dyslexia Using Eye Tracking during Reading. *PLoS ONE* **2016**, *11*, e0165508. [CrossRef]
11. Joseph HS, S.L.; Nation, K.; Liversedge, S.P. Using Eye Movements to Investigate Word Frequency Effects in Children's Sentence Reading. *Sch. Psychol. Rev.* **2013**, *42*, 207–222. [CrossRef]
12. Mani, N.; Huettig, F. Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *J. Exp. Child Psychol.* **2014**, *126*, 264–279. [CrossRef] [PubMed]
13. Tribushinina, E.; Mak, W.M. Three-year-olds can predict a noun based on an attributive adjective: Evidence from eye-tracking. *J. Child Lang.* **2016**, *43*, 425–441. [CrossRef] [PubMed]
14. Yarbus, A.L. *Eye Movements and Vision*; Plenum Press: New York, NY, USA, 1967.
15. Cooper, R.M. 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. *Cogn. Psychol.* **1974**, *6*, 84–107. [CrossRef]

16. Mackworth, N.H. The wide-angle reflection eye camera for visual choice and pupil size. *Percept. Psychophys.* **1968**, *3*, 32–34. [CrossRef]
17. Just, M.A.; Carpenter, P.A. A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* **1980**, *87*, 329–354. [CrossRef] [PubMed]
18. Allopenna, P.D.; Magnuson, J.S.; Tanenhaus, M.K. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *J. Mem. Lang.* **1998**, *38*, 419–439. [CrossRef]
19. Tanenhaus, M.K.; Spivey, M.; Eberhard, K.; Sedivy, J. Integration of visual and linguistic information in spoken language comprehension. *Science* **1995**, *268*, 1632–1634. [CrossRef]
20. Frazier, L.; Fodor, J.D. The sausage machine: A new two-stage parsing model. *Cognition* **1978**, *6*, 291–325. [CrossRef]
21. De Vincenzi, M. *Syntactic Parsing Strategies in Italian*; Kluwer: Dordrecht, The Netherlands, 1991.
22. De Vincenzi, M.; Job, R. An investigation of Late Closure: The role of syntax, thematic structure and pragmatics in initial and final interpretation. *J. Exp. Psychol. Learn. Mem. Cogn.* **1995**, *21*, 1303–1321. [CrossRef]
23. Tagliani, M. On Vision and Language Interaction in Negation Processing: The Real-Time Interpretation of Sentential Negation in Typically Developed and Dyslexic Adults. Ph.D. Dissertation, University of Verona, Verona, Italy, University of Göttingen, Göttingen, Germany, 2021.
24. Sedivy, J. Chapter 6. Using eyetracking in language acquisition research. In *Experimental Methods in Language Acquisition Research*; Blom, E., Unsworth, S., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2010; pp. 115–138. [CrossRef]
25. Altmann GT, M.; Kamide, Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* **1999**, *73*, 247–264. [CrossRef]
26. Nadig, A.S.; Sedivy, J.C. Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution. *Psychol. Sci.* **2002**, *13*, 329–336. [CrossRef]
27. Sedivy, J.C.; KTanenhaus, M.; Chambers, C.G.; Carlson, G.N. Achieving incremental semantic interpretation through contextual representation. *Cognition* **1999**, *71*, 109–147. [CrossRef] [PubMed]
28. Werker, J.F.; Tees, R.C. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* **1984**, *7*, 49–63. [CrossRef]
29. Saffran, J.R. Constraints on Statistical Language Learning. *J. Mem. Lang.* **2002**, *47*, 172–196. [CrossRef]
30. Hallé, P.A.; de Boysson-Bardies, B. Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behav. Dev.* **1994**, *17*, 119–129. [CrossRef]
31. Eimas, P.D.; Siqueland, E.R.; Jusczyk, P.; Vigorito, J. Speech perception in infants. *Science* **1971**, *171*, 303–306. [CrossRef] [PubMed]
32. Tsushima, T.; Takizawa, O.; Sasaki, M.; Shiraki, S.; Nishi, K.; Kohno, M.; Menyuk, P.; Best, C. Discrimination of English /r-l/and /w-y/by Japanese infants at 6-12 Months: Language specific developmental changes in speech perception abilities. In Proceedings of the International Conference of Spoken Language Processing Acoustical Society of Japan, Yokohama, Japan, 18–22 September 1994; pp. 1695–1698.
33. Bloom, L. One Word at a Time: The Use of Single Word Utterances before Syntax. In *One Word at a Time*; De Gruyter Mouton: Berlin, Germany, 2013. [CrossRef]
34. Fenson, L.; Marchman, V.A.; Thal, D.J.; Dale, P.S.; Reznick, J.S.; Bates, E. *MacArthur-Bates Communicative Development Inventories*, 2nd ed.; APA PsycTests: Washington, DC, USA, 2006; Available online: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft11538-000 (accessed on 22 December 2022).
35. Benedict, H. Early lexical development: Comprehension and production. *J. Child Lang.* **1979**, *6*, 183–200. [CrossRef]
36. Markman, E.M. *Categorization and Naming in Children: Problems of Induction*; MIT Press: Cambridge, MA, USA, 1989.
37. Golinkoff, R.M.; Ma, W.; Song, L.; Hirsh-Pasek, K. Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspect. Psychol. Sci.* **2013**, *8*, 316–339. [CrossRef]
38. Fantz, R.L. Pattern Vision in Newborn Infants. *Science* **1963**, *140*, 296–297. [CrossRef] [PubMed]
39. Fantz, R.L. Pattern Vision in Young Infants. *Psychol. Rec.* **1958**, *8*, 43. Available online: https://www.proquest.com/docview/1301204249/citation/9A76F464DBD14937PQ/1 (accessed on 22 December 2022). [CrossRef]
40. Fantz, R.L. Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science* **1964**, *146*, 668–670. [CrossRef] [PubMed]
41. Spelke, E. Infants' intermodal perception of events. *Cogn. Psychol.* **1976**, *8*, 553–560. [CrossRef]
42. Spelke, E.S. Perceiving bimodally specified events in infancy. *Dev. Psychol.* **1979**, *15*, 626–636. [CrossRef]
43. Thomas, D.G.; Campos, J.J.; Shucard, D.W.; Ramsay, D.S.; Shucard, J. Semantic Comprehension in Infancy: A Signal Detection Analysis. *Child Dev.* **1981**, *52*, 798–803. [CrossRef]
44. Ambridge, B.; Rowland, C.F. Experimental methods in studying child language acquisition. *WIREs Cogn. Sci.* **2013**, *4*, 149–168. [CrossRef]
45. Naigles, L. Children use syntax to learn verb meanings. *J. Child Lang.* **1990**, *17*, 357–374. [CrossRef]
46. Hoff, E. *Research Methods in Child Language: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
47. Golinkoff, R.M.; Hirsh-Pasek, K.; Mervis, C.B.; Frawley, W.B.; Parillo, M. Lexical principles can be extended to the acquisition of verbs. In *Beyond Names for Things: Young Children's Acquisition of Verbs*; Psychology Press: London, UK, 1995; pp. 185–222.
48. Reznick, J.S. Visual preference as a test of infant word comprehension. *Appl. Psycholinguist.* **1990**, *11*, 145–166. [CrossRef]

49. Golinkoff, R.M.; Hirsh-Pasek, K.; Cauley, K.M.; Gordon, L. The eyes have it: Lexical and syntactic comprehension in a new paradigm. *J. Child Lang.* **1987**, *14*, 23–45. [CrossRef]
50. Hirsh-Pasek, K.; Golinkoff, R.M. The intermodal preferential looking paradigm: A window onto emerging language comprehension. In *Methods for Assessing Children's Syntax*; The MIT Press: Cambridge, MA, USA, 1996; pp. 105–124.
51. Naigles, L.G.; Kako, E.T. First Contact in Verb Acquisition: Defining a Role for Syntax. *Child Dev.* **1993**, *64*, 1665–1687. [CrossRef]
52. Naigles, L.G.; Gelman, S.A. Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *J. Child Lang.* **1995**, *22*, 19–46. [CrossRef]
53. Golinkoff, R.M.; Hirsh-Pasek, K.; Bailey, L.M.; Wenger, N.R. Young children and adults use lexical principles to learn new nouns. *Dev. Psychol.* **1992**, *28*, 99–108. [CrossRef]
54. Gleitman, L. The structural sources of verb meaning. *Lang. Acquis.* **1990**, *1*, 3–55. [CrossRef]
55. Gleitman, L.R.; Cassidy, K.; Nappa, R.; Papafragou, A.; Trueswell, J.C. Hard words. *Lang. Learn. Dev.* **2005**, *1*, 23–64. [CrossRef]
56. Landau, B.; Gleitman, L.R. *Language and Experience*; Harvard University Press: Cambridge, MA, USA, 1985.
57. Lidz, J.; Waxman, S.; Freedman, J. What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition* **2003**, *89*, 295–303. [CrossRef] [PubMed]
58. Halberda, J.; Sires, S.F.; Feigenson, L. Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychol. Sci.* **2006**, *17*, 572–576. [CrossRef]
59. Markman, E.M.; Wachtel, G.F. Children's use of mutual exclusivity to constrain the meanings of words. *Cogn. Psychol.* **1988**, *20*, 121–157. [CrossRef]
60. Hollich, G.J.; Hirsh-Pasek, K.; Golinkoff, R.M.; Brand, R.J.; Brown, E.; Chung, H.L.; Hennon, E.; Rocroi, C.; Bloom, L. Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monogr. Soc. Res. Child Dev.* **2000**, *65*, 1–123. [PubMed]
61. Bates, E. Comprehension and production in early language development: Comments on Savage-Rumbaugh et al. *Monogr. Soc. Res. Child Dev.* **1993**, *58*, 222–242. [CrossRef]
62. Fernald, A.; Zangl, R.; Portillo, A.L.; Marchman, V.A. Looking while listening: Using eye movements to monitor spoken language. In *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*; John Benjamins Publishing: Amsterdam, The Netherlands, 2008; pp. 97–134.
63. Fernald, A.; McRoberts, G.; Herrera, C. The Role of prosodic features in early word recognition. In Proceedings of the 8th International Conference on Infant Studies, Miami, FL, USA, 7–10 May 1992.
64. Fernald, A.; Pinto, J.P.; Swingley, D.; Weinberg, A.; McRoberts, G.W. Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year. *Psychol. Sci.* **1998**, *9*, 228–231. [CrossRef]
65. Fernald, A.; Thorpe, K.; Marchman, V.A. Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cogn. Psychol.* **2010**, *60*, 190–217. [CrossRef]
66. Arias-Trejo, N.; Plunkett, K. The effects of perceptual similarity and category membership on early word-referent identification. *J. Exp. Child Psychol.* **2010**, *105*, 63–80. [CrossRef] [PubMed]
67. Mani, N.; Plunkett, K. Phonological specificity of vowels and consonants in early lexical representations. *J. Mem. Lang.* **2007**, *57*, 252–272. [CrossRef]
68. Swingley, D.; Aslin, R.N. Spoken word recognition and lexical representation in very young children. *Cognition* **2000**, *76*, 147–166. [CrossRef] [PubMed]
69. Swingley, D.; Aslin, R.N. Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychol. Sci.* **2002**, *13*, 480–484. [CrossRef] [PubMed]
70. Durrant, S.; Luche, C.D.; Cattani, A.; Floccia, C. Monodialectal and multidialectal infants' representation of familiar words. *J. Child Lang.* **2015**, *42*, 447–465. [CrossRef] [PubMed]
71. Redolfi, M. How children acquire adjectives: Evidence from three eye-tracking studies on Italian. Doctoral Dissertation, University of Verona, Verona, Italy, University of Konstanz, Konstanz, Germany, 2022.
72. Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; OUP Oxford: Oxford, UK, 2011.
73. Wass, S.V. The use of eye-tracking with infants and children. In *Practical Research with Children*; Routledge: London, UK, 2016; pp. 50–71. [CrossRef]
74. Tanenhaus, M.K.; Brown-Schmidt, S. Language processing in the natural world. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 1105–1122. [CrossRef]
75. Bergelson, E.; Swingley, D. At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3253–3258. [CrossRef] [PubMed]
76. Salverda, A.P.; Tanenhaus, M.K. The Visual World Paradigm. In *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
77. Salverda, A.P.; Kleinschmidt, D.; Tanenhaus, M.K. Immediate effects of anticipatory coarticulation in spoken-word recognition. *J. Mem. Lang.* **2014**, *71*, 145–163. [CrossRef]
78. Baayen, H.; Vasishth, S.; Kliegl, R.; Bates, D. The cave of shadows: Addressing the human factor with generalized additive mixed models. *J. Mem. Lang.* **2017**, *94*, 206–234. [CrossRef]

79. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
80. Magnuson, J.S. Fixations in the visual world paradigm: Where, when, why? *J. Cult. Cogn. Sci.* **2019**, *3*, 113–139. [CrossRef]
81. Venhuizen, N.J.; Crocker, M.W.; Brouwer, H. Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Process.* **2019**, *56*, 229–255. [CrossRef]
82. Dahan, D.; Tanenhaus, M.K. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychon. Bull. Rev.* **2005**, *12*, 453–459. [CrossRef]
83. Huettig, F.; Rommers, J.; Meyer, A.S. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychol.* **2011**, *137*, 151–171. [CrossRef]
84. Huizeling, E.; Alday, P.M.; Peeters, D.; Hagoort, P. Combining EEG and eye-tracking to investigate the prediction of upcoming speech in naturalistic virtual environments: A 3D visual world paradigm. In Proceedings of the 18th NVP Winter Conference on Brain and Cognition, Egmond aan Zee, The Netherlands, 28–30 April 2022.
85. Huettig, F.; Guerra, E. Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Res.* **2019**, *1706*, 196–208. [CrossRef]
86. Orenes, I.; Beltrán, D.; Santamaría, C. How negation is understood: Evidence from the visual world paradigm. *J. Mem. Lang.* **2014**, *74*, 36–45. [CrossRef]
87. Orenes, I.; Moxey, L.; Scheepers, C.; Santamaría, C. Negation in context: Evidence from the visual world paradigm. *Q. J. Exp. Psychol.* **2016**, *69*, 1082–1092. [CrossRef]
88. Orenes, I.; García-Madruga, J.A.; Espino, O.; Byrne, R.M. The Comprehension of Counterfactual Conditionals: Evidence from Eye-Tracking in the Visual World Paradigm. *Front. Psychol.* **2019**, *10*, 1172. [CrossRef]
89. Özge, D.; Küntay, A.; Snedeker, J. Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition* **2019**, *183*, 152–180. [CrossRef]
90. Bergelson, E.; Aslin, R. Semantic specificity in one-year-olds' word comprehension. *Lang. Learn. Dev.* **2017**, *13*, 481–501. [CrossRef]
91. Bergelson, E.; Swingley, D. Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Dev.* **2018**, *89*, 1567–1576. [CrossRef]
92. Mani, N.; Huettig, F. Prediction during language processing is a piece of cake—But only for skilled producers. *J. Exp. Psychol. Hum. Percept. Perform.* **2012**, *38*, 843. [CrossRef] [PubMed]
93. Borovsky, A.; Elman, J.L.; Fernald, A. Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *J. Exp. Child Psychol.* **2012**, *112*, 417–436. [CrossRef]
94. Mak, W.M.; Tribushinina, E.; Lomako, J.; Gagarina, N.; Abrosova, E.; Sanders, T. Connective processing by bilingual children and monolinguals with specific language impairment: Distinct profiles. *J. Child Lang.* **2017**, *44*, 329–345. [CrossRef]
95. Thompson, C.K.; Choy, J.J. Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *J. Psycholinguist. Res.* **2009**, *38*, 255–283. [CrossRef]
96. Sharma, S.; Kim, H.; Harris, H.; Haberstroh, A.; Wright, H.H.; Rothermich, K. Eye tracking measures for studying language comprehension deficits in aphasia: A systematic search and scoping review. *J. Speech Lang. Hear. Res.* **2021**, *64*, 1008–1022. [CrossRef]
97. Franzen, L.; Stark, Z.; Johnson, A.P. Individuals with dyslexia use a different visual sampling strategy to read text. *Sci. Rep.* **2021**, *11*, 6449. [CrossRef]
98. Robertson, E.K.; Gallant, J.E. Eye tracking reveals subtle spoken sentence comprehension problems in children with dyslexia. *Lingua* **2019**, *228*, 102708. [CrossRef]
99. Huettig, F.; McQueen, J.M. The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* **2007**, *57*, 460–482. [CrossRef]
100. de Groot, F.; Huettig, F.; Olivers, C.N.L. When meaning matters: The temporal dynamics of semantic influences on visual attention. *J. Exp. Psychol. Hum. Percept. Perform.* **2016**, *42*, 180–196. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.