

## Video Article

# Eye Tracking During Visually Situated Language Comprehension: Flexibility and Limitations in Uncovering Visual Context Effects

Camilo Rodríguez Ronderos<sup>1</sup>, Katja Münster<sup>1</sup>, Ernesto Guerra<sup>2</sup>, Helene Kreysa<sup>3</sup>, Alba Rodríguez<sup>1,4</sup>, Julia Kröger<sup>4</sup>, Thomas Kluth<sup>4</sup>, Michele Burigo<sup>4</sup>, Dato Abashidze<sup>5</sup>, Eva Nunnemann<sup>4</sup>, Pia Knoeferle<sup>1,6,7</sup>

<sup>1</sup>Institute of German Studies and Linguistics, Humboldt-Universität zu Berlin

<sup>2</sup>Center for Advanced Research in Education – CIAE, Universidad de Chile

<sup>3</sup>Institute of Psychology, Friedrich Schiller University Jena

<sup>4</sup>Cognitive Interaction Technology Excellence Cluster (CITEC), Faculty of Linguistics and Literature, Bielefeld University

<sup>5</sup>Department of Education, Concordia University

<sup>6</sup>Berlin School of Mind and Brain

<sup>7</sup>Einstein Center for Neurosciences Berlin

Correspondence to: Pia Knoeferle at [pia.knoeferle@hu-berlin.de](mailto:pia.knoeferle@hu-berlin.de)

URL: <https://www.jove.com/video/57694>

DOI: [doi:10.3791/57694](https://doi.org/10.3791/57694)

Keywords: Behavior, Issue 141, Visually situated language comprehension, eye tracking, psycholinguistic methods, visual context effects, lifespan, flexibility and limitations of method, linking hypothesis

Date Published: 11/30/2018

Citation: Rodríguez Ronderos, C., Münster, K., Guerra, E., Kreysa, H., Rodríguez, A., Kröger, J., Kluth, T., Burigo, M., Abashidze, D., Nunnemann, E., Knoeferle, P. Eye Tracking During Visually Situated Language Comprehension: Flexibility and Limitations in Uncovering Visual Context Effects. *J. Vis. Exp.* (141), e57694, doi:10.3791/57694 (2018).

## Abstract

The present work is a description and an assessment of a methodology designed to quantify different aspects of the interaction between language processing and the perception of the visual world. The recording of eye-gaze patterns has provided good evidence for the contribution of both the visual context and linguistic/world knowledge to language comprehension. Initial research assessed object-context effects to test theories of modularity in language processing. In the introduction, we describe how subsequent investigations have taken the role of the wider visual context in language processing as a research topic in its own right, asking questions such as how our visual perception of events and of speakers contributes to comprehension informed by comprehenders' experience. Among the examined aspects of the visual context are actions, events, a speaker's gaze, and emotional facial expressions, as well as spatial object configurations. Following an overview of the eye-tracking method and its different applications, we list the key steps of the methodology in the protocol, illustrating how to successfully use it to study visually-situated language comprehension. A final section presents three sets of representative results and illustrates the benefits and limitations of eye tracking for investigating the interplay between the perception of the visual world and language comprehension.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/57694/>

## Introduction

Psycholinguistic research has highlighted the importance of eye-movement analyses in understanding the processes implicated in language comprehension. The core of inferring comprehension processes from the gaze record is a hypothesis that links cognition to eye movements<sup>5</sup>. There are three major types of eye movements: saccades, vestibulo-ocular movements, and smooth pursuit movements. Saccades are fast and ballistic movements that happen mostly unconsciously and have been reliably associated with shifts in attention<sup>6</sup>. The moments of relative gaze stability between saccades, known as fixations, are considered to index current visual attention. Measuring the locus of the fixations and their duration in relation to cognitive processes is known as the 'eye-tracking method'. Early implementations of this method served to examine reading comprehension in strictly linguistic contexts (see Rayner<sup>7</sup> for a review). In that approach, the duration of inspecting a word or sentence region is associated with processing difficulty. Eye tracking has, however, also been applied to examine spoken language comprehension during the inspection of objects in the world (or on a computer display<sup>2</sup>). In this 'visual world' eye-tracking version, the inspection of objects is guided by language. When the comprehenders hear *the zebra*, for instance, their inspection of a zebra on the screen is taken to reflect that they are thinking about the animal. In what is known as the visual world paradigm, a comprehender's eye gaze is taken to reflect spoken language comprehension and the activation of associated knowledge (e.g., listeners also inspect the zebra when they hear *grazing*, indicating an action performed by the zebra)<sup>2</sup>. Such inspections suggest a systematic link between language-world relations and eye movements<sup>2</sup>. A common way to quantify this link is by computing the proportion of looks to different pre-determined regions on a screen. This allows researchers to directly compare (across conditions, by participants and items) the amount of attention given to different objects at a particular time and how these values change in millisecond resolution.

Research in psycholinguistics has exploited eye tracking in visual worlds to tease apart competing theoretical hypotheses regarding the architecture of the mind<sup>1</sup>. Eye fixations on depicted objects have, moreover, revealed that comprehenders can—assuming a sufficiently restrictive linguistic context—perform incremental semantic interpretation<sup>8</sup> and even develop expectations about upcoming characters<sup>9</sup>. Such eye gaze data have also shed light on a range of further comprehension processes, such as lexical ambiguity resolution<sup>10,11</sup>, pronoun resolution<sup>12</sup>, the disambiguation of structural and thematic role assignment by means of information in the visual context<sup>13,14,16</sup>, and pragmatic processes<sup>15</sup>, among many others<sup>4</sup>. Clearly, the eye movements to objects during language comprehension can be informative of the implicated processes.

The eye-tracking method is non-invasive and can be used with infants, young, and older language users. A key advantage is that, unlike punctate responses to probes or response button presses in verification tasks, it provides insights over time in a millisecond resolution into how language guides attention and how the visual context (in the form of objects, actions, events, a speaker's gaze, and emotional facial expressions, as well as spatial object configurations) contributes to language processing. The continuity of the measure during sentence comprehension complements well with other post-sentence and post-experiment measures such as from overt picture/video-sentence verification, comprehension questions, and memory recall tasks. Overt responses in these tasks can enrich the interpretation of the eye gaze record by providing insight into the outcome of the comprehension process, memory, and learning<sup>2</sup>. Combining eye tracking with these other tasks has uncovered to which extent different aspects of the visual context modulate visual attention and (immediate, as well as delayed) comprehension across the lifespan.

The presentation of language (spoken or written) and scenes can be either simultaneous or sequential. For instance, Knoeferle and collaborators<sup>17</sup> presented the scene 1,000 ms before the spoken sentence, and it remained present during comprehension. They reported evidence that clipart depictions of action events contribute to the resolution of local structural ambiguity in German subject-verb-object (SVO) compared with object-verb-subject (OVS) sentences. Knoeferle and Crocker<sup>18</sup> presented a clipart scene before a written sentence and tested the incremental integration of clipart events during sentence comprehension. They observed incremental congruence effects, meaning that the participants' reading times of sentence constituents were longer when these mismatched (vs. matched) the event depicted in the preceding scene. In another stimulus presentation variant, the participants first read a sentence describing a spatial relationship and then saw a scene of a particular spatial arrangement involving object line drawings<sup>19</sup>. This study assessed the predictions of computational spatial language models by asking the participants to rate the fit of the sentence given to the scene, with the eye movements being recorded during scene interrogation. The participants' gaze patterns were modulated by the shape of the object that they were confronted with—partially confirming the model predictions and providing data for model refinement.

While many studies have used clipart depictions<sup>17,18,19,20,21,22,23</sup>, it is also possible to combine real-world objects, videos of these objects, or static photographs with spoken language<sup>1,21,24,25,26,46</sup>. Knoeferle and colleagues used a real-world setting<sup>24</sup> and Abashidze and colleagues used a videotaped presentation format for examining action events and tense effects<sup>25</sup>. Varying the precise content of the scenes (e.g., depicting actions or not)<sup>22,27,38</sup> is possible and can also reveal visual context effects. A related study by Rodríguez and collaborators<sup>26</sup> investigated the influence of videotaped visual gender cues on the comprehension of subsequently presented spoken sentences. Participants watched the videos displaying either male or female hands performing a stereotypically gender-related action. Then, they heard a sentence about either a stereotypically male or female action event while simultaneously inspecting a display showing two photographs side by side, one of a man and the other of a woman. This rich visual and linguistic environment allowed the authors to tease apart the effects of language-mediated stereotypical knowledge on comprehension from the effects of the visually presented (hand) gender cues.

A further application of this paradigm has targeted developmental changes in language processing. Eye movements to objects during spoken language comprehension revealed the effects of depicted events in 4 - 5-year olds<sup>27,28</sup> and in older adults<sup>29</sup> in real time, but somewhat delayed compared to young adults. Kröger and collaborators<sup>22</sup> examined the effects of prosodic cues and case marking within an experiment and compared these across experiments in adults and children. Participants inspected an ambiguous action-event scene while listening to a related unambiguously case-marked German sentence. Eye movements revealed that the distinct prosodic patterns helped neither the adults nor the 4- or 5-year-olds when disambiguating who-does-what-to-whom. Sentence-initial case marking, however, influenced adults' but not children's eye movements. This suggests that 5-year-olds' understanding of case marking is not sufficiently robust to enable thematic role assignment (see the study by Özge and collaborators<sup>30</sup>), at least not when action events did not disambiguate thematic role relations. These results are interesting, given that they are in contrast with previous results of prosodic effects on thematic role assignment<sup>31</sup>. Kröger and collaborators<sup>22</sup> proposed that the (more or less supportive) visual context is responsible for the contrasting findings. To the extent that these interpretations hold, they highlight the role of context in language comprehension across the lifespan.

The eye-tracking method combines well with the measures from picture- (or video-)sentence verification tasks<sup>18,20,26</sup>, picture-picture verification tasks<sup>32</sup>, corpus studies<sup>24</sup>, rating tasks<sup>19</sup>, or post-experimental recall tasks<sup>25,33</sup>. Abashidze and collaborators<sup>34</sup> and Kreysa and collaborators<sup>33</sup> investigated the interplay of speaker gaze and real-world action videos<sup>34</sup> and speaker gaze and action depictions<sup>33</sup>, respectively, as cues for upcoming sentence content. By combining the tracking of eye gaze in a scene during language comprehension with a post-experimental memory task, they gained a better understanding of the way in which the listeners' perception of a speaker's gaze and the depicted actions interact and affect both immediate language processing and memory recall. The results revealed the distinct contribution of actions *versus* speaker gaze to real-time comprehension *versus* post-experiment memory recall processes.

While the eye-tracking method can be employed with great flexibility, certain standards are key. The following protocol summarizes a generalized procedure that can be adjusted to different types of research questions according to the researchers' specific needs. This protocol is a standardized procedure employed in the Psycholinguistics Laboratory at the Humboldt-Universität zu Berlin, as well as in the former Language and Cognition Laboratory at the Cognitive Interaction Technology Excellence Cluster (CITEC) at Bielefeld University. The protocol describes a desktop and a remote setup. The latter is recommended for use in studies with children or older adults. All experiments mentioned in the **Representative Results** use an eye tracker device which has a sampling rate of 1,000 Hz and is used together with a head stabilizer, a PC for testing the participants (Display PC) and a PC for monitoring the experiment and the participants' eye movements (Host PC). The main difference of this device to its predecessor is that it allows for binocular eye tracking. The protocol is intended to be sufficiently general for use with other eye-tracking devices that include a head stabilizer and make use of a dual PC setup (Host + Display). However, it is important to keep in mind that other setups will likely have different methods for handling problems such as calibration failures or track loss, in which case the experimenter should refer to the user manual of their specific device.

## Protocol

This protocol follows the ethics guidelines of the institution where the data was collected, *i.e.*, the Cognitive Interaction Technology Excellence Cluster (CITEC) of Bielefeld University and Humboldt-Universität zu Berlin. The experiments conducted at Bielefeld University were approved individually by the Bielefeld University's ethics committee. The psycholinguistics laboratory at the Humboldt-Universität zu Berlin has a laboratory ethics protocol that was approved by the ethics committee of the Deutsche Gesellschaft für Sprachwissenschaft, the German Linguistics Society (DGfS).

### 1. Desktop Setup

NOTE: The following are the key steps in an eye-tracking experiment.

#### 1. Instrument preparation

1. Turn on the eye-tracking camera and the Host PC (see **Figure 1** for an illustration of the in-lab data collection and the pre-processing setup).
2. Initiate the eye tracker software.
3. Turn on the Display PC. Open the folder that contains the experiments.
4. Start the 'deployed' version with a .exe extension by double-clicking it.
5. Depending on the setup, a prompt for selecting an experimental list might appear, which is typically a tab-delimited .csv/.txt file containing the conditions, sound file names, written sentences, and/or images that a participant will be exposed to. Select the list, and the program will start. Name the output file where the results will be stored.  
NOTE: The output file is saved on the Display PC and a back-up is saved on the Host PC.
6. Leave the first screen of the experiment (the introduction or welcome screen) open.
7. Remove the protective cover from the eye tracker camera.
8. For hygienic purposes, place a tissue on the chin holder.
9. Optional: Prepare a response pad and/or a keyboard as required by the experiment.
10. Optional: If presenting auditory stimuli, adjust the volume of the speakers and test them prior to running the experiment. Alternatively, test the functionality of the headphones to be used by the participant.
11. Prepare the necessary forms to be signed by the participant.
12. Adjust the light intensity in the laboratory in order to have a dimly lit room with a constant luminance for the entire duration of the data collection process.

#### 2. Preparation for the participants

1. After the participants have arrived, introduce yourself.
2. Place a **Do not disturb** sign on the laboratory door.
3. Ask the participant to take a seat.
4. Guide the participant through the information sheet and the necessary consent and demographic forms.
5. Let the participant read and sign the consent form.
6. Briefly explain the general aspects of the experiment and its duration. Do not provide too much information prior to the experiment, since this might influence the participant's gaze behavior and the resulting data.
7. Provide written instructions and give the participant the opportunity to ask questions.
8. Briefly explain the function of the eye tracker.
9. If necessary, clarify the task and point out any buttons/keys that will need to be pressed during the experiment.
10. Explain how the chin rest is intended to minimize head movement during the experiment. Mention that the device functions best if the participant avoids any movements.

#### 3. Setting up the eye tracker

1. Prepare the participant for the experiment: ask them to sit at the table and place their chin on the chin rest. Ask them to lean their forehead against the headrest.
2. Ask the participant to adjust the height and position of the chair if necessary: the participant should feel comfortable with their chin on the chin rest and their forehead against the headrest.
3. Explain that it is common to accidentally move the head during the experiment and that this should be avoided. Explain that the head posture might have to be corrected during the experiment.
4. If the experiment requires a button press (*e.g.*, via the response pad), instruct the participant to leave their fingers resting on the buttons to be pressed and to avoid looking down at the response pad when pressing a button.
5. Ask the participant to read the welcome/introduction screen.
6. Sit in front of the Host PC. If the screen is not already on **Camera setup**, click on **Camera setup** to get to the right screen. If the protective cover has been removed from the camera lens and the participant is correctly positioned on the chin rest, the screen should show three images of the participant's eyes: a bigger image at the top, and two smaller ones at the bottom on the left and right. These smaller images show one eye each, corresponding to the left and the right eye.
7. Select the eye to be tracked. It is customary to track the participant's dominant eye. If the participant does not know which eye is dominant, conduct an ocular dominance test (step 1.3.8).
8. Determine ocular dominance. Ask the participant to stretch out one arm and align the thumb with a distant object with both eyes open. Ask the participant to alternate closing the left or right eye. The dominant eye is the one for which the thumb remains aligned with the object when the eye is open<sup>35,36</sup>.
9. Click on **Image | Display pc** (or press **Enter**) and the images described in step 1.3.6 will also appear on the participant's display screen, but only one at a time.

10. Press the left/right arrow on either the experimenter's or the participant's computer keyboard to switch between the smaller and bigger image of the eye. Focus on the smaller image (the eye).
11. Press **A (Align eye window)** on either keyboard in order to center the search limits box on the pupil position. Then, a red square around the eye should appear with a turquoise circle [the 'corneal reflection', (cr)] near the bottom of the pupil. The pupil itself should be blue.
12. Ensure that two crosses ('crosshairs') appear on the screen—one in the center of the pupil and one in the center of the corneal reflection. The red box and the two crosses mean that the eye tracker is detecting the pupil and the cr.  
NOTE: If the red box or the crosses are absent, the eye is not tracked—if this happens, a **no pupil** will appear on the experimenter's computer.
13. Adjust the camera's focus by manually turning the focus lens. Be careful not to touch the front of the lens. Turn the lens until the best focus point is reached.  
NOTE: The best focus point is reached when the turquoise circle (the corneal reflection) is as small as possible (*i.e.*, when this circle is in focus).
14. Set the pupil threshold. Ensure that only the image of the pupil is blue (*e.g.*, the eyelashes should not be blue) on the Host PC. Also, ensure that the entire pupil (not just its central part) is blue. Only worry about what is inside the red square.
15. Press **A**. This automatically sets the pupil threshold. If the pupil is not accurately displayed in blue, adjust the threshold manually using the **UP** key to increase and the **DOWN** key to decrease the amount of blue part of the image surface.  
NOTE: Mascara (which is usually black, like the pupil) may interfere with setting the threshold—the eye tracker might take the dark eyelashes for the pupil. In this case, ask the participant to remove their makeup by providing them with makeup remover tissues.
16. Set the **cr threshold**. If **A** was pressed in step 1.3.15, the **cr threshold** should have automatically been set.  
NOTE: Numeric values for all threshold settings should be visible. If a question mark appears in any of them, there was a problem in one of the preceding steps and the thresholds should be set manually.

#### 4. Calibrating the eye tracker

NOTE: Check whether the eye tracker can consistently identify the position of the eye when the participant looks at other parts of the screen.

1. Ask the participant to look at the four corners of the screen one at a time while the camera setup window is in view. Carefully look for any irregular reflections (these will show up as turquoise 'blobs' on the screen) that interfere with the corneal reflection when the eye gaze is directed at the corner.
2. Ask the participant to look at the center of the screen and then direct their gaze to the problematic corner if the red box around the eye and either of the crosshairs are not visible at any point during step 1.4.1. This will help determine the source of the problem.
3. Re-adjust the position of the participant's head and check if this yields any improvement. Repeat this step if necessary. If the device is still unable to accurately track the participant's gaze after several attempts, abort the experiment.
4. Inform the participant that the eye tracker will be calibrated and that they are going to see a black circle (with a small grey dot) moving to different parts of the screen. Instruct the participant to fixate the circle until it moves to a new location. Instruct the participant to avoid straining their eyes and to focus on the small grey dot inside the black circle for optimal results.
5. Tell the participant that it is important to keep still and not to try to anticipate the position of the next circle during the calibration. Instruct them to follow the circles with their eyes and not their head. Click on **Calibrate** to start the calibration process. Normally, a 9-point calibration procedure is used in which the black circle moves to nine locations in a serial manner.
6. For an automatic calibration, press **ENTER** after the participant has accurately fixated the first dot in the middle of the screen. For manual calibration (*e.g.*, when there are problems tracking the participant's eye or when dealing with special participant groups such as children), accept each fixation by pressing **ENTER** (or by clicking **Accept fixation**/pressing the **Spacebar**).  
NOTE: At the end of the calibration process, an almost rectangular pattern on the experimenter's screen should be visible. This represents the eye gaze patterns of the participant. Additionally, the results of a good calibration should be highlighted in green. If they are not, repeat the calibration procedure (*i.e.*, click on **Calibration**).
7. Validate the results. Tell the participant to go through the same procedure (looking at dots) in order to validate the results of the calibration process. Remind them to look at the dot and to be still.  
NOTE: The process of validation is similar to that of calibration and both results are compared by the eye tracker software to ensure that the eye is being tracked accurately.
8. Click **Validate**.
9. Accept each fixation by pressing **ENTER** (or by clicking **Accept fixation**/pressing the **Spacebar**).
10. After the validation, the results will appear on the experimenter's screen. Pay particular attention to the two error measures, the average error (*e.g.*, 0.23°) and the maximal error (*e.g.*, 0.70°). These represent the degrees to which the tracked image deviates from the actual gaze position of a participant.
11. When using a visual world paradigm, keep the average error (the first number) under 0.5° and the maximum error (the second number) under 1°.
12. If the error values are above the threshold, ask the participant to readjust the position of their head and restart the calibration procedure. If no improvement is observed, abort the experiment.  
NOTE: It is common to observe high calibration errors when the participant wears contact lenses. The participant should be asked in advance to bring their prescription glasses instead of their contact lenses.
13. After successfully completing the calibration process, click **Output/Record** to start the experiment. Inform the participant that the experiment will now begin.

#### 5. During the experiment

NOTE: During the experiment (depending on how an individual experiment is programmed, but typically before each experimental trial), a **drift check** or **drift correct** screen will display a dot in the center of the screen. Its purpose is to report the calculated fixation error for the current trial, and, depending on the eye tracker model, to automatically adjust it.

1. During the instruction phase, the participant has been told to fixate the central dot whenever it appears. Make sure that during every drift check/drift correct phase, participants are again fixating the dot. This can be done by following the participant's gaze on the Host PC, where the participant's gaze appears as a moving green circle.

2. Once the drift check/drift correct has been completed, press **ENTER** (or the **Spacebar**) to make the dot disappear and display the next trial.
  3. Determine how to deal with failure to pass the drift check phase according to the model of the eye tracker used. Depending on the model, the eye tracker will either perform an automatic **drift correction** by adjusting the gaze's coordinates to match those of the central dot, or it will make a beeping noise to prompt the experimenter to re-calibrate before continuing the experiment (**drift check**).
  4. If using an automatic drift correction, keep in mind that too many drift corrections in consecutive trials and/or too large a degree of drift correction will distort the results and require a re-calibration of the device.
6. **Re-calibration during the experiment**
1. It is possible to re-calibrate at any time during the experiment. During the presentation of the **Drift correction/drift check** screen, click on **Camera setup**, then click on **Calibrate**. Go through the calibration and validation processes until a satisfactory value is reached, then click on **Output/record**. The experiment will resume from the exit point in the trial sequence.
7. **After the experiment**
1. Provide the participant with a questionnaire to assess if they were able to guess the key experimental manipulations. Here, it is also important to ask about potential strategies that could have been developed throughout the experiment.
  2. Debrief the participant about the purpose of the experiment. Thank them for their participation and provide the necessary monetary compensation or assign course credit if applicable.

## 2. Remote Setup: Adjusting the Setup for Studies with Children and Older Adults

NOTE: This section describes only the differences between a remote setup and a desktop setup as described in step 1. Points not explicitly mentioned here should be assumed to be identical to the procedure described in step 1.

1. Exchange the standard 35 mm eye tracker camera lens for a 16 mm lens.
2. Connect all necessary equipment (Host PC, speakers, eye tracker, and laptop if a laptop is used).
3. Place the laptop on a laptop stand and the eye tracker in front of it (the participant should be able to see the top 75% of the screen).
4. Place a "target sticker" (available from the tracker's manufacturer) on the participant's forehead (above the eyebrow of the right eye or on the right cheek if the forehead is too small (*i.e.*, in case of infants); this sticker replaces the chin rest of the desktop setup and it allows the eye tracker to accurately determine the position of the participant's head).
5. Make sure that the participant is seated 550 - 600 mm away from the camera (the distance of the target sticker to the camera).
6. Make sure that the words "big angle" or "near eye" do not appear on the Host PC. If they do, it means that the position of the target sticker is not ideal. In this case, readjust the target sticker. It can also mean that the participant's forehead is particularly small. If this is the case, place the sticker on the participant's cheek.  
NOTE: If the sticker is placed too close to the ear, the message "big angle" is likely to appear and the sticker must be repositioned.
7. Before starting the calibration process described in step 1.4, make sure that the participant is sitting comfortably. Ask them to maintain the same position throughout the whole experiment and explain that the eye tracker is very sensitive to bodily movements. Insist that they keep their head still. If the participant moves too much, the device will emit a humming noise.  
NOTE: If the remote eye tracker setup is being used to test sensitive populations, such as children or older adults, it is advisable to use a manual calibration procedure.

## 3. Adjusting the Setup for Reading Studies

NOTE: When investigating visual context effects on reading, it is necessary to pay particular attention to the calibration and re-calibration processes. As opposed to visual world studies, eye tracking during reading requires a much higher degree of device precision, given the accuracy needed to track word-to-word and letter-to-letter reading patterns.

1. Make sure that both the average and maximum error displayed during the validation phase remain under 0.5°.
2. Make sure to use at least a 9-point scale for calibration. This assures a more accurate tracking of the eye gaze position, which is essential, given the small size of the areas of interest during reading.

## Representative Results

A study by Münster and collaborators<sup>37</sup> investigated the interplay of sentence structure, depicted actions, and facial emotional cues during language comprehension. This study is well-suited to illustrate the advantages and limitations of the method, as it showed both robust depicted action effects and marginal effects of facial emotional cues on sentence comprehension. The authors created 5 s videos of a woman's facial expression that changed from a resting position into a happy or into a sad expression. They also created emotionally positively valenced German Object-Verb-Adverb-Subject (OVAdvS) sentences of the form 'The [object/patient<sub>accusative case</sub>] [Verb] [positive Adverb] the [subject/agent<sub>nominative case</sub>].' Via the positive adverb, the sentences matched the 'happy' video and mismatched the 'sad' video, permitting in principle anticipation of the agent (who was smiling and described as acting happily by the positive adverb). Following the speaker video, the sentence appeared with one of two versions of an agent-patient-distractor clipart scene. In one version, the agent was depicted as performing the mentioned action on the patient, while the distractor character performed a different action. The other scene version depicted no actions between the characters. Eye movements in the scene revealed the effects of the actions and of the speaker's facial expression on sentence comprehension.

The action depiction rapidly affected the participants' visual attention, meaning that the participants looked more at the agent than at the distractor when the mentioned action was (vs. was not) depicted. These looks were anticipatory (*i.e.*, occurring before the agent was mentioned), suggesting that the action depiction clarified the agent before the sentence did. The earliest effect of the action depiction emerged during the verb (*i.e.*, the verb mediated the action-associated agent). By contrast, whether the preceding speaker smiled or looked unhappy had no clear effect on the agent anticipation (**Figure 2**). The latter result could reflect the more tenuous link between a speaker's smile and a positive sentential adverb relating to the depicted agent's action (compared with a direct verb-action reference mediating an action-associated agent). Alternatively, it could be specific to the task and stimulus presentation: perhaps emotion effects would have been more pronounced in a more socially interactive task or in one that presents the speaker's face during (rather than before) sentence comprehension. However, presenting a smiling speaker's face during comprehension might cause participants to focus on the face at the expense of other scene content, perhaps masking otherwise observable effects of the manipulated variables (source: unpublished data).

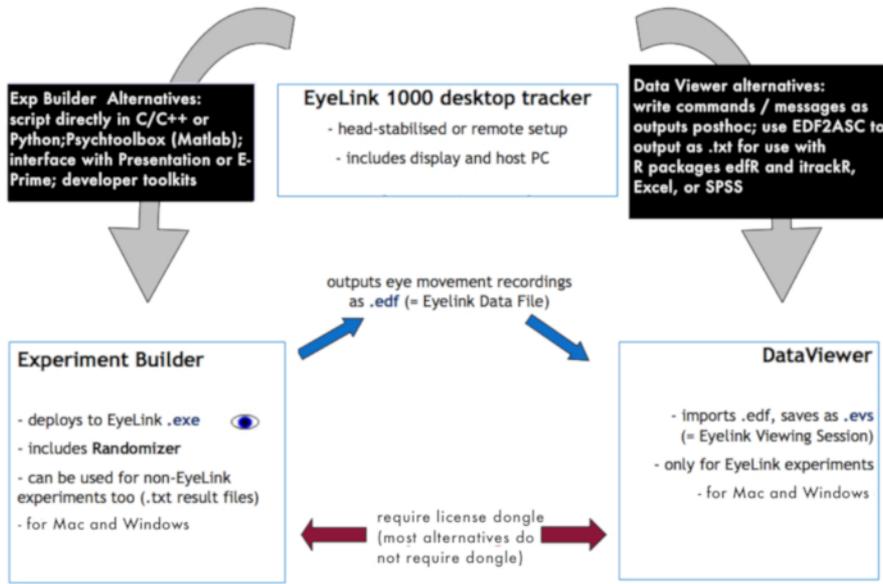
In a further variant of the paradigm, Guerra and Knoeferle<sup>32</sup> asked whether spatial-semantic world-language relations can modulate the comprehension of abstract semantic content during reading. Guerra and Knoeferle borrowed the idea from the conceptual metaphor theory<sup>38</sup> that spatial distance (*e.g.*, proximity) grounds the meaning of abstract semantic relations (*e.g.*, similarity). In line with this hypothesis, participants read coordinated abstract nouns faster when they were similar (vs. opposite) in meaning and had been preceded by a video conveying proximity (vs. distance; playing cards move closer together vs. farther apart). In a second set of studies<sup>39</sup>, sentences described the interaction between two people as intimate or unfriendly, leading to the discovery that videos of two cards approaching one another sped up the reading of sentence regions that conveyed social proximity/intimacy. Note that the spatial distance affected sentence reading rapidly and incrementally even when the sentences did not refer to the objects in the video. The videos distinctively modulated reading times as a function of the congruence between spatial distance and semantic as well as social aspects of sentence meaning. These effects appeared both in first-pass reading times (the duration of the first inspection of a pre-determined sentence region) and the total time spent on that sentence region (see **Figure 3** for an illustration of the results from the studies by Guerra and Knoeferle)<sup>32</sup>. However, the analyses also revealed substantial variation between participants, leading to the conclusion that such subtle card-distance effects may not be as robust as the effects of verb-action relations, to mention one example.

A further set of studies illustrates how variation in sentence structure can help evaluate the generality of visual context effects. Abashidze and collaborators<sup>34</sup> and Rodríguez and collaborators<sup>26</sup> examined the effects of recent actions on the subsequent processing of spoken sentences. In both studies, participants first inspected an action video (*e.g.*, an experimenter flavoring cucumbers, or female hands baking a cake). Next, they listened to a German sentence that was either related to the recent action or to another action that might be performed next (flavoring tomatoes<sup>34</sup>, building a model<sup>26</sup>). During comprehension, the participants inspected a scene showing two objects (cucumbers, tomatoes)<sup>34</sup> or two photographs of agent faces (a female and a male agent face, named 'Susanna' and 'Thomas', respectively)<sup>26</sup>. In the study by Abashidze and collaborators<sup>34</sup>, the speaker first mentioned the experimenter and then the verb (*e.g.*, flavoring), eliciting expectations about a theme (*e.g.*, the cucumbers or the tomatoes). In the study by Rodríguez and collaborators<sup>26</sup>, the speaker first mentioned a theme (the cake), and then the verb (baking), eliciting expectations about the agent of the action (female: Susanna, or male: Thomas, depicted *via* photos of a female and a male face).

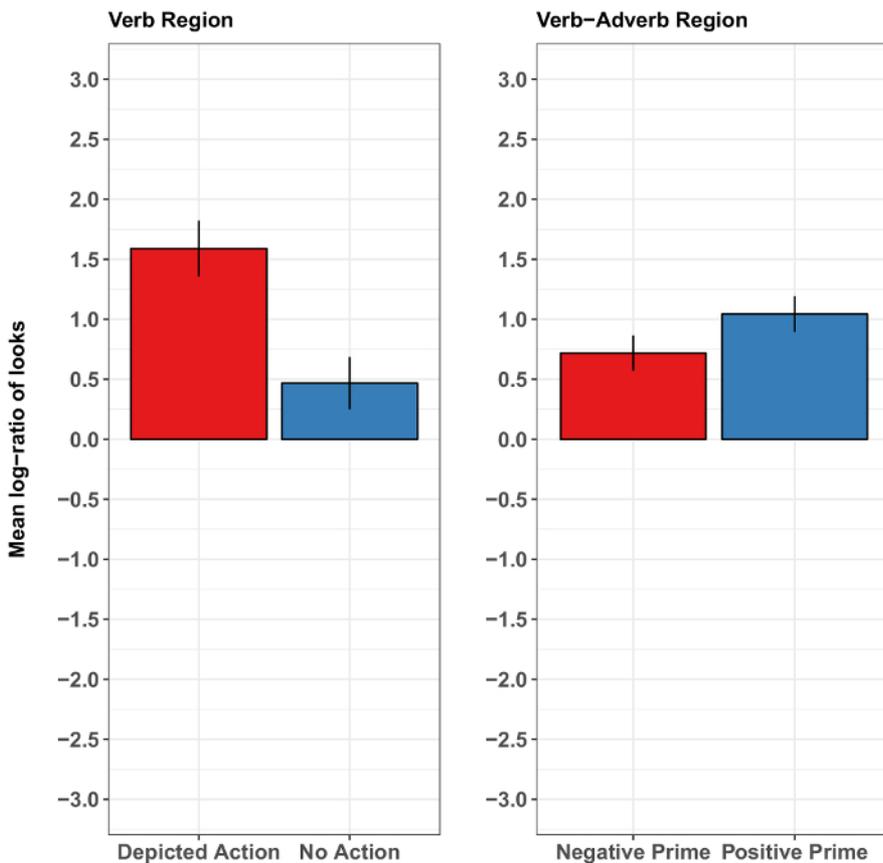
In both studies, the question at issue was whether people would (visually) anticipate the theme/agent of the recent action event or the other theme/agent based on further contextual cues provided during comprehension. In Abashidze and collaborators<sup>34</sup>, the experimenter's gaze cued the future theme object (the tomatoes) from the onset of the verb (literally translated from German: 'The experimenter-agent flavors soon ...'). In Rodríguez and collaborators<sup>26</sup>, gender knowledge of stereotypical actions became available when the theme and the verb were mentioned (*e.g.*, the literal translation of the German stimuli was: 'The cake-theme bakes soon ...'). In both sets of studies, the participants preferentially inspected the target/agent of the recent action (the cucumbers<sup>34</sup> or Susanna<sup>26</sup>) over the alternative (future/other-gender) target (the tomatoes<sup>34</sup> or Thomas<sup>26</sup>) during the sentence.

This so-called 'recent-event preference', thus, appears to be robust across substantial variation in sentence structure and experimental materials. It was modulated, however, by visual constraints from the concurrent scene<sup>26</sup>, such that presenting plausible photographs of verb themes in addition to gendered photographs of agents reduced the reliance on the recently-inspected action events and modulated attention based on gender-stereotype knowledge conveyed by language. **Figure 4** illustrates the main findings of the experiments by Rodríguez and collaborators<sup>26</sup>.

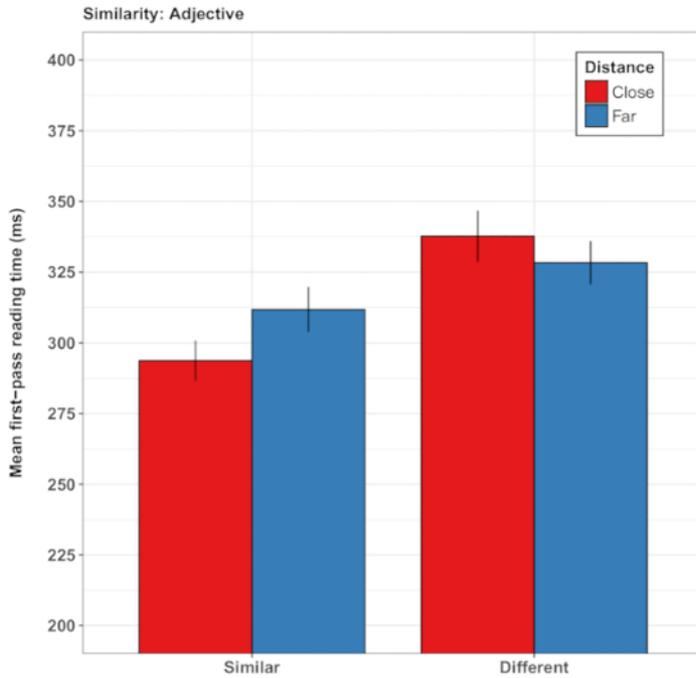
While this version of the visual-world paradigm yielded robust results, other studies have highlighted the complexity (and limitations) of the linking hypothesis. Burigo and Knoeferle<sup>20</sup> uncovered that participants—when listening to utterances such as *Die Box ist über der Wurst* ('The box is above the sausage')—mostly followed the utterance in inspecting clipart depictions of these objects. But on a proportion of trials, the participants' gaze decoupled from what was mentioned. Upon hearing 'sausage' and having inspected the sausage at least once, the participants' next inspection returned to the box on approximately 21% of the trials in speeded verification (Experiment 1) and in 90% of the trials in post-sentence verification (Experiment 2). This gaze pattern suggests that the reference (hearing 'sausage') guided only some (inspecting the sausage) but not all eye movements (inspecting the box). This sort of design could be used to tease apart lexical-referential processes from other (including sentence-level interpretative) processes. However, researchers must be careful when making claims about different levels of linguistic processing based on relative differences in eye-movement proportions, given the ambiguity in linking eye gaze to cognitive and comprehension processes.



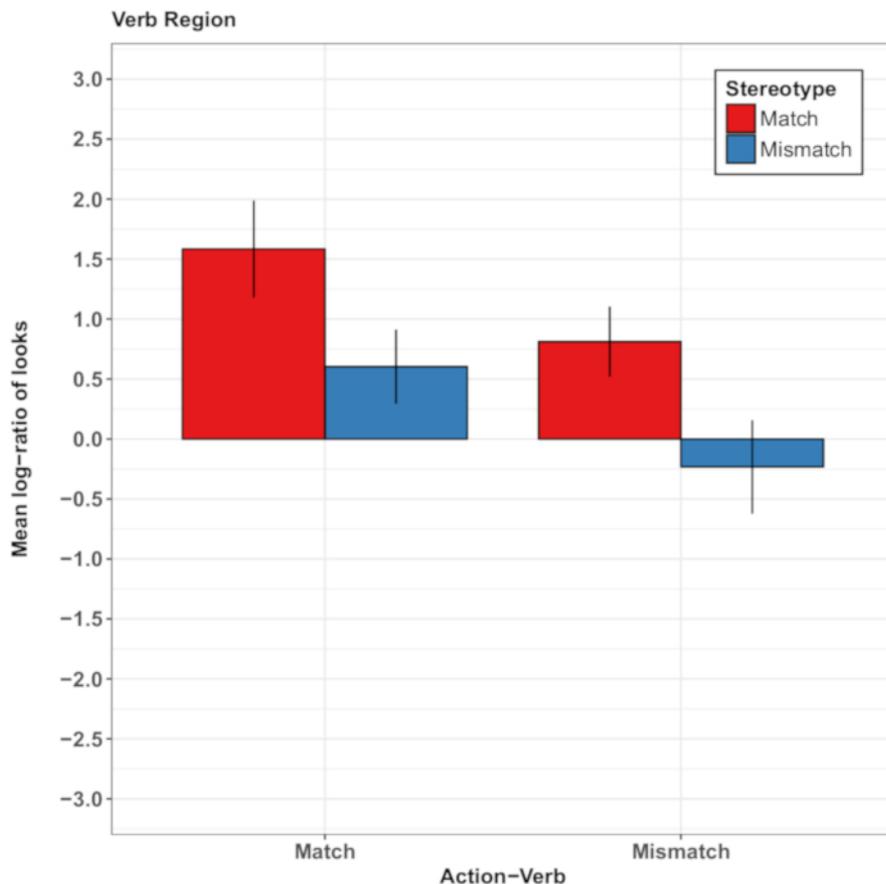
**Figure 1: Overview of the data collection environment.** The graph shows how the different software and hardware elements used for data collection and preprocessing relate to one another. [Please click here to view a larger version of this figure.](#)



**Figure 2: Representative results of Münster and collaborators<sup>37</sup>.** These panels show the by-subject mean log-gaze probability ratios per condition in the verb region (for the depicted action) and in the verb-adverb region combined (for the effect of emotional prime). The results show a considerably higher proportion of looks towards the target image when the action mentioned in the sentence was depicted than when it was not depicted. The results for the effect of emotional facial primes were less conclusive: they suggest only a slight increase in the mean-log ratio of looks to the target when the facial prime had a positive emotional valence (a smile) than when it had a negative valence (a sad face). The error bars represent the standard error of the mean. This figure has been modified from Münster *et al.*<sup>37</sup>. [Please click here to view a larger version of this figure.](#)



**Figure 3: Representative results of Guerra and Knoeferle<sup>32</sup>.** This panel shows the mean first-pass reading times (in milliseconds) of the (dis)similarity adjective. The results show shorter reading times for similarity sentences after seeing two playing cards move closer together compared with farther apart. The error bars represent the standard error of the mean. This figure has been modified from Guerra and Knoeferle<sup>32</sup>. [Please click here to view a larger version of this figure.](#)



**Figure 4: Representative results of Rodríguez and collaborators<sup>26</sup>.** This panel shows the by-subject mean log gaze probability ratios per condition in the verb region. The proportions of looks above 0 show a preference for the target agent image and the proportions below 0 show a preference for the competitor agent image. The results show that participants were more likely to inspect the target agent image when the object and the verb mentioned in the sentence matched the previous videotaped events than when they did not. Additionally, there were more looks toward the target agent image when the action described by the sentence conformed to gender stereotypes. The error bars represent the standard error of the mean. This figure has been modified from Rodríguez *et al.*<sup>26</sup>. [Please click here to view a larger version of this figure.](#)

## Discussion

In summary, the reviewed variants of eye tracking in visual contexts have uncovered many ways in which a visual scene can affect language comprehension. This method provides crucial advantages compared to methods such as measuring reaction times. For instance, ongoing eye movements provide us with a window into language comprehension processes and how these interact with our perception of the visual world over time. Moreover, participants are not necessarily required to perform an explicit task during language comprehension (such as judging the grammaticality of a sentence *via* a button press). This allows researchers to use the method with populations that might struggle with overt behavioral responses other than eye gaze, such as infants, children and, in some cases, older adults. Eye tracking is ecologically valid in that it reflects participants' attention responses—not unlike humans' visual interrogation of communication-relevant things in the world around them during more or less attentive listening to unfolding utterances.

One of the boundaries (or perhaps characteristics) of the visual world paradigm is that not all events can be depicted straightforwardly and unambiguously. Concrete objects and events can, of course, be depicted. But how abstract concepts are best depicted is less clear. This can limit (or define) insights into the interaction between language processing and the perception of the visual world using an eye-tracking visual world paradigm. Further challenges relate to the linking hypotheses between observed behavior and comprehension processes. Eye fixations are a single behavioral response that likely reflects many subprocesses during language comprehension (*e.g.*, lexical access, referential processes, language-mediated expectations, visual context effects, among others). Given this insight, researchers must be careful not to over- or misinterpret the observed gaze pattern. To address this problem, prior research has highlighted the role of comprehension subtasks to clarify the interpretation of the gaze record<sup>40</sup>.

One way to enhance the interpretability of eye movements is to integrate them with other measures such as event-related brain potentials (ERPs). By investigating the same phenomenon with two methods that are comparable in their temporal granularity and complementary in their linking hypotheses, researchers can rule out alternative explanations of their results and enrich the interpretation of each individual measure<sup>41</sup>. This approach has been pursued across experiments<sup>43</sup>, but, more recently, also within a single experiment (albeit in strictly linguistic contexts)<sup>44</sup>. Future research could benefit greatly from such methodological integration and continued combination with post-trial and post-experimental tasks.

The eye-tracking method can replicate established results, as well as test new hypotheses, about the interaction of visual attention in scenes with language comprehension. The procedure outlined in the protocol must be carefully followed since even minor experimenter errors can affect data quality. In reading studies, for example, the relevant analysis regions are often individual words or even letters, meaning that even small calibration errors might distort the results (see the article by Raney and colleagues<sup>42</sup>). Steps 1.4 and 1.5 of the protocol, the calibration of the eye tracker and the drift check/drift correct, are of particular importance since they directly impact the recording accuracy. Failure to correctly calibrate the eye tracker can result in the tracker not accurately tracking eye movements to the pre-established areas of interest. Such tracking failure will lead to missing data points and a loss in statistical power which can be problematic when investigating world-to-language relationships that are very subtle and yield small statistical effect sizes (see the description of the experiments by Guerra and Knoeferle<sup>32</sup> and Münster and colleagues<sup>37</sup> among the **Representative Results**).

Given the need to maximize power and sensitivity of the equipment, it is important that experimenters know how to deal with problems that routinely occur during an experimental session. For example, the pupil position and movement of participants wearing glasses can result in calibration difficulties due to light reflections on the lenses of a participant's glasses. One way to solve this problem is to mirror the image of the participant's eye on the Display PC and encourage them to move their head until the reflection of light on the glasses is no longer visible on their screen, meaning that it is no longer captured by the camera. A further cause of calibration failure can be pupil constriction, which may be a consequence of an overexposure to light. In that case, dimming the light in the laboratory will increase pupil dilation and, thus, help the eye tracker in accurately detecting the pupil.

As a final thought, we would like to address the potential that the visual-world paradigm has for research on second-language learning. The paradigm has already been successfully used in psycholinguistic research to investigate phenomena such as cross-language lexical and phonological interaction<sup>46,47,48</sup>. In addition, the close link between visual attention and language learning has frequently been highlighted in the applied-linguistics literature on second-language acquisition<sup>49,50,51</sup>. Future research on second-language learning will likely continue to benefit from the advantageous position of eye tracking as a method that provides an index of visual attention in millisecond resolution.

## Disclosures

The authors have nothing to disclose.

## Acknowledgements

This research was funded by the ZuKo (Excellence Initiative, Humboldt-Universität zu Berlin), the Excellence Cluster 277 'Cognitive Interaction Technology' (German Research Council, DFG), and by the European Union's Seventh Framework Program for research, technological development, and demonstration under grant agreement n°316748 (LanPercept). The authors also acknowledge support from the Basal Funds for Centers of Excellence, Project FB0003 from the Associative Research Program of CONICYT (Government of Chile), and from the Project "FoTeRo" in the Focus center XPrag (DFG). Pia Knoeferle provided a first draft of the article informed by a laboratory protocol that Helene Kreysa instantiated at Bielefeld University and that continues to be used at the Humboldt-Universität zu Berlin. All authors contributed to the contents by providing input on methods and results in one form or another. Camilo Rodríguez Ronderos and Pia Knoeferle coordinated the input from the authors and, in two iterations, substantially revised the initial draft. Ernesto Guerra produced **Figures 2 - 4** based on input from Katja Münster, Alba Rodríguez, and Ernesto Guerra. Helene Kreysa provided **Figure 1** and Pia Knoeferle updated it. Parts of the reported results have been published in the *Proceedings of the Annual Meeting of the Cognitive Science Society*.

## References

1. Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C. Integration of visual and linguistic information in spoken language comprehension. *Science*. **268** (5217), 1632-1634 (1995).
2. 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. *Cognitive Psychology*. **6** (1), 84-107 (1974).
3. Knoeferle, P., Guerra, E. What is non-linguistic context? A view from language comprehension. In *What is a Context? Linguistic Approaches and Challenges*. edited by Finkbeiner, R., Meibauer, J., Schumacher, P. B., 129-150, John Benjamins Publishing Company. Amsterdam, The Netherlands (2012).
4. Knoeferle, P., Guerra, E. Visually situated language comprehension. *Language and Linguistics Compass*. **10** (2), 66-82 (2016).
5. Just, M. A., Carpenter, P. A. A theory of reading: From eye fixations to comprehension. *Psychological Review*. **87** (4), 329-354 (1980).
6. Deubel, H., Schneider, W. X. Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*. **36** (12), 1827-1837 (1996).
7. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*. **124** (3), 372-422 (1998).
8. Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., Carlson, G. N. Achieving incremental semantic interpretation through contextual representation. *Cognition*. **71** (2), 109-147 (1999).
9. Kamide, Y., Scheepers, C., Altmann, G. T. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*. **32** (1), 37-55 (2003).
10. 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. *Journal of Memory and Language*. **38** (4), 419-439 (1998).
11. Tanenhaus, M. K., Magnuson, J. S., Dahan, D., Chambers, C. 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 (2000).
12. Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., Trueswell, J. C. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*. **76** (1), B13-B26 (2000).
13. Chambers, C. G., Tanenhaus, M. K., Magnuson, J. S. Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **30** (3), 687-696 (2004).

14. Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., Sedivy, J. C. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*. **45** (4), 447-481 (2002).
15. Huang, Y. T., Snedeker, J. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*. **58** (3), 376-415 (2009).
16. Trueswell, J. C., Sekerina, I., Hill, N. M., Logrip, M. L. The kindergarten-path effect: studying online sentence processing in young children. *Cognition*. **73** (2), 89-134 (1999).
17. Knoeferle, P., Crocker, M. W., Scheepers, C., Pickering, M. J. The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye movements in depicted events. *Cognition*. **95** (1), 95-127 (2005).
18. Knoeferle, P., Crocker, M. W. Incremental Effects of Mismatch during Picture-Sentence Integration: Evidence from Eye-tracking. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. 1166-1171. Stresa, Italy (2005).
19. Kluth, T., Burigo, M., Schultheis, H., Knoeferle, P. The role of the center-of-mass in evaluating spatial language. *Proceedings of the 13th Biannual Conference of the German Society for Cognitive Science*, 11-14. Bremen, Germany (2016).
20. Burigo, M., Knoeferle, P. Visual attention during spatial language comprehension. *PLoS One*. **10** (1) (2015).
21. Münster, K., Carminati, M. N., Knoeferle, P. How Do Static and Dynamic Emotional Faces Prime Incremental Semantic Interpretation? Comparing Older and Younger Adults. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. 2675-2680. Quebec City, Canada (2014).
22. Kroeger, J. M., Münster, K., Knoeferle, P. Do Prosody and Case Marking influence Thematic Role Assignment in Ambiguous Action Scenes? *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. 2463-2468. London, UK (2017).
23. Knoeferle, P., Crocker, M.W. The influence of recent scene events on spoken comprehension: evidence from eye movements. *Journal of Memory and Language*. **57** (4), 519-543 (2007).
24. Knoeferle, P., Carminati, M. N., Abashidze, D., Essig, K. Preferential inspection of recent real-world events over future events: evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology*. **2**, 376 (2011).
25. Abashidze, D., Knoeferle, P., Carminati, M. N. How robust is the recent-event preference? *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. 92-97. Quebec City, Canada (2014).
26. Rodríguez, A., Burigo, M., Knoeferle, P. Visual constraints modulate stereotypical predictability of agents during situated language comprehension. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. 580-585. Philadelphia, USA (2016).
27. Zhang, L., Knoeferle, P. Visual Context Effects on Thematic Role Assignment in Children versus Adults: Evidence from Eye Tracking in German. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2593-2598. Boston, USA (2012).
28. Zhang, Lu, Kornbluth, L., Knoeferle, P. The role of recent versus future events in children's comprehension of referentially ambiguous sentences: Evidence from eye tracking. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 1227-1232. Boston, USA (2012).
29. Münster, K., Knoeferle, P. Situated language processing across the lifespan: A review. *International Journal of English Linguistics*. **7** (1), 1-13 (2017).
30. Özge, D. et al. Predictive Use of German Case Markers in German Children. *Proceedings of the 40th Annual Boston University Conference on Language Development*. 291-303. Somerville, USA (2016).
31. Weber, A., Grice, M., Crocker, M. W. The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*. **99** (2), B63-B72 (2006).
32. Guerra, E., Knoeferle, P. Effects of object distance on incremental semantic interpretation: similarity is closeness. *Cognition*. **133** (3), 535-552 (2014).
33. Kreysa, H., Knoeferle, P., Nunnemann, E. Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. 2513-2518. Quebec City, Canada (2014).
34. Abashidze, D., Knoeferle, P., Carminati, M. N. Eye-tracking situated language comprehension: Immediate actor gaze versus recent action events. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. 31-36. Pasadena, USA (2015).
35. Crovitz, H. F., and Zener, K. A Group-Test for Assessing Hand- and Eye-Dominance. *The American Journal of Psychology*. **75** (2), 71-276 (1962).
36. Ehrenstein, W.H., Arnold-Schulz-Gahmen, B.E., & Jaschinski, W. Eye preference within the context of binocular functions. *Graefes Arch Clin Exp Ophthalmol*. **243** 926 (2005).
37. Münster, K., Carminati, M. N., Knoeferle, P. The Effect of Facial Emotion and Action Depiction on Situated Language Processing. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. 1673-1678. Pasadena, USA (2015).
38. Lakoff, G., Johnson, M. *Metaphors We Live By*. University of Chicago Press. Chicago, USA (1980).
39. Guerra, E., Knoeferle, P. Visually perceived spatial distance affects the interpretation of linguistically mediated social meaning during online language comprehension: An eye tracking reading study. *Journal of Memory and Language*. **92**, 43-56 (2017).
40. Kreysa, H., Knoeferle, P. Effects of speaker gaze on spoken language comprehension: task matters. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 1557-1562. Boston, USA (2012).
41. Knoeferle, P. Language comprehension in rich non-linguistic contexts: Combining eye tracking and event-related brain potentials. In *Towards a cognitive neuroscience of natural language use*. edited by Roel, W., 77-100, Cambridge University Press. Cambridge, UK (2015).
42. Raney, G. E., Campell, S. J., Bovee, J. C. Using eye movements to evaluate cognitive processes involved in text comprehension. *Journal of Visualized Experiments*. (83), e50780 (2014).
43. Knoeferle, P., Habets, B., Crocker, M. W., Muentz, T. F. Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*. **18** (4), 789-795 (2008).
44. Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., Kliegl, R. Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*. **140** (4), 552-572 (2011).
45. Knoeferle, P., Kreysa, H. Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*. **3**, 538 (2012).
46. Chambers, C. G., Cooke, H. Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **35** (4), 1029 (2009).
47. Duyck, W., Van Assche, E., Drieghe, D., Hartsuiker, R. J. Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **33** (4), 663 (2007).
48. Weber, A., Cutler, A. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*. **50** (1), 1-25 (2004).

49. Verhallen, M. J., Bus, A. Young second language learners' visual attention to illustrations in storybooks. *Journal of Early Childhood Literacy*. **11** (4), 480-500 (2011).
50. Robinson, P., Mackey, A., Gass, S. M., Schmidt, R. Attention and awareness in second language acquisition. *The Routledge Handbook of Second Language Acquisition*. 247-267 (2012).
51. Tomlin, R. S., Villa, V. Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*. **16** (2), 183-203 (1994).