5

VISUAL WORLD EYE-TRACKING

Paola E. Dussias, Jorge Valdés Kroff, and Chip Gerfen

Preliminary Issues

Our goal in this chapter is to describe the basic experimental design of a visual world study and to discuss the effects that researchers test for, including how these help address questions in second language processing. In part, we cover these points by illustrating two studies that have been particularly influential in the field: Allopenna, Magnuson, & Tanenhaus (1998) and Lew-Williams and Fernald (2007). Several chapters and articles have been written that extensively cover the visual world paradigm in depth (for thorough technical reviews concerning the paradigm as mainly applied to monolingual research, see; Altmann, 2011b; Huettig, Rommers, & Meyer, 2011; Tanenhaus, 2007; Tanenhaus & Trueswell, 2006). We also discuss some of the recent work in the second language (L2) literature to highlight how the method has helped researchers inform key issues in second language acquisition (SLA).

Before reviewing how to carry out a visual world study, we introduce four important core elements that will partially determine the decisions researchers make when designing visual world experiments. These decisions will depend greatly on a number of factors including but not limited to: the resources available to the researcher in terms of equipment, the population that the researcher wants to test (e.g., children versus adults), the sampling rate of the system, and the instructions to participants. All of these are likely to be dependent on the research questions that the researcher wants to address, while one relates to equipment. As discussed in Chapter 4 (Keating, this volume), an eye-tracking setup is considerably less expensive than an ERP setup, but more expensive than typical behavioral methods which require a single PC and perhaps a button box and/or a microphone and some software (e.g., self-paced reading, see Jegerski, Chapter 2, this volume). Eye-tracking

systems vary in terms of the type of hardware they use and consequently in terms of the software necessary to develop an experiment and to extract and analyze data. Studies involving children (e.g., Snedeker & Trueswell, 2004), as well as those that utilize a paradigm referred to as the *looking-while-listening* paradigm (Fernald, Perfors, & Marchmann, 2006), employ commercial video cameras. In one version of these studies, participants sit in front of an inclined podium, where a video camera is hidden beneath the podium. The podium has a hole in the center to allow the lens of the camera to focus on the participant's face (see Figure 5.1).

In each quadrant of the podium, there is a prop that is used by participants to perform certain actions. Participants hear a prerecorded command (e.g., "put the doll in the box) and are asked to perform the action. A second camera is placed behind the participant to record the actions and the location of the props. Using hidden cameras as a method of recording eye movements is desirable with small children because the method is not invasive, it is less expensive than commercially-available systems, and is more portable (Snedeker & Trueswell, 2004). An alternative to the basic video camera setup is to employ one of the many models of experimental eye-tracking systems, developed by a variety of companies, which come with computer eye-tracking algorithms to measure fixations. The most common are head-mounted, desk-mounted, and tower-mounted systems, although lately technological advances have been made toward the development

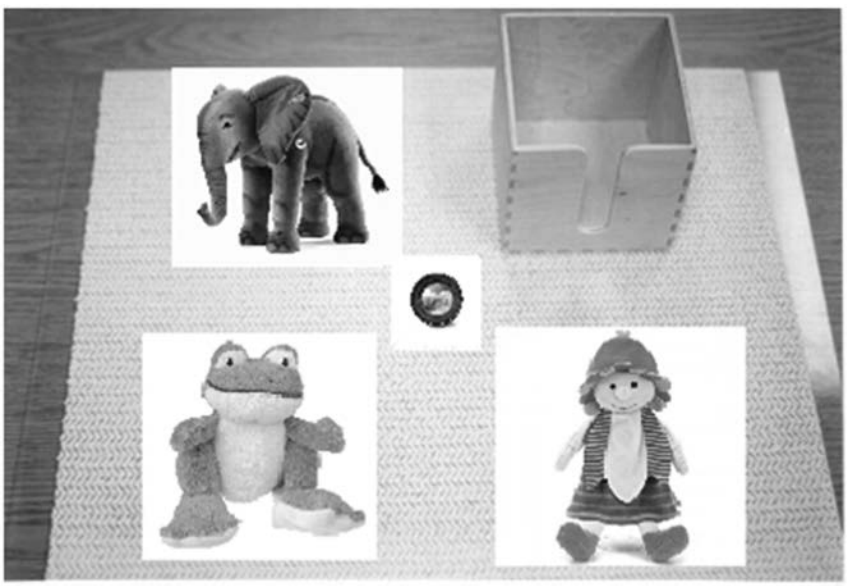

FIGURE 5.1 Sample visual scene using a hidden camera setup and real objects for an action-based experiment. Participants would hear a recorded stimulus such as "Put the doll in the box," and follow the instructions using their hands to manipulate the objects.

of eye-tracking goggles. Most language labs are likely to have an eye-tracker developed by Applied Science Laboratories, SensoMotoric Instruments (SMI), SR Research, or Tobii Technology. Head-mounted eye-trackers require that a participant wear a padded headband with miniature cameras mounted on the headband to record eye movements. This system requires more participant set-up and training time than other systems. Some eye-trackers are directly embedded into specialized computer monitors. These systems are generally more portable (although they can still be cumbersome) and require a less intense participant set-up. Eye-trackers also come as small desk-mounted or tower-mounted devices, generally set at a fixed position just below a computer monitor. These devices are easier to set up than head-mounted eye-trackers and do not produce the discomforts associated with head-mounted devices.

Cost restrictions, portability, and population of interest also influence the kind of eye-tracker used. As mentioned earlier, commercial video cameras are considerably cheaper than eye-trackers specifically designed for experimental research; however, they require intense manual data codification and extraction (explained in more detail below). In contrast, commercially available eye-trackers are generally more expensive, but come with experimental software specifically created to conduct eye-tracking research and with technical support staff who have expertise with the visual world and other research paradigms (e.g., the support group at SR Research). Commercial eye-trackers also allow experimenters to track the progress of a participant and make any necessary adjustments if the eye is not being properly tracked during the course of an experimental session.

A third issue to consider is the sampling rate of the system. Sampling rates determine the frequency with which a data point is recorded and are indicated in hertz (Hz) but are easily converted into time measurements by dividing the value into 1000 (milliseconds). For example, an eye-tracker with a sampling rate of 500 Hz records a data point every 2 milliseconds (1000/500 = 2) whereas an eye-tracker with a sampling rate of 1000 Hz records a data point every millisecond (1000/1000 = 1). In essence, there is a tradeoff between the sampling rate and the degree of movement that the participant can engage in. The higher the sampling rate, the more stable the participant's head must be. Because of this, eye-trackers with higher sampling rates are not particularly well-suited for young children. If a particular research question aims to investigate fairly subtle changes in the time-course of sentence processing, such as whether second language speakers are able to process in a native-like fashion a vowel contrast that does not exist in their native language (e.g. *bit* versus *bet* versus *beet* for Spanish learners of English), a higher sampling rate will be necessary to have sufficient data.

Finally, visual world studies differ with regard to the instructions that participants are given. These can be broadly defined as falling into two categories—action-based and passive listening (Tanenhaus & Trueswell, 2007). As the name implies, action-based instructions require participants to carry out an action related to the linguistic stimuli that they have just heard. Under this version of a

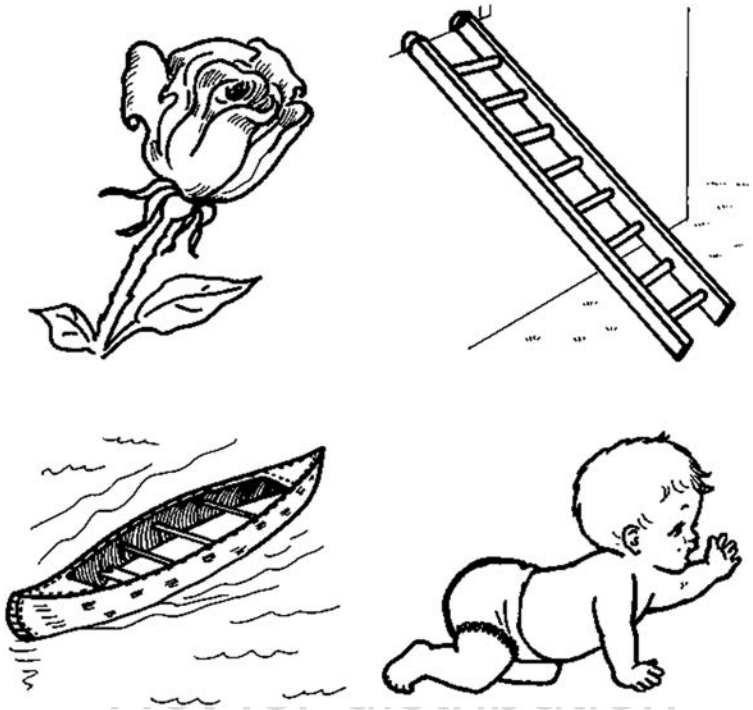

FIGURE 5.2 Sample visual scene employed in an action-based visual world experiment using a commercial eye tracking system and digital images. Participants would hear a recording such as "Click on the flower," and follow the instructions using the computer mouse.

visual world study, participants generally see a series of objects (animate or inanimate) either presented in realia (i.e., stuffed animals, plastic toys, or other types of props; Figure 5.1) or as images on a computer screen, as illustrated in Figure 5.2.

Objects are thus presented without any contextual visual information (i.e., no visual scene). In the simplest and most classic version of this design (e.g., Allopenna et al., 1998), participants hear a sentence like "*Put the doll in the box*" or "*Click on the flower*" and then carry out the action, either by physically moving the target object or by using a computer mouse while their eye movements are recorded. In contrast, in a passive listening task, objects are embedded within a contextually rich visual scene (see Figure 5.3). In one of the first studies to employ this design, Altmann & Kamide (1999) presented participants with a picture of a boy seated in a room surrounded by various objects including a cake and a toy car. Here, participants heard a sentence such as "*The boy will eat the cake.*" Instead of clicking on any named objects, participants were instructed to respond *yes* or *no* if the visual scene was compatible with what they heard. Participants responded both vocally and via a button box but, most importantly, the critical

FIGURE 5.3 Sample visual scene employed in a passive listening visual world experiment. Participants would hear a recorded stimulus such as "The boy will eat the cake," and indicate whether the statement was consistent with the image via both a verbal yes/no response and a button press.

eye movements were those produced while they listened to the linguistic stimuli (i.e., before a response was made).

There are advantages and disadvantages to each approach but, broadly speaking, action-based tasks produce cleaner data (i.e., with less variation) because all participants are instructed to carry out the same task and therefore their eye movements follow a similar trajectory. During passive listening, participants are inspecting a context-rich visual scene that is closely linked to the linguistic stimuli that they are hearing. Because there are differences in the way in which individuals inspect visual scenes, data are more variable and may require more trials and/or participants in order to perform statistical analyses (Altmann, 2011b).

History of the Method

Cooper (1974) is widely cited as the first scholar advocating for the use of speech and a visual field containing objects semantically related to the speech signal to study real-time perceptual and cognitive processes. However, it was not until the

publication of Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) that researchers took notice of the strong link between eye movements and comprehension. In that study, Tanenhaus et al. presented participants with one of two visual scenes, each containing four objects—for example, a towel, a towel with an apple on it, a pencil, and an empty box. Participants were asked to carry out a simple task by following an auditory instruction such as "*Put the apple on the towel in the box*." The auditory instruction is syntactically ambiguous at the point participants hear "towel" because it could be interpreted as the goal (i.e., move the apple to the empty towel) or as a modifier (i.e., move the apple that is on the towel to a to-be-named location). The ambiguity is resolved at the moment that participants hear the actual goal, "box." While listening to experimental instructions, participants' eye movements were recorded by way of a head-mounted eye-tracker. Tanenhaus et al. found that upon encountering the region of ambiguity, participants' eye movements toward the incorrect target location (e.g., the empty towel) increased. In other words, participants' eye movements reflected the local syntactic ambiguity, thereby confirming that eye movements are closely time-locked to the unfolding auditory signal. Tanenhaus et al. compared eye movements of this so-called *one-referent* scene to eye movements to a scene that included two apples, one on a towel, as in the previous scene, and one on a napkin, thus creating a *two-referent* scene. When participants were given the same instructions, looks to the empty towel (i.e., the incorrect target location) were significantly reduced and looks to the potential target location (i.e., the empty box) increased as compared to the one-referent scene. Thus, participants were more likely to interpret the towel as a modifier NP in the two-referent scene and subsequently anticipate the probable goal location. Tanenhaus et al. interpreted this finding as suggesting that participants were able to integrate contextual information (i.e., how many referents were present in a visual scene) as a relevant cue to modulate syntactic ambiguity. This at-the-time novel approach to understanding the spoken language processing of syntactic ambiguity proved highly fruitful in informing debates on syntactic modularity (Huettig et al., 2011).

The last decade has seen an impressive growth of experimental approaches using the visual world paradigm. This experimental paradigm has successfully been used to answer research questions related to virtually any area of monolingual language comprehension, including studies of subphonetic variation (e.g., Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough, 2007); phonological and phonetic processing, including effects of word frequency, cohort density, neighborhood density, lexical stress, and voice onset time (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; McMurray, Tanenhaus, & Aslin, 2002; Reinisch, Jesse, & McQueen, 2010); the influence of semantic and syntactic context on the activation of word meanings (e.g., Dahan & Tanenhaus, 2004); morphosyntactic processing (e.g., Lew-Williams & Fernald, 2007); early influence of multiple contextual cues during sentence processing (e.g., Chambers, Tanenhaus, & Magnuson, 2004; Snedeker & Trueswell, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, &

Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999); predictive processing and event representation (e.g., Altmann & Kamide, 1999; Altmann & Kamide, 2009; Kamide, Altmann, & Haywood, 2003); pragmatic inferencing (e.g., Engelhardt, Bailey, & Ferreira, 2006; Sedivy, Tanenhaus, Chambers, & Carlson, 1999); and linguistic relativity (e.g., Huettig, Chen, Bowerman, & Majid, 2010; Papafragou, Hulbert, & Trueswell, 2008). More recently, researchers have been able to extend the use of the visual world paradigm into the realm of language production and message generation to answer questions about temporal links between eye movements and speech planning (Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000). The paradigm has also been used in research with nontraditional populations such as young children and individuals with aphasia (e.g., Trueswell et al., 1999; Snedeker & Trueswell, 2004, for children; Yee, Blumstein, & Sedivy, 2008; Thompson & Choy, 2009; Mirman, Yee, Blumstein, & Magnuson, 2011, for aphasic populations).

To our knowledge, the first study to use the visual world paradigm to investigate second language processing was Spivey and Marian (1999). The study examined the parallel activation of words in both the native and second language when participants heard a word in one language only (for related work, see Ju & Luce, 2004; Marian & Spivey, 2003a, 2003b; Weber & Cutler, 2004). L1 Russian, L2 English participants sat in front of a display that contained four real objects. The critical manipulation was the presence of an object in the visual scene which shared word onset between the two languages. For example, participants heard the Russian instruction *Poloji marku nije krestika* "Put the stamp below the cross", while viewing a display containing a stamp (the target), a marker (which bares phonetic similarity with *marku*—the Russian word for stamp), a keychain, and a quarter. The findings showed that participants made eye movements to the between-language phonological competitor (e.g., marker), suggesting that lexical items in both languages were activated simultaneously, despite the fact that only one language was being used during the experimental session. Furthermore, the results provided evidence that even a second language learned in adulthood can influence spoken language comprehension in the native language.

Compared to the study of native/monolingual language comprehension, relatively little work has been conducted using the visual world paradigm within the field of SLA (Blumenfeld & Marian, 2005; Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Hopp, 2012; Ju & Luce, 2004; Lew-Williams & Fernald, 2007, 2010; Marian & Spivey, 2003a, 2003b; Perrotti, 2012; Spivey & Marian, 1999; Weber & Cutler, 2004, Weber & Paris, 2004). The paucity of SLA studies using the method is attributable in part to the relative infancy of the visual world paradigm relative to other experimental methods used in SLA research. However, the method's wide applicability is particularly appealing to SLA research. Specifically, the use of eye movements as an index of comprehension is a covert dependent measure with high ecological validity (i.e., it does not rely on overt responses such as button responses or linguistic judgments as the dependent

measure). Its focus on spoken language comprehension (and production) with visual nonlinguistic input circumvents the need to depend on literacy as a means to test second language processing. Hence, the visual world paradigm has the potential to broaden the traditional scope of SLA research by including typically underrepresented research populations such as immigrant language learners who may not have full command of a second language in its written domain, yet successfully communicate in a second language.

What is Looked at and Measured

The visual world combines experimental designs typically employed in spoken language comprehension studies with eye-tracking, a research tool that is becoming increasingly ubiquitous at research institutions. The previous chapter introduced the different states associated with eye movements. To briefly summarize, although our experience as viewers suggests that eye movements proceed smoothly as we examine a visual scene, we are in fact engaging in a series of short, ballistic movements known as *saccades.* These saccades occur between moments of measurable stability known as *fixations.* Of course, individuals also engage in *blinks* during any given period of time. As in the case of eye-tracking studies that involve the reading of text, in visual world studies researchers are also interested in the fixations. Some researchers, particularly Altmann and colleagues, additionally argue that saccades should be measured and analyzed (Altmann 2011a, b); however, this view is a minority position within the field. Because visual inspection of nonlinguistic stimuli is not as "linear" as in reading studies, in visual world studies researchers are more interested in the fixations that occur in aggregate over a timescale rather than in the different kinds of fixations that may occur over time (e.g., *first-pass fixation*, *regression*, *total fixation*; see Keating, Chapter 4, this volume).

To determine whether participants fixated on a particular object in a visual world study, *regions of interest* need to be defined amongst the different objects displayed in the visual scene. Generally speaking, in experiments that make use of experimental eye-trackers and visual scenes presented on a computer monitor, regions of interest are predetermined by way of the tools included in the experimental software. To illustrate, if a scene consists of a four-picture display of line-drawing objects (e.g., a canoe, a baby, a flower, and a ladder), the regions of interest would be defined by drawing geometric shapes around each object using the experimental software. It is important to note that these regions of interest are invisible to the participant. Because participants do not always directly fixate on the center of objects, regions of interest are drawn larger than the image itself. Researchers differ on how large they may draw the region of interest. One common practice is to draw square boxes that contain the entire image in use. Thus, for the image of the *canoe* in Figure 5.2, the region of interest would include the picture of the canoe as well as the surrounding white space. Another approach is to split the monitor into a grid and define each region as a quadrant (as in the case

of a four-picture display) in the grid. Of these two stances, we lean more towards the first approach, as it allows researchers to label fixations that occur at a distance from the objects as "outside looks." These outside looks can capture very broadly the variability that may exist in the data (explained in greater detail below). Other researchers, however, are more conservative and draw the regions of interest to match the shape of the object itself (i.e., having a free-form canoe-like region drawn around the canoe).

Visual world designs that use moving video images or that make use of commercial video cameras present some unique challenges for determining where a participant is fixating. For video images, the issue is that the region of interest itself moves dynamically during the video. Some experimental software comes with tools that can define dynamic regions of interest; however, they track moving images with varying success and thus may necessitate labor-intensive check-ups on the part of the researcher to ensure that the region of interest is well-defined. For researchers who rely upon commercial video cameras, typically the camera is trained upon the participants' eyes and, therefore, the data coding relies more upon which direction the eye is looking. It is important for the participant to be seated at a fixed distance from the visual scene and for the distance between objects to be fairly large in order to increase the likelihood that an eye movement be associated with any particular region of interest. Absent any custom software, researchers must examine video recordings of these eye movements frame by frame (with a standard frame refresh rate of 33 ms) in order to manually code in which direction the eye is looking. This approach is highly labor-intensive and is best suited for large research teams in order to provide the necessary resources to code such data and to provide inter-rater reliability in the data coding.

Perhaps one of the most challenging aspects of a visual world study is understanding how the raw data extracted from experimental software or manually coded is then transformed into proportional data that is plotted as curvilinear time-course plots. Fixations at any given time and in any given region are binary and exclusive: either there is a fixation (1) or there is not (0), and if there is a fixation in Region A, then there cannot be a fixation in Region B. Thus, an individual's fixation record for any object in one experimental trial consists of a series of 1s and 0s for long stretches. Researchers differ in the exact protocol that they follow to extract eye-tracking data (largely dependent on the experimental software available to them). In Table 5.1, we illustrate a typical raw data file from our lab, generated by DataViewer, a data-extraction software program provided as a part of the Eyelink system (SR Research).

At the moment of data extraction, a *sample report* is created containing the region of interest in which the eye has been tracked, and whether the eye is in a state of blinking or saccadic movement. If the eye is not blinking or launching a saccade, then it is counted as *in fixation*. The spreadsheet in Table 5.1 includes two main types of variables: those that identify the data and those that constitute measurements. The variables RECORDING SESSION LABEL, CONDITION,

TABLE 5.1 Sample spreadsheet of extracted eye-tracking data before conversion to proportion data

Recording session label	*Sample message*	*Cond*	*Timestamp*	*Trial label*	*Left interest area id*	*Left in blink*	*Left in saccade*	*Response*	*File*
Part01	SOUND _WORD ONSET	2	15477730	Trial: 1	1	0	0	incorrect	1
Part01	.	2	15477732	Trial: 1	1	0	0	incorrect	1
Part01	.	2	15477734	Trial: 1	1	0	0	incorrect	1
Part01	.	2	15477736	Trial: 1	1	0	0	incorrect	1
Part01	.	2	15477738	Trial: 1	1	0	0	incorrect	1
...	...	...	...	...	...	...	...	...	...
Part01	.	4	15559754	Trial: 10	1	0	0	correct	1
Part01	.	4	15559756	Trial: 10	1	0	0	correct	1
Part01	.	4	15559758	Trial: 10	1	0	0	correct	1
Part01	.	4	15559760	Trial: 10	1	0	1	correct	1
Part01	.	4	15559762	Trial: 10	1	0	1	correct	1
...	...	...	...	...	...	...	...	...	...
Part05	SOUND _WORD ONSET	1	21629970	Trial: 53	2	0	0	correct	2
Part05	.	1	21629972	Trial: 53	2	0	0	correct	2
Part05	.	1	21629974	Trial: 53	2	0	0	correct	2
Part05	.	1	21629976	Trial: 53	2	0	0	correct	2
Part05	.	1	21629978	Trial: 53	2	0	0	correct	2

TIMESTAMP, TRIAL LABEL, and FILE are identifying variables. Other variables are the measurement variables and include LEFT INTEREST AREA ID, LEFT IN BLINK, LEFT IN SACCADE, and RESPONSE. The measurement variables are composed of three columns related specifically to eye-tracking data (all of those beginning with LEFT) and one column, RESPONSE, associated with the behavioral response to a secondary task (i.e., clicking on a target item with a computer mouse). SAMPLE MESSAGE (column 2) represents the point in the recording session at which the data was extracted. During the programming of a visual world experiment, the experimenter must flag the specific moment (in milliseconds) in each sound file that marks the beginning of the critical region, which in this example was a noun onset. Thus the onset of each trial should simply state that the sample message was SOUND_WORDONSET. This column is normally included as a means to verify that the data have been extracted from the appropriate noun onset in each trial. Its inclusion in the sample report is not strictly necessary, but it is a useful step to ensure correct data extraction.

Returning to the sample data points in Table 5.1, the first five rows (labeled Part01) show data for Participant 1 (RECORDING SESSION LABEL) on Trial 1 (TRIAL LABEL) corresponding to a trial from Condition 2 (CONDITION). Turning to the specific eye-tracking columns, the participant's eye is tracked in Region 1 (LEFT INTEREST AREA ID) and is not blinking or in saccadic movement—as indicated by all 0 values in both LEFT IN BLINK and LEFT IN SACCADE. However, note that in the RESPONSE column, the trial is marked as "incorrect." This response indicates that the participant has clicked on the wrong item for this trial. Following established practice for unimpaired, native language participants, this trial would typically be excluded from the data analysis. The second set of five sample data points also comes from Participant 1 but now from Trial 10. This trial is identified as a trial representing Condition 4. Here, the participant's eye is tracked in Region 1, and the participant has identified the correct target item, thus the data will be included for data analysis. Note that the last two rows of this sample set indicate that the participant began to launch a saccade (indicated by the 1 value found in LEFT IN SACCADE. Therefore, these rows will not be counted as fixations. Finally, in the last set of five sample data points, the data come from Trial 53 from a different participant, Participant 5. This trial is another trial labeled "Condition 1." Here, the participant's eye is tracked in Region 2; none of the rows reveal any blinks or saccades and the participant has clicked on the correct item, so all of the data points will be included.

How do these binary values get converted into proportional data? Simply put, the conversion happens as an aggregate of all binary values for a given time point for each experimental condition. To illustrate, let's use a simple two-picture display, a design which results in two predefined regions of interest—a target region of interest and a distractor region of interest. According to our lab protocol, this setup further results in a third region for fixations that fall *outside* either region of interest. Let's suppose that there are 10 trials per condition and

the data presented corresponds to the raw data file for Participant 1. In Condition A, at time 0 (onset of the temporal region of interest), the eye-tracking device records three fixations in the target region, four fixations in the distractor region, and three fixations falling outside either region. Then for this single time point, proportions are determined by dividing the number of actual fixations in any given region out of the total number of fixations observed. Thus, for the target region, the proportion of fixations is 0.3 (three fixations in the target region divided by 10 observed fixations); for the distractor region, the proportion of fixations is 0.4 (four fixations in the distractor region divided by 10 observed fixations); and for fixations to the outside region, the proportion of fixations is 0.3 (three fixations to neither target nor distractor regions divided by 10 observed fixations). Recall, however, that the eye may be in saccadic movement or blinking. These states would also be coded in the raw data file, which may result in less fixations for any given time point than the total number of possible fixations. To continue with our example, let's assume that in the same condition but at time 200 ms, Participant 1 had four fixations and two saccades in the target region, two2 fixations in the distractor region, one fixation in the outside region, and one blink on a trial. Following our lab protocol, the corresponding proportions would be calculated out of a total of seven fixations, resulting in the following proportions: 4/7 = 0.571 for the target region, 2/7 = 0.286 for the distractor region, and 1/7 = 0.143 for fixations outside of either region. Other labs may continue to calculate the proportions out of 10 *possible* fixations. Regardless of the method to determine proportions used, it is important to remain consistent throughout all calculations.

The description above should give a sense of the sheer amount of data that is being processed in a visual world experiment. Because of this, calculating proportions of fixations manually is not efficient. Depending on the experimental software, it is possible to extract fixations directly onto an Excel spreadsheet (or a text file which can later be opened in Excel). With some basic knowledge of Excel, *macros* can be created to calculate proportional data. One drawback of this approach is that with very large data sets, Excel can become slow to respond, may crash frequently, or may not have enough rows to accommodate the entire data set. Under these circumstances, the most convenient way of performing these calculations is via the use of R (R Development Core Team, 2008), an open-source statistical software package. Learning to use R can require a steep learning curve, as actions are carried out through command-line prompts. Despite this, the program has many benefits, as it can be used to calculate proportions, to generate time-course plots, and to perform statistical tests on the data. With sufficient expertise, a lab assistant can write scripts that are specific to any given experimental design, which make proportion calculation and data visualization easy. An adequate description of how to create these scripts is beyond the scope of this chapter, but Table 5.2 shows a sample subset of what a proportional data file looks like using R.

TABLE 5.2 Sample spreadsheet of proportional data aggregated over conditions and binned into 20 ms time bins

Subject	*Condition*	*Time*	*Prop target*	*Prop distractor*	*Prop nothing*
1	1	20	0.215053763	0.430107527	0.35483871
1	1	40	0.204081633	0.489795918	0.306122449
1	1	60	0.21978022	0.43956044	0.340659341
1	1	80	0.210526316	0.421052632	0.368421053
1	1	100	0.222222222	0.444444444	0.333333333
...	...	...	...	...	...
24	2	920	0.942408377	0.031413613	0.02617801
24	2	940	0.949238579	0	0.050761421
24	2	960	0.95	0	0.05
24	2	980	0.95	0	0.05
24	2	1000	0.95	0	0.05

The data file in Table 5.2 consists of three columns that identify each row: participant, condition, and time. These columns are followed by the proportional values. In our sample case, there are three columns containing proportional data, one for the proportion of fixations to target items (Prop Target), one for distractor items (Prop Distractor), and one for the proportion of fixations that fell outside of either region (Prop Nothing). Conceivably, the data could be arranged where the dependent measure (i.e., the proportional values) are all contained in one column, so long as a second column identifies the region of the proportional value. Notice that the timescale (Time) is in increments of 20 ms. This sample subset was extracted from an eye-tracker with a 500 Hz sampling rate. As discussed earlier, this sampling rate produces a data point every 2 ms. To make the data files more manageable and to prevent complications in the time-course plots caused by overlapping data points, data have been aggregated into 20 ms time bins. This is accomplished by taking the mean of proportional values that fall within each 20 ms time bin. Thus, in Table 5.2, the first row of data represents the mean of the first 10 proportions per region of interest.

Whereas the original Tanenhaus et al. (1995) study calculated the proportion of total trials on which participants looked at one region versus another plotted as simple bar graphs, contemporary studies also include time-course information. In time-course plots, total proportion of fixations over trials and participants are calculated and plotted over a millisecond timescale. Figure 5.4 is a sample graph of what a typical time-course plot looks like. The sample data are taken from a pilot study where a group of L2 English speakers whose native language was Spanish were asked to listen to variable sentences that named one of two objects presented on a computer screen (e.g., a visual display that showed a picture of a hammer on the left and a mug on the right while participants heard "*The man told his daughter to photograph the hammer on the table*"). The y-axis represents the total proportion of

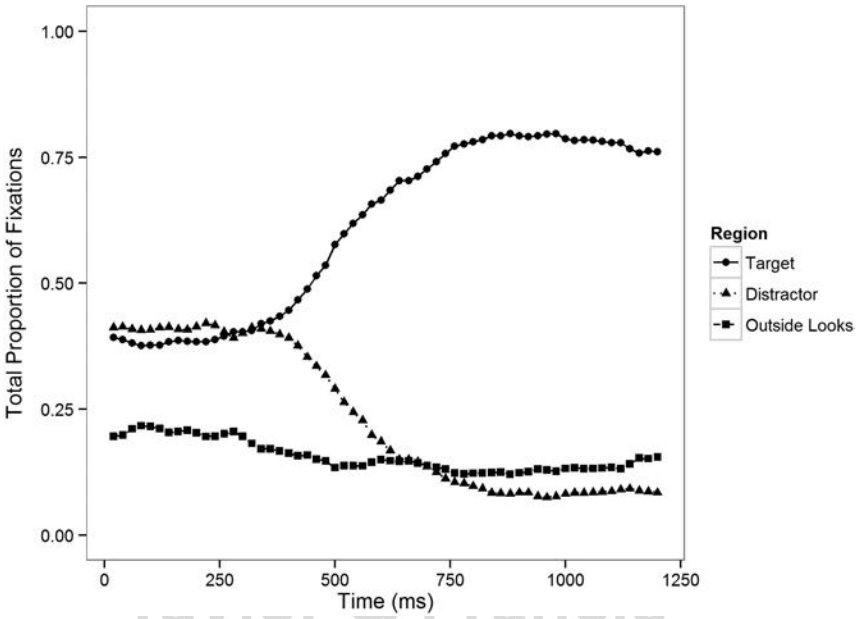

FIGURE 5.4 Sample time-course plot from a visual world experiment.

fixations; the x-axis plots time in milliseconds. Because the data are proportional, the y-axis is by definition bounded between the values of 0 and 1. Time at x = 0 typically represents the onset of a target region of interest. For this sample data, the region of interest starts at the onset of the noun pictured in the display (e.g., when participants hear "*hammer*").

This sample pilot study used a two-picture display; hence there is only one distractor. Many designs use four-picture displays; thus, a graph may show more lines than those shown in the sample graph. However, even in four-picture displays, some researchers simply aggregate distractors that are not important to the particular research question. In the sample plot, the fixations to the target noun are plotted as solid circles, fixations to the distractor noun are plotted as solid triangles, and any fixations that fell outside of either region are plotted as solid squares. From this plot, we can determine that fixations to both the target and distractor were equally prevalent until approximately 400 ms after the noun onset, where an increasing number of participants fixated on the target item. Fixations to target items continue to increase after that. One point to note from the sample plot is that fixations to target items never reach a proportion of 1, even if all participants fixate on the target item at some point during the region of interest. This somewhat counterintuitive observation results because individuals are idiosyncratic in how they will view a visual scene. Some people are faster to look at named objects than others; some individuals may be attracted to some sort of visual feature that

is irrelevant to the linguistic information presented. Nevertheless, the time-course plot shows that a majority of individuals looked at the target item at some point while it was named in a majority of the trials. Although not necessarily standard in published time-course plots, it is useful to plot fixations to outside regions as a means to observe whether a majority of participants in fact fixate on any displayed objects. Here, the line that corresponds to fixations outside of the two pictures remains relatively low and stable, never going above a proportion of 0.2. This line thus represents a random factor—that is, individuals at any given point may be blinking, transitioning to another picture, or examining the visual scene in its entirety. However, if the proportion of fixations to outside regions were high and not stable, then the plot would indicate a more serious problem, such as individuals who strategically did not look at either object (explained in more detail below), a high amount of variability amongst all individuals, or even a failure to detect the eye on the part of the eye-tracking device.

Researchers using the visual world paradigm look for the presence (or absence) of *competitor* and *anticipatory* effects. Broadly, these effects are taken to reflect delayed or facilitated processing, respectively. In the illustrative pilot study introduced above, the experimental condition contained a two-picture display with phonological competitors. For example, the target item, a hammer, was paired with a picture of a hammock (instead of the picture of a mug). Both items overlap in phonology in the first syllable /hæm/. When compared to items that do not compete phonologically, such as when a hammer is displayed with a picture of a mug, studies have shown that participants take longer to identify target items (Allopenna et al., 1998). Compare the previous Figure 5.4 with Figure 5.5 (the legend is the same). Whereas Figure 5.4 shows clear divergence between target and distractor items roughly around 400 ms, in Figure 5.5, participants are showing consideration of the distractor item (i.e., the phonological competitor) much later in the time-course. Although some separation appears to happen between 400 ms and 600 ms, we do not see a reliable increase in divergence between target and distractor until after 600 ms. This later divergence is the *competitor* effect.

In contrast to the competitor effect, an effect is said to be anticipatory when eye movements to a target object are launched significantly before the presentation of the linguistic input that is predicted to initiate looks to that target. Suppose, for example, that we want to find out whether participants listening to Spanish premodifiers marked for gender (e.g., the definite article *el* and *la*) use gender information to facilitate the processing of upcoming nouns (e.g., Lew-Williams & Fernald, 2007). The experimental setup would consist of some trials in which a masculine-gendered object (micrófono/microphone$_{MASC}$) is presented alongside a feminine-gendered object (vela/candle$_{FEM}$). These are the different-gender trials. Proportion of looks to the target in different-gender trials are compared to proportion of looks to the target in same-gender trials—trials consisting of two same-gendered objects presented alongside one another (e.g., micrófono/microphone$_{MASC}$ presented next to zapato/shoe$_{MASC}$). An anticipatory effect occurs

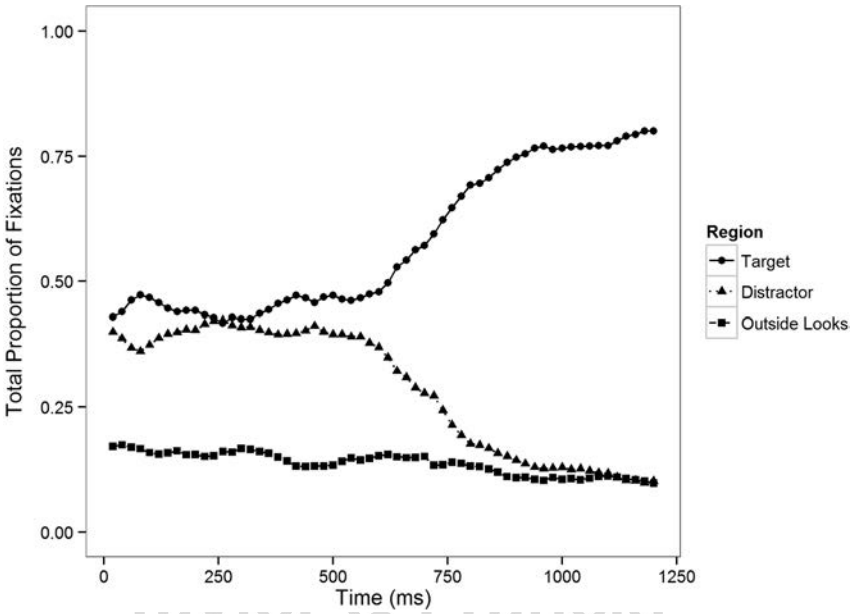

FIGURE 5.5 Sample time-course plot showing a competitor effect.

if after hearing *Encuentra la vela* "Find the candle," participants orient their eyes towards the target object *vela* more quickly on different-gender trials (i.e., when the gender information encoded in the article is informative) than on same gender trials (i.e., when participants need to wait to hear the named object before clicking on it). Figure 5.6 illustrates a time-course plot for an anticipatory effect. When comparing the panel on the top (Spanish Monolinguals, Feminine Different) to the panel on the bottom (Spanish Monolinguals, Feminine Same), we see that looks to targets on different-gender trials occur significantly before looks to targets on same-gender trials.

Although it is plausible to assume that competitor and anticipatory effects are opposite effects, in fact they are not. They are effects that can only be determined relative to a neutral baseline. For example, in the case of the Spanish grammatical gender described, if the gender information encoded in the article is not informative, the basic task is one of target word identification. This is precisely what goes on in the same-gender trials, and therefore, these trials constitute the *neutral baseline*. The effect of interest is whether the gender information present on the definite article will affect the time-course of a target relative to the neutral baseline. As described above, in the case of Spanish, it does and does so by quickening the time-course, hence the presence of an anticipatory effect.

A final potentially confusing point is the means by which researchers determine whether a competitor or anticipatory effect is present. In the Allopenna et al.

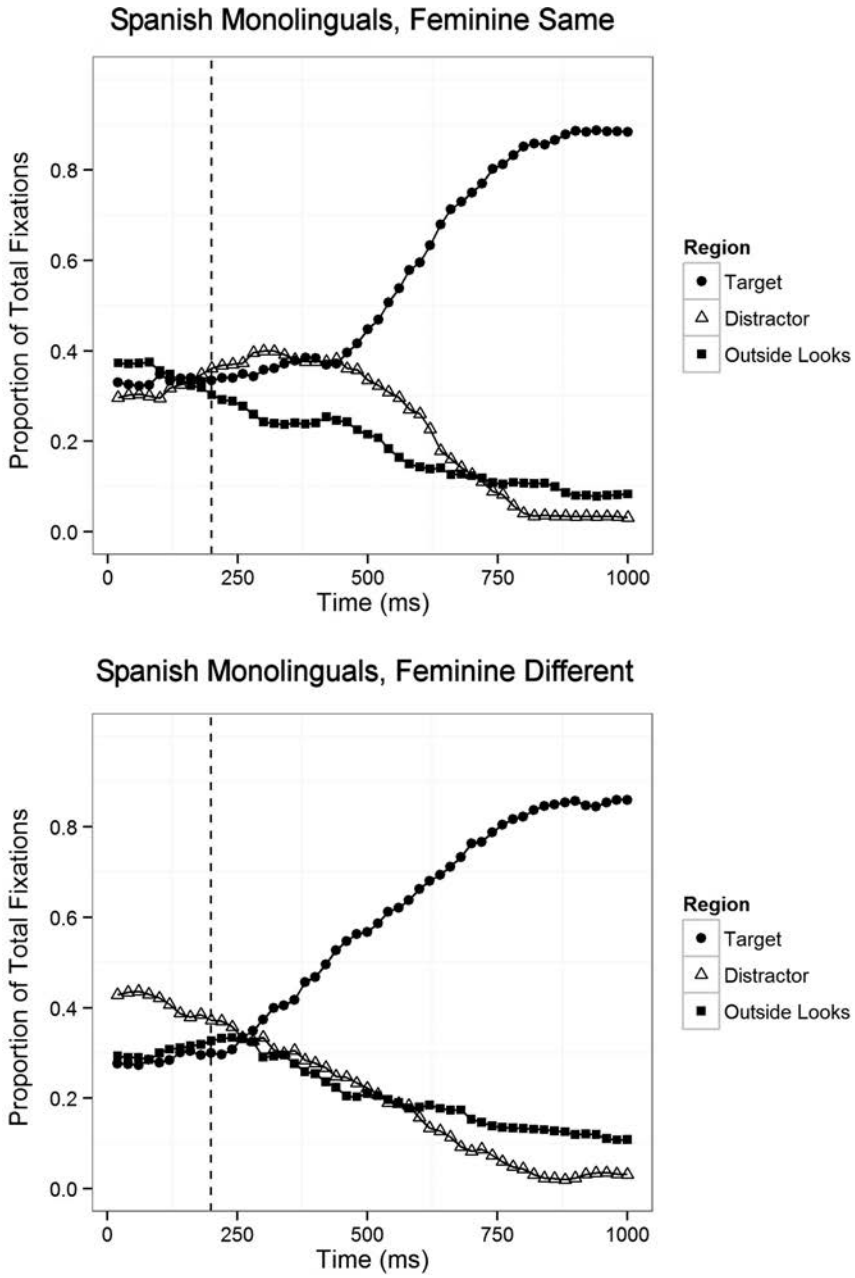

FIGURE 5.6 Sample time-course plots showing an anticipatory effect.

(1998) study, a competitor effect was determined relative to the cohort distractors that were copresent in the same visual scene. That is, the time-course of fixations to the target items was compared to the time-course of fixations to a phonological cohort, a rhyme cohort, and a nonphonological control. A competitor effect was determined by statistically comparing the distractor proportion of fixations to the proportion of fixations of the target item. In the case of the phonological cohort distractor (e.g., *beetle* for target word *beaker*), the proportion of fixations to the two items was not statistically different until roughly around 400 ms from the onset of the target noun. Nevertheless, the overall time-course for the target item was ultimately different from the time-course plot of all other distractor candidates, which also looked different from each other, meaning they all affected spoken word recognition in different ways. In contrast, Lew-Williams and Fernald (2007) compared the time-course of proportion of fixations only to target items but in separate conditions. In other words, an anticipatory effect was determined because a shift in a critical mass of looks to the target item happened faster in the different gender trials than in same gender trials. However, both trials had a similar time-course plot overall.

Visual World Studies and SLA

A few studies are beginning to emerge which use the visual world paradigm to ask longstanding questions in second language acquisition regarding the degree to which adult second language speakers recruit different types of information during real-time L2 comprehension. A recent study by Lew-Williams and Fernald (2010; see also Hopp, 2012) investigated the integration of L2-specific morphosyntactic information during on-line processing by asking whether adult L2 speakers of Spanish use grammatical gender encoded in definite articles to facilitate the processing of upcoming nouns. In a series of experiments modeled after Lew-Williams and Fernald (2007), they presented L2 learners of Spanish (L1 English) with two-picture visual scenes in which the pictured objects either matched or differed in gender. Participants heard instructions that asked them to find an object. In three experiments, they showed that when listening to sentences naming both familiar and newly-learned objects, native speakers were able to orient their eyes towards target objects more quickly on different gender trials (i.e., when the gender information in the article was informative) than on same gender trials, showing an anticipatory effect. L2 speakers of Spanish, on the other hand, waited to hear the noun to initiate a gaze shift. These findings suggested that non-native listeners were not able to integrate abstract gender information during on-line processing the way that native speakers did.

Other studies have examined transfer effects from the L1 to the L2. For example, Weber and Paris (2004) investigated whether the gender of words in the L1 exerts an effect on the recognition of words spoken in the L2. In this study, the eye movements of L1 French speakers, who had acquired German as a second

language during adulthood, were monitored while they heard spoken instructions in German to click on pictures displayed on a computer screen (e.g., *Wo befindet sich die Perle* "Where is the pearl?"). The critical manipulation included target and distractor pictures in German that either shared phonological onset and grammatical gender with French words, or that shared phonological onset with French words but differed in grammatical gender. For example, in the same-gender pairs, the feminine German target *perle* "pearl" (*perle* in French, also feminine) was paired with the feminine German distractor *perücke* "wig" (*perruque* in French, also feminine). In the different-gender pairs, however, the feminine German target *kassette* "cassette" (cassette in French, also feminine) was paired with the feminine German distractor *kanone* "cannon" which was masculine in French (*canon*). Weber and Paris found that when target and competitor pairs were the same gender in German but were different gender in French (i.e., different-gender pairs), fixations to the competitor picture were reduced compared to when target and competitor pairs were the same gender in both German and French. What these results suggest is that for L1 French speakers who had acquired German as adults, the grammatical gender of French nouns modulated the processing of determiner + noun combinations in their L2, German.

The visual world paradigm has also been used to answer critical questions about the brain's ability to accommodate multiple languages, lexical activation and competition, and mechanisms of language (non)selectivity (see, for example, Blumenfeld & Marian, 2005; Canseco-Gonzalez, Brehm, Brick, Brown-Schmidt, Fischer, & Wagner, 2010; Cutler, Weber, & Otake, 2006; Ju & Luce, 2004; Marian & Spivey, 2003a, 2003b; Spivey & Marian, 1999; Weber & Cutler, 2004). These studies are not discussed here because they are not centrally related to current issues in SLA.

Issues in the Development and Presentation of Stimuli

A number of important decisions need to be made when designing visual world experiments. The first deals with the selection of the visual display. Depending on the research question, displays can vary, consisting of black-and-white line drawings or colored pictures of concrete objects displayed on a computer screen (e.g., Allopenna et al., 1998; Perrotti, 2012; Weber & Paris, 2004), arrays of objects laid out in a work space (e.g., Spivey & Marian, 1999; Snedeker & Trueswell, 2004), or line drawings or colored drawing of semirealistic scenes (Altmann & Kamide, 1999; Griffin & Bock, 2000; Arnold & Griffin, 2007).

The obvious advantage of black-and-white line drawings is that there are large repositories of pictures that have been normed for naming agreement, familiarly, and visual complexity with native-speaking children and adults as well as with some groups of L2 learners. Among these are the pictures employed in the *Boston Naming Test* (Kaplan, Goodglass, Weintraub, & Segal, 1983). The test contains 60 line drawings, graded in difficulty from easy and high frequency (e.g., *bed*) to

difficult and low frequency (e.g., *abacus*), which have been normed in English and Spanish. Three other popular sources are the set normed for adults by Snodgrass and Vanderwart (1980), which contains 260 pictures normed for name agreement, image agreement, familiarity, and visual complexity, the English version of the *Peabody Picture Vocabulary Test* (Dunn & Dunn, 1997) and its Spanish equivalent, the *Test de Vocabulario en Imágenes Peabody* (TVIP; Dunn, Dunn, & Arribas, 2006). One source of line drawings that has proven to be extremely useful is the *International Picture Naming Project* (Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E., 2004; available at http://crl.ucsd.edu/~aszekely/ipnp/). The website provides access to 520 black-and-white drawings of common objects and 275 concrete transitive and intransitive actions. A special feature of these drawings is that they have been normed in seven languages (English, German, Mexican Spanish, Italian, Bulgarian, Hungarian, and Mandarin) and with children and adults.

In the context of visual world experiments, the most significant drawback with black-and-white line drawings is that it can sometimes prove to be difficult to find pictures with the desired linguistic characteristics across a bilingual's two languages. It is because of this that many bilingualism researchers resort to other sources. Colored line drawings and photographs of real objects are the most popular. Colored line drawings are available through many sources, including IMSI MasterClips Image Collection (IMSI, 1990), and pictures of real objects abound over Google Images. One advantage of color images is that recognition of the objects is enhanced by color and texture information present in the visual display. Therefore, norming for naming agreement is greatly facilitated. A second advantage is that color images can be manually manipulated using Microsoft Paint (or a similar paint tool) to remove distracting patterns, to crop an image such that the target item is centered or to adjust the size of an image. One potential disadvantage is that visual complexity varies greatly among pictures, which can influence eye gaze.

A second important issue to consider is the position of the objects in the visual scene. Readers of left-to-right languages like English and French have a left-gaze bias (i.e., when presented with a visual scene on a computer screen, they will first direct their gaze to the upper left corner). Therefore, it is important to counterbalance the position of each item on the visual display. For example, if in one presentation list a target picture appears on the left side position and a distractor picture on the right side position of the computer screen, then a separate experimental list should have these positions reversed. It is important to note that counterbalancing of this sort results in double the number of experimental lists.

Studies have shown that eye movements generated during the recognition of a spoken word are mediated by a number of factors. It is important to be aware of these to avoid confounds in the experimental design. We have already mentioned phonological overlap as one factor (Allopenna et al., 1998). Eye movements to a display containing the picture of a *beetle* (target), a *beaker* (phonological

competitor), a *speaker* (a rhyme competitor) and a *carriage* (unrelated word) generate more eye movements to the phonological competitor upon hearing the referent's name than to either of the distractor pictures in the visual display. Importantly, phonological competition is modulated by frequency effects. Dahan et al. (2001) found that when the picture of a referent such as *bench* (the target picture) was presented along with a high frequency phonological competitor (*bed*) and a low frequency phonological competitor (*bell*), a participant's early eye movements were equally likely to the target word and to the high frequency phonological competitor. The authors also found that the frequency effects during the earliest moments of lexical access could be obtained even when none of the competitors were phonologically similar to the target. Fixation latencies to targets with high-frequency names (*horse)* were shorter than those to targets with low-frequency names (*horn*). Structural similarity of the pictures is another variable. Dahan and Tanenhaus (2005) showed that when participants heard the words *snake* being spoken in the presence of a display containing a snake (target), the picture of a *rope* (visual competitor), an *umbrella,* and a sofa (two distractor objects), participants were more likely to fixate the visual competitor *rope* than either of the distractor objects.

One last issue that requires careful attention is the preparation of sound files. Because eye movements to the objects are closely time-locked to the unfolding speech signal, spoken instructions need to be carefully controlled at the critical region of interest in order to facilitate data analysis. To illustrate, the creation of sound files similar to those employed in the grammatical gender processing studies described earlier would require a number of steps. First, a speaker is asked to record several versions of a simple invariant carrier phrase (e.g., *Encuentra la/ Encuentra el* "Find the" masculine/feminine). Typically, the speaker repeats each carrier phrase in a normal declarative intonation between five to ten times. From these, measurements of the duration of each definite article are taken and an average length for each article is calculated. Subsequently, the best carrier phrase is selected and the duration of the article within each carrier phrase is matched, using Praat or other similar type of software (e.g., the duration of the definite articles *el* and *la* are matched so that each is, say, 200 ms long). Finally, the same speaker is asked to name each experimental stimulus five times. From each set of repeated tokens, the best exemplar is selected to be inserted into the appropriate carrier phrase. This procedure avoids the presence of coarticulation information in the article, which is known to influence noun recognition.

Scoring, Data Analysis, and Reporting Results

There is currently no consensus on how best to analyze eye-tracking data collected from visual world studies. Part of the issue is that generally the dependent measure in visual world studies is total proportion of fixations (although see e.g., Altmann, 2011a, for analyses done with saccadic measures). As a consequence, the

dependent measure is bounded between 0 and 1, unlike other dependent measures typically used in behavioral studies. Additionally, because proportional data is plotted over time, the independent measure, time, is continuous. Altmann (2011b) describes the fundamental issues surrounding analysis:

> An entirely different class of statistical modeling needs to be carried out for analysing time-course data . . . how can one determine that any pair of curves are different from one another? How can one determine where the peak is located for any such curve (given that aggregating data for the purposes of such [time-course] plots hides the true underlying distribution of the data across subjects and trials)? And most importantly, perhaps, how can one model the dynamic changes to fixation proportions across time when successive time points are not independent of one another? (Altmann, 2011b, p. 996)

These issues are all tempered by the decisions that researchers must make on the mode of presentation of the visual scene, which further impacts how the data are analyzed. In some experiments, participants are allowed free view of the visual scene prior to the target region of interest. This protocol is in contrast to other work, which requires participants to remain on a fixation point or fixation cross until the onset of the target region of interest. Both methods have their advantages and disadvantages. Allowing free view of the visual scene represents a more ecologically valid task reflective of what participants would presumably do in nonexperimental settings. Therefore, free view presentation offers an ecological advantage over fixed visual presentation. On the other hand, free view presentation aggravates one potentially problematic issue in data analysis that is attenuated in fixed visual presentations. Specifically, because participants are idiosyncratic in the manner in which they view a visual scene prior to hearing a named object, free view presentation greatly increases the likelihood for baseline effects. Briefly, baseline effects are represented on a time-course plot by the y-intercept (or value of y at $x = 0$). The greater the magnitude of difference between the y-intercept of the target and any distractors, the greater the baseline effect, which subsequently represents a random effect in eye-tracking data. We illustrate an example of a time-course plot with baseline effects in Figure 5.7. Notice how the two lines representing fixations to target items and distractor items are already separated from the beginning of the time-course. In other words, at $x = 0$ (the onset of the critical region), the proportion of fixations to distractor items is approximately 0.477 whereas it is around 0.341 for target items. There are considerably more fixations to distractor items already present from this onset. However, this difference is not related to the experimental manipulation itself. Planned eye movements occur roughly 150 to 200 ms after their initiation. Therefore, any differences that already are present at the onset of the critical region are from planned movements occurring before the experimental manipulation. These planned movements are random effects which

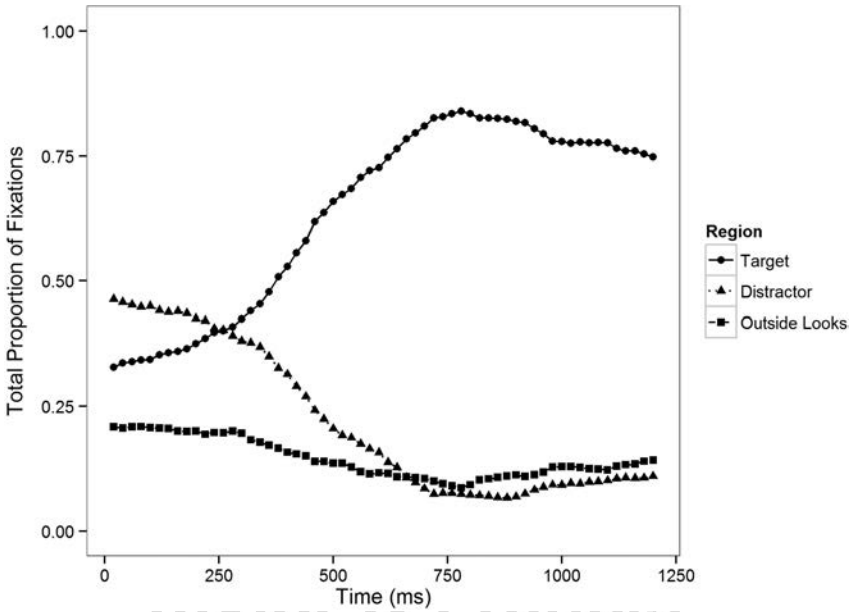

FIGURE 5.7 Sample time-course plot showing baseline effects.

are unique to the individuals and the trials. If baseline effects are very strong, then they may mask any effects driven by the experimental manipulation.

In fixed visual presentations, participants do not begin looking at the visual scene until the onset of the target region of interest; consequently, baseline effects are neutralized. In other words, all proportional data begin at 0 at the onset of the target region of interest. Although fixed visual presentations are more tenable in action-based tasks where participants are instructed in the auditory stimuli to manipulate a target item either by moving it to a new location (e.g., Tanenhaus et al., 1995) or by clicking on it with a computer mouse (e.g., Allopenna et al., 1998), the decision between a fixed visual presentation and free viewing must ultimately be made within the context of the goals of the experiment.

Researchers have implemented various approaches to analyze data which attempt to take into consideration the issue of baseline effects (see in particular special issue 59 of the Journal of Memory and Language, especially Barr, 2008, and Mirman, Dixon, & Magnuson, 2008). One such approach involves a target-contingent-based analysis (Tanenhaus, 2007). Here, researchers simply remove any trials in which participants are already on the target item at the onset of the critical region of the auditory stimulus. As in the fixed visual presentation, this technique of data trimming reduces proportional data to 0, but one disadvantage is that it can result in a high amount of data loss. In studies that employ four-picture displays, data loss may be more manageable than in experiments that employ two-picture

displays, where this method of data trimming is not suitable. Another approach to data analysis involves growth curve analysis (Mirman et al., 2008) which fits nonlinear (i.e., polynomial) models to time-course data. One major advantage to this approach is that the derived models functionally describe changes over time while preserving the original data points (i.e., there is no need to aggregate the data over time bins or over trials and participants). Second, because growth curve analysis is a regression technique, models can be hierarchical and subsequently can account for random effects such as the baseline effects described above (Mirman et al., 2008; Baayen, Davidson, & Bates, 2008). However, it remains unclear how interpretable higher-order polynomial coefficients derived from the models are. Researchers would need some level of expertise in statistics and programming to create these nonlinear regressions. Another approach employed by many visual world researchers is to follow the more traditional analyses using *t*-tests and ANOVAs. Under this approach, proportional data are aggregated over time regions and subsequent analyses are carried out within each region. Researchers look for the initial time region when the proportion of fixations to target items is significantly higher than fixations to distractors. For example, researchers can conduct paired-samples *t*-tests for each condition on fixation proportions to target and distractor items in sequential 100 ms regions from 0 ms to a predetermined amount of time (typically 800 to 1000 ms). Planned eye movements generally take about 150 to 200 ms to execute. Therefore, the earliest moment in which target stimuli could affect real-time processing would be roughly 150 to 200 ms after the target onset. Then, a paired-samples *t*-test in each region would indicate an initial moment when a significantly greater proportion of fixations occurs on target items than on distractor items. Furthermore, sequential tests would reveal whether this difference is sustained throughout the rest of the time-course. In the examples presented in Figures 5.4 and 5.5, we calculated the initial time region of divergence for fixations to target items in Figure 5.4 was at Time Region 400 (i.e., the proportion of fixations to target items was significantly higher than for distractor items). In contrast, the initial time region of divergence occurred in Time Region 600 for Figure 5.5. These results lead to the interpretation that the condition in Figure 5.5 (recall that it was phonological competition) results in delayed processing (i.e., a longer time to show significant looks to the target item) when compared to trials on which there was no phonological competition. As stated previously, this is the classic competitor effect. A final approach which we have used in our lab implements a *change point* analysis (Cudeck & Klebe, 2002), described in more detail in the Exemplary Study section below.

An Exemplary Study

Dussias, Valdés Kroff, Guzzardo Tamargo, and Gerfen (2013) employed the visual world paradigm to address two questions. First, do L1 English speakers who are highly proficient in Spanish show effects of prenominal gender marking on

the identification of subsequent Spanish nouns? Second, does the presence of a gender system in the L1 that overlaps significantly with the gender system of the L2 determine the extent to which grammatical gender processing in the L2 is native-like?

To investigate these questions, three groups of participants were recruited: functionally monolingual native speakers of Spanish (native controls) from the University of Granada (Spain), L1 English-L2 Spanish speakers from a large U.S. institution, and L1 Italian-L2 Spanish participants completing a year of university study in Granada. In a language history questionnaire, the native group reported having studied English or French in high school and none had spent over one month in a country where the second language was spoken. The 18 English-Spanish speakers were divided into two proficiency groups based on their performance on a standardized test of Spanish (Diploma de Español como Lengua Extranjera [Diploma of Spanish as a Foreign Language], DELE).

To assess knowledge of gender agreement in a comprehension task, participants completed a written picture identification task, which exploited the availability of nominal ellipsis in Spanish, to assess whether learners were able to select morpho-syntactically appropriate nouns to complete sentence fragments. To assess gender agreement in production, participants were also administered a picture naming task in which they produced article + noun fragments to pictures displayed on a computer screen. The high mean of correct responses in these two tasks suggested that gender agreement in Spanish for these participants proved largely unproblematic. Participants knew the agreement rules in Spanish and applied them with a high degree of accuracy in a production task and a comprehension task. One remaining question was whether these same participants could access this knowledge during on-line processing of grammatical gender in Spanish.

The experiment included 112 color pictures of highly familiar concrete objects. The pictures were previously normed for naming agreement. Ten participants who did not participate in the experiment were presented with a picture and were asked to name it aloud in English. Only pictures in which there was 100% naming agreement were selected. Half of the pictures represented Spanish object names with feminine gender and half with masculine gender. Each picture served as the target on two trials and as the distractor on two additional trials. Because readers of left-to-right languages have a looking bias to direct eye gaze to the left of the screen first, the presentation side of target items was counterbalanced. To investigate whether a gender-facilitatory effect occurred when participants processed rich sentence contexts (instead of invariable sentence frames such as *Encuentra la/el . . .* "Find the . . . "), we embedded the picture names in variable sentences and distributed the target items evenly so that half appeared in the middle of the sentence (e.g., for *la espada* "the sword": *El estudiante estaba dibujando la espada que vio ayer* "The student was drawing the sword that he saw yesterday") and half at the end (e.g., *El niño miraba a su hermano mientras fotografiaba la espada* "The boy watched his brother taking a picture of the sword"). To conceal

the main purpose of the experiment, after listening to each sentence, participants performed a plausibility judgment task. Half of the sentences were plausible (exemplified above) and half implausible (e.g., *El señor compró la espada para la piedra* "The man bought the sword for the rock").

A native speaker recorded each experimental sentence between three and five times at a comfortable speaking rate in a sound attenuated chamber. The sentences were produced using standard, broad-focus intonation (i.e., no narrow focus or other emphasis was produced on any of the target noun phrases). From the master recordings, one token was selected for inclusion in the experiment. To precisely match the durational properties of the masculine and feminine articles for all of the experimental items, the article preceding the target noun in each selected sentence was hand-edited to a duration of 147 ms ± 3 ms using Praat. This duration was chosen by sampling the master recordings and calculating a mean duration of the masculine and feminine articles. In this way, the duration of the acoustic signal conveying grammatical gender prior to the onset of the target noun was identical across all items.

Data analysis entailed a change point analysis, in which we implemented a multiphase mixed-effects regression model (Cudeck & Klebe, 2002; see also Section 6.4 "Regression with breakpoints" in Baayen, 2008). The basic feature of this analysis is that any number of phases, each uniquely modeled by its own function, can be united into a more complex whole (described in more detail in Cudeck & Klebe, 2002). An advantage to this approach is that it allows researchers to estimate a point in the time course (i.e., the change point) in which there is a shift between phases. The change point describes the moment in time when one rate of change switches to a different one. In terms of a visual world experiment, researchers are interested in when participants launch their eye movements in response to the linguistic input that they hear. This observation suggests that there is a moment prior to the linguistic input that subsequently changes based on that linguistic input. To that effect, we modeled a three-phase regression model, with each phase described by a linear function, and termed these phases the preconvergence phase, the convergence phase, and the postconvergence phase. The preconvergence phase corresponds to eye movements that are not directly impacted by the critical region in the auditory stimuli; rather, they include random baseline effects due to participants' free view of the visual scene and the time dedicated to launching eye movements towards target items. The convergence phase represents the period of time whereby participants' eye movements shift towards target items. Finally, the postconvergence phase corresponds to the stage in real-time processing where participants are no longer uniformly affected by the experimental stimuli. That is, participants begin to return to a random state of free view.

Experimentally, we were interested in the first change point. The first change point between the preconvergence and convergence phase indicates the point in time when a critical mass of participants begins to shift fixations to the target item. We can then compare change points across conditions. Specifically, by

conducting simple paired *t*-tests, we can determine whether one change point occurs significantly earlier (or later) than another change point. This method reveals whether an experimental condition induces a facilitatory or delayed effect when compared to a baseline condition. For the current study, we were, therefore, interested in whether the change point for different gender conditions happened significantly earlier than for same gender conditions.

As mentioned earlier, the minimum latency to plan and launch a saccade has been estimated to be approximately 200 ms (e.g., Saslow, 1967). Thus, approximately 200 ms after target onset is the earliest point at which one expects to see fixations driven by acoustic information from the target word. In line with previous findings, (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Lew-Williams & Fernald, 2007), results for the native Spanish speaker group showed evidence of the use of gender marking on articles to anticipate upcoming nouns in contexts where two pictured objects belonged to different gender classes. Analyses showed that feminine and masculine different gender trials had an earlier change point than same gender trials, indicating that Spanish monolinguals used information encoded in the article as a facilitatory cue in real-time speech. Results for the two groups of late English-Spanish learners revealed sensitivity to gender marking on Spanish articles similar to that found in native speakers, but this sensitivity was modulated by level of proficiency. The higher proficiency English-Spanish group was quicker to orient to both feminine and masculine target pictures when the article was informative (i.e., in different gender trials) than when it was not (i.e., in same gender trials). The low proficiency group did not show evidence of using grammatical gender anticipatorily, despite being highly accurate in gender assignment and gender agreement in two offline tasks. Finally, for the Italian learners, the change point in the feminine-different trials was significantly earlier than in the feminine-same gender trials. For the feminine condition, then, the results suggest that the Italian participants exploited the presence of grammatical gender on the article as a means of predicting the identity of an upcoming noun. By contrast, there was no significant difference between the change points for the same and different gender displays in the masculine article conditions, indicating that the Italian participants did not use gender as a cue in predicting the identity of a following noun when the determiner carried masculine grammatical gender.

Pros and Cons in Using the Method

Pros

- A particular advantage of the visual world paradigm is its high ecological validity. As with other eye-tracking techniques, the dependent measure in a visual world study does not require an overt behavioral response, such as a button push. Rather, visual world eye-tracking provides direct insights into the interpretation of linguistic input in real time. Specifically, given the plausible linking hypothesis, which maintains that the link between eye movements

and linguistic processing is closely timed (Tannenhaus & Trueswell 2006), the visual world method allows researchers to make inferences directly, without secondary behavioral responses, via the tracking of the locus of visual fixations as the speech signal unfolds in real time.

- A second, related benefit is that the visual world paradigm allows for the examination of auditory comprehension in a controlled, laboratory setting. Much of the research on comprehension, given historical methodological constraints, has been conducted on reading. Visual world eye-tracking affords an effective means of examining auditory comprehension, providing researchers with an invaluable tool for testing the conclusions drawn from reading studies and for incorporating prosodic cues that are necessarily absent from written presentation. Arguably, for second language learners especially, for whom reading and listening skills may be vastly different, the visual world paradigm provides an important means of directly testing processing-while-listening.
- A third advantage is that very few psycholinguistic techniques allow researchers to examine both comprehension and production using a single method. It is important to note that visual world eye-tracking can also be used to test speech production, as noted above, in designs in which participants' eye movements are tracked as they speak while viewing a visual array or scene.

Cons

- Visual world eye-tracking is costly. Although the cost of eye-trackers is coming down, the technology is expensive compared, for example, to the cost of running a self-paced reading task (see Jegerski, Chapter 2, this volume), which most often requires a PC and software for constructing the experiment. Two widely used eye-trackers for language processing research are made by SR Research and by Tobii Technology.
- One challenge in eye-tracking studies involves the selection/creation of materials. Researchers must decide on whether to employ line drawings, full pictures, or even real-life objects. One obvious limitation is that the method generally works most simply with objects that are easily pictured. In second language research with the visual world paradigm, the researcher must take pains to assure not only that the target stimuli are easily pictured but also that the participants know the names for the objects, constraints which can sometimes severely limit the set of usable target stimuli.
- Another challenge in visual world studies involves programming the experiments, which can be challenging for researchers not well versed in constructing complex experiments. In a visual world study, researchers must tightly align the timing of the presentation of auditory input with the appearance of a visual display. In addition, there are a number of different experiment

builders, each with its own programming language. To cite the example of SR Research's Eyelink, which we use in our lab, researchers are faced with the choice of programming the experiment in SR Research's proprietary Experiment Builder software or with building the experiment in E-Prime and interfacing with the tracker hardware.

- Another challenge in visual world studies involves data collection and extraction. Current setups such as SR Research's Eyelink 1000 collect eye movement data at a sampling rate of 1000 Hz. For students unfamiliar with sampling issues, a 1000 Hz sampling rate means that 1000 data points will be collected per second. Thus, a single minute of eye movement data will contain 60,000 data points. With such a large quantity of data, it becomes necessary to automate data extraction, usually something that can be achieved by writing a program in Matlab or R. This adds an additional layer of work to visual world studies.
- As is discussed above, perhaps the largest current disadvantage in carrying out a visual world study involves the lack of consensus in the field regarding how to best analyze visual world eye-tracking data. Unlike reaction time data in a button press experiment or reading time on a word (or series of words) in a self-paced reading task, both of which provide researchers with a single numerical measure, visual world eye-tracking data represent proportions of fixations to a target versus a distractor or competitor item over some unit of time. In simple terms, it is not altogether clear how to best analyze the large amount of data that such designs yield, and researchers have attempted a number of approaches, ranging from binning the data and conducting successive *t*-tests at regular temporal intervals throughout a trial to more sophisticated growth curve analyses, which essentially attempt to model response patterns by fitting polynomial equations to the aggregated eye movements across time.

Discussion and Practice

Questions

1) Give two examples of types of second language learners for whom the visual world method would be a better approach than would a reading study for examining sentence processing. Be specific in discussing your answer.
2) Think of a research question (for example, the investigation of a particular issue in the processing of a particular grammatical structure) in which a reading study would be a preferable method, and explain your choice.
3) Define the following terms: a) anticipatory effect and b) competitor effect. Discuss examples of how these effects have been exploited to inform theoretical issues in language processing.

4) Discuss the research design of a published paper in which visual world eye-tracking is employed to examine issues in bilingual language processing. Make sure to explicitly address the following aspects of the article: a) the research question that motivated the study; b) how the predictor variables were selected; c) what criteria were used in selecting the participants and what data were collected on their language experience and proficiency, along with any other individual difference measures; d) how the data were presented and analyzed statistically. Finally, provide a critique of the study in which you identify at least one open question for future research that cannot be addressed by the results of the experiment(s).
5) What is meant by the claim that the visual world method has high ecological validity for examining language processing? For example, discuss why research such as Altmann and Kamide (1999) on scene processing provides a compelling example of the ecological validity of the method.

Research Project Option

A good way to begin to do research is via replication. Spivey and Marian (1999) examined late Russian-English bilinguals and showed that even in a monolingual Russian experimental context, native Russian's eye-movements reveal a competitor effect for objects whose English translation equivalents exhibit phonological overlap with Russian target items. Design and carry out a replication of this study using another language pair. If you cannot replicate the study, consider the factors that may underlie your null finding (e.g., the participants' language levels and dominance, the language pair, the stimuli, or some other factor). If you can replicate the finding, speculate on how it relates to Ju and Luce's (2004) finding that fine-grained phonetic detail in the acoustic signal is sufficient to constrain parallel activation and allow for selective lexical access in bilinguals.

Author Note

The writing of this chapter was supported in part by NSF Grant BCS-0821924 to P. E. Dussias and Chip Gerfen and NSF Grants BCS-0955090 and OISE-0968369 to J. F. Kroll and P. E. Dussias. J. Valdés Kroff was supported by a National Science Foundation Graduate Fellowship.

Suggested Readings

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German Gender. *BUCLD 36: Proceedings of the 36th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137,* 151–171.

Ju, M., & Luce, P. A. (2004). Falling on Sensitive Ears—Constraints on Bilingual Lexical Activation. *Psychological Science, 15,* 314–318.

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

Spivey, M., & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284.

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. In K. Forbus, D. Gentner, & T. Tegier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Altmann, G. T. M. (2011a). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137,* 190–200.

Altmann, G. T. M. (2011b). The mediation of eye movements by spoken language. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 979–1004). Oxford, UK: Oxford University Press.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111,* 55–71.

Arnold, J., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56,* 521–536.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457–474.

Blumenfeld, H. K., & Marian, V. (2005). Covert bilingual language activation through cognate word processing: An eye-tracking study. In *Proceedings of the twenty-seventh annual meeting of the cognitive science society* (pp. 286–291). Mahwah, NJ: Lawrence Erlbaum.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48,* 653–685.

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

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 687–696.

Cooper, R. (1974). The control of eye-fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory and language processing. *Cognitive Psychology, 6,* 84–107.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods, 7,* 41–63.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34,* 269–284.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42,* 465–480.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30,* 498–513.

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

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *PPVT-III Peabody: Test de vocabulario en imágenes.* Madrid, Spain: Tea, D. L.

Dussias, P., Valdés Kroff, J., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looks go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35,* 353–387.

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

Eyelink 1000 [Apparatus and software]. Mississauga, Ontario: SR Research.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. *Developmental Psychology, 42,* 98–116.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. *BUCLD 36: Proceedings of the 36th Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture, 10,* 39–58.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137,* 151–171.

IMSI Master Clips Premium Image Collection [Computer software]. (1990). Novata, CA: IMSI Inc.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science, 15,* 314–318.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156.

Kaplan, E., Goodglass, H., Weintraub, S., & Segal, O. (1983). *Boston Naming Test.* Philadelphia, PA: Lea & Febiger.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18,* 193–198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63,* 447–464.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31,* 1–24.

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24,* 173–193.

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within and between-language competition. *Bilingualism: Language and Cognition, 6,* 97–115.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* 33–42.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59,* 475–494.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in Aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language, 117,* 53–68.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108,* 155–184.

Perrotti, L. (2012). *Grammatical gender processing in L2 speakers of Spanish: Does cognate status help?* (Unpublished Undergraduate Honors Thesis). Penn State University, University Park, PA.

R [Open Source Computer Software]. (2008). Vienna, Austria: R Development Core Team. Retrieved from http://www.r-project.org/.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology, 63,* 772–783.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition, 105,* 466–476.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57,* 1030–1033.

Sedivy, J. E., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental interpretation through contextual representation: Evidence from the processing of adjectives. *Cognition, 71,* 109–147.

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 147–215.

Spivey, M., & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10,* 281–284.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E., (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51,* 247–250.

Tanenhaus, M. K. (2007). Spoken language comprehension: Insights from eye movements. In M. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics* (pp. 309–326). Oxford, UK: Oxford University Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science, 286,* 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movement and spoken language comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 835–862). Amsterdam: Elsevier.

Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in Agrammatic Aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research, 38,* 255–283.

Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1–25.

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. In K. Forbus, D. Gentner, & T. Tegier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Yee, E., Blumstein, S. E., & Sedivy J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's Aphasia: Evidence from eye movements, *Journal of Cognitive Neuroscience, 20,* 592–612.