3

USING THE LOOKING-WHILE-LISTENING PROCEDURE FOR SECOND LANGUAGE RESEARCH

Casey Lew-Williams

Introduction

This chapter provides a practical, accessible guide to implementing the looking-while-listening (LWL) procedure in second language research, and concludes with an overview of research using LWL to investigate real-time processing in native and non-native speakers. The procedure capitalizes on the basic human perceptual capacities of (1) using eyes to look, and (2) using ears to listen. Its simple name reflects its use of basic perceptual abilities and its relatively simple implementation.

In a fundamental respect, LWL is indistinguishable from other looking-time methods. Participants look at pictures (sometimes a single picture, sometimes two to four pictures, and sometimes a complex scene), and they listen to speech that has the potential to direct their attention to some aspect of the visual scene (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). Like automated eye trackers and the intermodal preferential looking paradigm, LWL generates moment-to-moment data about eye gaze, but the differences exist in how eye-movement data are collected, processed, and analyzed. I will not focus here on the many merits of automated eye trackers (Tobii, EyeLink, ASL, SMI, etc.), but instead on the merits associated with LWL, which has several advantages over other options for collecting data related to eye movements. First, many labs cannot afford automated eye trackers; LWL is cheaper than most other options. Second, not every lab has personnel who are sufficiently skilled in programming to successfully implement automated eye-tracking studies; LWL requires minimal programming skills. Third, recalibration in automated eye-tracking methods sometimes leads to momentary data loss; LWL is minimally prone to data loss. This is particularly advantageous for research on young children, whose inattention—needless to say—contributes to attrition and high rates of data loss. Fourth, the technology needed for portable LWL setups is straightforward, making it a feasible means for collecting data in diverse locations outside of university communities and outside of first-world countries. However, there is one main disadvantage associated with LWL relative to automated methodologies: Coding of eye movements takes time. Using specialized software, the burden is not as great as one might imagine, but LWL data do require more postprocessing relative to automated eye trackers.

The important commonality between researchers who employ different methodologies for tracking eye movements is that they value real-time measures as a window into learning and processing—or more broadly, as a window into cognition. The *means* of acquiring information about participants' eye movement behaviors are not of critical importance—experimental rigor is maintained regardless of researchers' reliance on online vs. offline coding. After all, many different eye-tracking methodologies yield strings of 1s and 0s that can be analyzed carefully in R, SPSS, JMP, and other statistical analysis programs. LWL is one tried-and-true approach, and it has been used to address a range of questions about real-time language processing. I will begin this chapter with a brief overview of how the LWL procedure has been used to address questions about language learning in young children. Then, I will provide information about implementing LWL, with special attention to issues that require consideration in second language (L2) research. Finally, I will describe research on Spanish gender-marked and number-marked articles as an example of how the LWL procedure can uncover precise time-course information about the emergence of second language processing.

LWL as a Measure of Young Children's Language Processing

LWL has yielded many exciting findings in research on language processing in young children, revealing that specific referential contexts interact with children's perceptual capacities in real time to shape interpretation of incoming speech. Eye movements are a natural and unschooled behavior, and visual fixations occur rapidly in both adult and child participants. Listeners develop hypotheses quickly as words unfold, using portions of the incoming speech stream to interpret what word is likely to come next, often time-locked to scanning of the visual field. Once sufficient information about the incoming word is gathered, listeners tend to move their eyes rapidly to potentially relevant targets, and once that target has been visually and/or acoustically processed, listeners either fixate or move on to other referents.

As children get older and collect more experience listening to language, they become faster and faster to move their eyes and establish reference, needing to hear less and less of a word to initiate a response. In one study (Fernald et al., 1998), the time course of familiar word recognition was tested in 15-, 18-, and 24-month-olds. Participants looked at two pictures (e.g., a ball and a shoe) and heard a very simple sentence referring to one picture (e.g., "Where's the ball?"). The dependent measure was the speed with which children moved their eyes from the distracter picture (the shoe) to the target picture (the ball). With age, children in this cross-sectional study became increasingly faster at finding the appropriate referent, with reaction times decreasing from 995 to 827 to 679 ms at 15, 18, and 24 months of age, respectively. By the second birthday, children responded before the offset of the target noun itself, showing that young children gain impressive efficiency in recognizing familiar words. While this research is based on English-learning children, studies investigating the emergence of incremental processing in Spanish-learning and bilingual children have begun to show comparable growth in spoken word recognition (Hurtado, Marchman, & Fernald, 2007).

LWL has also been effective in uncovering the emergence of predictive language processing in young children, and many studies have adapted real-time measures and experiment designs used previously in adult research. In adult studies, a variety of methods have been used to tap the time course of language processing, including lexical decision, auditory naming, and eye tracking, among others. For example, when adults hear the verb 'eat' in the presence of edible and inedible objects, they instantly shift their attention to food items, as if predicting what a speaker is most likely to talk about (Altmann & Kamide, 1999). Prior to hearing the full name of a referent, adults also exploit *partial phonetic information* to predict the completion of the noun (Marslen-Wilson & Zwitserlood, 1989), *prenominal adjectives* to find objects that differ in color (Sedivy, Tanenhaus, Chambers, & Carlson, 1999), *filled pauses* like 'uh' and 'um' to find objects that are new to the discourse (Arnold, Tanenhaus, Altmann, & Fagnano, 2004), *sentence frames* such as 'Look at the' to recognize a sentence-final word (Lieberman, 1963), *number information* to predict single referents vs. sets of referents (Lew-Williams & Fernald, 2009), *gender-marked pronouns* to identify male vs. female characters (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000), and *gender-marked articles* to identify objects whose names differ in grammatical gender (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Lew-Williams & Fernald, 2010). Using the LWL procedure, young children have also shown skill in using each of these structures to efficiently process words that come next (verbs: Fernald, Zangl, Portillo, & Marchman, 2008; partial phonetic information: Swingley, Pinto, & Fernald, 1999; prenominal adjectives: Fernald, Thorpe, & Marchman, 2010; filled pauses: Kidd, White, & Aslin, 2011; sentence frames: Fernald & Hurtado, 2006; number information: Lukyanenko & Fisher, 2014; gender-marked pronouns: Arnold, Brown-Schmidt, & Trueswell, 2007 [using eye tracking, not LWL]; grammatical gender: Lew-Williams & Fernald, 2007). AuQ014

These studies share an important finding: Young children can use currently heard speech to more rapidly identify what words come next. Time 'saved' is on the order of milliseconds, but this predictive processing capacity has consequences for learning: In one study, 36-month-old children who were slower to interpret a sentence-medial adjective/noun combination were less successful in learning the name of a novel word that appeared at the end of the sentence (Fernald, 2009).

Critically, understanding the time course of young children's language interpretation has proven valuable in its use as a predictive technique. A recent longitudinal study demonstrated that children's efficiency in language processing powerfully predicts language and cognitive outcomes in elementary school, above and beyond knowing about children's vocabulary levels (Marchman & Fernald, 2008). And recent work shows that efficiency in processing is an important mechanism behind the established link between quantity of child-directed speech in the child's environment and later vocabulary knowledge (Weisleder & Fernald, 2013).

Implementing LWL

Fernald et al. (2008) and Swingley (2011) provided detailed guides for setting up and using the LWL procedure, including information about designing experiments, coding eye movements, and analyzing data. Here, I will include many of the same topics but devote extra attention to issues pertaining to data collection and analysis with adult (particularly L2) participants.

Designing experiments. Anybody with corrected vision and hearing can be a participant in a LWL study, including infants and elderly individuals (but note that the method could also be used to study the effects of hearing impairments on language processing, e.g., Grieco-Calub, Saffran, & Litovsky, 2009). In a simple LWL study, researchers may want to assess processing of familiar nouns, in which case there may be only one trial type, but multiple participant groups (e.g., L1 vs. L2 learners, or children ages 18 vs. 24 months). Other studies manipulate subtle aspects of speech or visual stimuli to create two or more within-subjects conditions. For example, a study about using verbs to 'listen ahead' could include some sentences with semantically informative verbs (e.g., "Eat the cookie" while viewing a cookie and a shoe) and some with semantically uninformative verbs (e.g., "Look at the cookie" while viewing the same two pictures). But, as in any eye-tracking studies, it is possible to design experiments with substantially more complex sentences, as in studies investigating relative clause ambiguity in L2 learners using automated eye trackers (Dussias, 2004).

Studies can last any desired total amount of time, but researchers should note that there is an art to designing studies that are engaging for different populations. Studies with children ages 1–3 years should typically last no more than 5–6 minutes, which would involve 24–32 test trials plus a fun filler trial occurring every 4–5 trials. Filler trials can be anything related to or unrelated to the goal of the experiment; for example, filler trials could include colorful pictures of visually engaging stimuli (e.g., jellyfish), short videos that re-engage attention (e.g., fireworks or a baby laughing), or information about when the study will be over (e.g., "You're doing great. We're almost done."). Studies with young children that last longer than 5–6 minutes typically result in fussiness, inattention, and data loss. But studies with older children (e.g., 5- to 12-year-olds) can be longer, perhaps up to 10 minutes, but filler trials should be designed to appeal to this age group. For research with adults, studies can be significantly longer in order to accommodate greater diversity in auditory and visual stimuli. For each of these groups, multiple studies can be interleaved, such that fillers for one study are experimental trials for the other. The risk is that this will likely yield fewer trials for any particular study or condition, leaving the researcher with less statistical power than would be ideal.

On each test trial—prior to hearing the auditory stimulus—participants typically have two seconds to inspect the pictures in silence. This 2-second period is intended to reduce the likelihood that eye movements will result from superficial attention to visually appealing objects and to increase the likelihood that participants will respond meaningfully to the speech signal, but we do not currently know if it is an essential aspect of the research design. Following this brief inspection period, participants hear the speech stimulus, which consists of a word or sentence depending on the goals of the experiment. Then, participants are usually given 1–3 seconds to continue viewing the pictures after the speech stimulus has ended.

The experimenter should document the contents of each trial (trial number, left picture, center picture, right picture, sound stimulus, target location, condition, onset of critical words in ms, etc.) in carefully counterbalanced orders. Specifically formatted order files are needed for linking to individual participant's eye movement behaviors, as described below.

Instructions to participants. When testing adults in the LWL procedure, instructions are essential. Given that the experiments are sometimes designed for young children, there is a possibility that adults will outsmart the study and apply top-down strategies. This can be avoided. Experimenters should not provide instruction to 'look as fast as possible at the correct picture.' Instead, experiments should use simple instructions such as the following: "This study was designed for young children. You will see pictures on a screen. When you see the pictures, you will also hear a sentence. Your job is to look at the pictures and listen carefully." Instructions can be modified depending on the nature of the experiment. Adult participants often ask if they are allowed to move their heads or if they are supposed to look anywhere in particular, and the experimenter should simply encourage them to respond naturally. As the session begins, the experimenter should ensure that the participant is seated at a reasonably comfortable viewing distance. The operating principle for viewing distance should be: the closer the better. The participant's eyes should be approximately 60 cm away if using a large flatscreen TV (e.g., 140 cm/55 in. in size) and will need to be 30–45 cm away if using a laptop.

Hardware. The hardware required for successful implementation of LWL depends on the setting of the research. In the end, the goal is to get a recording of participants' faces as they look at pictures and listen to sentences, and the means of achieving this goal are numerous. The most basic approach to using LWL involves a laptop computer and an external camera positioned above the laptop, facing the participant. This simple, portable setup is particularly useful for field research. A full-screen video of a sequence of test trials is played to each participant, ideally starting with a momentary 'beep' so that the video of the participant's eye movements can be synchronized to the stimulus video after the fact. Currently, cameras built into laptops are not reliable. They drop frames and can be problematic for coding eye movements after data have been collected, so an external camera is essential.

While laptops offer portability, most LWL studies have taken place in lab settings using a substantially more complicated network of equipment. There are two main advantages to conducting research in the lab: The experimenters have more control over visual and auditory distractions in the surrounding environment, and the ability to present stimuli on a larger screen generates larger gaze shifts that enhance inter-rater reliability. In general, the following hardware is needed for setting up LWL (excluding cables): computer for controlling the presentation of stimuli, flatscreen television for presenting visual stimuli, external speakers for presenting auditory stimuli, video camera for recording eye movements, computer for recording eye movements, microphone for recording sound in the testing room, and quad splitter for combining the participant video and stimulus video into one file. Depending on the use of digital or analog equipment, various converters will also be necessary between these devices. Note that lighting in the testing room should be bright enough to see eye movements, and brightness varies substantially between different cameras.

Software. For controlling stimuli and timing, different labs have used a variety of software programs, including Matlab, PsyScope, and Habit. But the laptop version of LWL requires nothing more than playing a digital video (e.g., in Quicktime). Video editing software such as Adobe Premiere is also needed after data collection in order to superimpose a timestamp over the recording of the participant. For coding eye movements, researchers have used custom software called EyeCoder, originally developed in Anne Fernald's lab at Stanford University. EyeCoder allows for experimenters to code at each 33-ms frame whether the participant is looking left vs. right vs. center, shifting between the pictures, or looking away from the pictures. Other software has been used for similar purposes, such as SuperCoder, originally developed in George Hollich's lab at Purdue University. A novice coder typically takes about 60 minutes to code a five-minute video of a participant, and a skilled coder typically takes about 30 minutes. Thus, a study with 24 participants can be coded in as few as 12 hours, which is quite manageable when distributed over days or weeks. Inter-rater reliability should also be assessed, which involves a second researcher coding 15–25% of participant videos and determining (1) the proportion of frames on which the two coders agree on their responses (give or take one frame), and (2) the proportion of frames *surrounding shifts* on which the two coders agree. EyeCoder automatically calculates inter-rater reliability. After eye movements have been coded for the desired number of participants, a program called DataWiz exports gaze location and links it to an experimenter-created order file containing information about each trial. This output links frame-by-frame, trial-by-trial content (e.g., target location) with specific information about each participant (e.g., sex, age, or status as an L1 or L2 learner).

Measures of efficiency in language processing. DataWiz generates reaction time information on each trial for each participant defined from the onset of relevant acoustic information. This onset is determined by linking the timing information entered into each line of the order file with the manually entered responses for each 33-ms frame on each trial. Reaction time is determined automatically by DataWiz, and an advantage of LWL over methods that rely on summary measures of looking time (e.g., head-turn preference) is that it yields highly precise time course data. As with other methods that monitor eye movements, data can be noisy for an individual participant; thus, a single reaction time is almost meaningless and not sufficiently indicative of processing speed for a particular condition. Child participants can be prone to distraction, even in the middle of a trial; thus, researchers should determine a strict criterion for the number of usable trials per condition that would yield a meaningful mean. Three trials is a minimum criterion, although more trials will yield more reliable estimates of RT. With adults, nearly all trials should be usable, and therefore RT data should be abundant in any given experimental condition.

DataWiz also enables easy calculation of accuracy, or the reliability of looking to the target referent (as a proportion of looking to either the target or distracter), during a specific window of time. In the output from DataWiz, each row contains relevant trial and participant information plus a frame-by-frame string of 1s, 0s, dots, and dashes (1 = looking at target picture, 0 = looking at distracter picture, dot = shifting between the pictures, dash = looking away from both pictures, and if a center picture is included in the experiment design, .5 = looking at center picture). Determining accuracy requires deciding upon an appropriate analysis window, i.e., the window of time when eye movements are most likely to be meaningfully related to the speech signal. The same reasoning applies to reaction time, as researchers must select which reaction times to exclude for being either too quick or too delayed. Generally, researchers need to proceed with caution when selecting analysis windows because it is easy to find a small effect *somewhere* in the seconds following the onset of relevant speech, and it may not be fair to capitalize on a fleeting, momentary effect. Different publications have selected analysis windows in different ways. After taking into account the time it takes to initiate the motor plan for an eye movement in response to speech, which is usually on the order of 150–367 ms (Haith, Wentworth, & Canfield, 1993; Fernald et al. 2008; Matin, Shao, & Boff, 1993; Saslow, 1967), researchers sometimes analyze accuracy out to a fixed endpoint of 1,800 or 2,000 ms, sometimes in incremental chunks (e.g., 500–1,000 ms, 1,000–1,500 ms, 1,500–2,000 ms, and so on), and sometimes in specific windows corresponding to the lengths of words in the sentence (e.g., article window, noun window, postnoun window).

A third and underused dependent measure is known as shift tendency, used in only one publication using LWL (see Thorpe & Fernald, 2006). This measure provides a window into participants' uncertainty about identifying the target vs. distracter picture. Shift tendency is nothing but a tally of the number of shifts occurring during a particular analysis window. If participants have poorer representations of target words, they tend to shift back and forth before landing on an interpretation (or they may not resolve the meaning behind the sentence at all).

Cleaning data. The DataWiz output file may need to be cleaned up in order to be ready for statistical analyses. This can be done manually in SPSS or using R scripts (or equivalents). First, unusable trials should be identified prior to inclusion in participant means for accuracy or reaction time. Trials are usually eliminated from analysis if a participant is inattentive, which occurs more often for child than adult participants. Study-specific criteria can be used to eliminate trials, such as requiring participants to be attentive for a particular proportion of each trial. A common standard is to eliminate a trial if participants look away for longer than 500 ms during the window of interest (i.e., 15 consecutive frames, assuming a camera captures 30 frames per second).

A particularly critical issue in research with adult participants is deciding how to characterize different kinds of looking behaviors in LWL. In the DataWiz output file, trials are categorized by where the participant was looking at the onset of relevant acoustic information (0 ms). Most trials are labeled as either distracter-initial (D-initial) or target-initial (T-initial). D-initial trials are used to calculate reaction time—the time it takes for a participant to initiate the shift from the distracter to target picture. Both D- and T-initial trials are used to calculate accuracy. In typical two-alternative LWL studies, there is no centrally located fixation point, and participants are not instructed to fixate anything in particular before initiating an eye movement. A central fixation point is common in adult eye-tracking studies, but in many ways, it violates the naturalness of looking and listening; rarely in life are we forced to look at an object until we are ready to make a decision about where to shift next. The lack of a fixation point is an advantage of the LWL procedure.

However, in the experimental context, some adults are indeed hesitant to freely look and listen, and they fixate the center of the screen—even in the absence of a stimulus—until they decide to look left or right. In some cases they do not look at the pictures prior to hearing the stimulus sentence and instead rely on peripheral vision. In essence, they are applying some kind of top-down strategy, and in my experience, 10–15% of adult participants adopt this behavior consistently within an experimental session. The question is: should these trials be included in analyses? In the DataWiz output, they appear as Away trials (A-trials), which are typically excluded from analyses. But for adults, many of these trials are potentially usable and do not reflect being inattentive; instead, they may reflect being hyper-attentive. Researchers should take the time to watch each of these A-trials individually in real time and make a judgment about whether the participant is on-task or off-task. This is somewhat subjective, but in most cases, it is clear whether the participant is strategizing vs. not focusing on the experiment. For the on-task trials, researchers should manually enter reaction time information in the appropriate column and re-label the trials (e.g., 'C-initial trials', or center fixation trials). C-initial trials often have slower reaction times than D-initial trials, possibly because adults are waiting until they reach a confident decision about looking left vs. right. Then, researchers should make a sound decision about whether the C-initial trials should be included in data analyses. In general, if C-trials represent a small minority of trials, it is probably safer to exclude them from analyses, as they deviate from the participants' typical looking behavior. But a participant does not necessarily need to be excluded if he/she deploys this strategy throughout the experimental session.

Individual differences in language proficiency. With sufficient test trials per condition, LWL has the power to yield reliable measures of efficiency in language processing that can be compared to other components of a research project. Child language researchers are often interested in whether LWL measures of reaction time and accuracy correlate with children's cumulative vocabulary or grammatical knowledge (see Lew-Williams & Fernald, 2007). L2 researchers are often interested in how real-time processing interacts with measures of language proficiency. Researchers typically gather measures of self-reported proficiency in speaking, understanding, reading, and writing; measures of off-line language proficiency such as performance on grammaticality judgment tasks or standardized language assessments; measures of day-to-day or week-to-week language use with friends and family; measures of classroom language experience; and measures of age of acquisition (see Grüter, Lew-Williams, & Fernald, 2012). These measures collectively offer an exciting opportunity to uncover how maturation and experience interact to support L2 learning and how processing efficiency emerges from and/or paves the way for proficiency.

A Case Study of LWL as a Measure of Second-Language Processing

To conclude this chapter, I will overview a series of studies that compare the language-processing capabilities of first and second language learners using the LWL procedure. This work focuses on Spanish grammatical gender as a case study for understanding the potential of LWL to uncover subtle aspects of L2 learning. Collectively, these experiments reveal that specific referential contexts shape efficiency in first and second language processing.

Gender-marking languages present an interesting case for spoken word recognition, especially those languages with high co-occurrence statistics between gender-marked articles and the nouns that succeed them, like Spanish. Lew-Williams and Fernald (2007) monitored the eye movements of Spanish-learning 2- and 3-year-old children from low-income families as they looked at pairs of objects with names of either the same or different grammatical gender. (Note that all stimuli for this study are available for free download on the IRIS Digital Repository). On same-gender trials, participants looked at two pictures with names of the same gender (e.g., *pelota*, 'ball$_{[f]}$', *galleta*, 'cookie$_{[f]}$') and heard a sentence directing them to one picture (e.g., *Encuentra la pelota*, 'Find the ball'). In this case, the gender-marked article *la* provided no information about the identity of the subsequent noun; not until the acoustic onset of the noun did the target referent become clear. On different-gender trials, children viewed pictures depicting objects with names of different grammatical gender (e.g., *pelota*, *zapato*, 'shoe$_{[m]}$') and heard an identical sentence. On these trials, *la* was potentially informative about the identity of the upcoming referent. Their native Spanish-speaking parents were also tested in this paradigm. While listening to simple sentences, both children and adults were faster to orient to the target picture when the gender-marked article was useful, i.e., on different-gender trials. Adults showed faster and more accurate looking than children, but the two groups were comparable in the efficiency of establishing reference. Those children who demonstrated more efficient processing were the same children who showed more advanced development in productive vocabulary and grammar measures, as has been demonstrated in other studies (Fernald et al., 2006). In this task testing the use of morphosyntactic cues in online comprehension, listeners took advantage of a subtle prenominal cue when their visual field lent itself to this advantage, revealing how the young child learning a richly inflected language makes use of co-occurrences in language and makes progress in 'becoming a native listener' (Werker, 1989).

The next goal of this project was to assess whether making use of *la* and *el* as predictive cues in real-time language processing is more difficult for L2 learners of Spanish than for native Spanish-learning children. Lew-Williams and Fernald (2010) tested L2 Spanish-learning adults in an identical LWL paradigm: a short, child-friendly study that involved looking at highly familiar objects and listening to simple Spanish sentences. L2 Spanish speakers in this study had started learning Spanish at 12.8 years on average and had learned Spanish in a classroom for five years on average. Despite these moderate levels of proficiency, the L2 Spanish-learning adults were not able to take advantage of informative gender-marked articles to identify familiar nouns. That is, they were not faster to shift to the target referent when the article did vs. did not provide a prenominal cue to its identity.

While findings indicate that L2 Spanish learners are unable to process local morphosyntactic relations in 'native-like' ways, I next examined a diverse range of article-noun relations in order to examine the nature of L2 learners' difficulty. In one experiment, L1 and L2 adults (with comparable Spanish proficiency as adults in Lew-Williams & Fernald, 2010) viewed female and male humans belonging to either the same or different *biological* gender categories. For example, they listened to the sentence *Encuentra la niña* ('Find the girl$_{[f]}$') while looking either at pictures of a male and female (on different-gender trials) or a girl and a woman (on same-gender trials). Like native Spanish speakers, L2 adults succeeded in using informative gender-marked articles to more rapidly orient to female vs. male faces. Similar efficiency in processing was shown in an experiment testing processing of *number*-marked articles, in which L1 and L2 adults viewed pictures showing either the same or different numbers of objects (e.g., one cat vs. one dog; one cat vs. five dogs) and heard sentences referring to them (e.g., *Encuentra el perro* vs. *Encuentra los perros*). Parallel to the findings about biological gender-marked articles, both L1 and

L2 adults took advantage of number-marked articles to interpret the number of referents in the visual scene. Thus, L2 adults did not achieve native-like efficiency in processing article-noun phrases when articles revealed *only* grammatical gender information, but native-like processing was observed when other kinds of information were marked in the article: biological gender and number. Importantly, test sentences in the biological gender study and the number study were identical in structure, consisting of a sentence frame, an article, and a noun. Given the similar demands across experiments, this work documented a disparity in the predictive processing skills of L2 learners.

These studies left open the question of why L2 Spanish-speaking adults failed to exploit articles that index membership in largely arbitrary noun classes. One possibility is that native Spanish speakers had heard the article-noun pairings many more times than L2-learners; thus, the findings could result from differential frequency of exposure to particular article-noun co-occurrences. Lew-Williams and Fernald (2010) asked: Will L1 and L2 adults still differ if they receive equal exposure to novel nouns, even when processing article-noun pairings they have never heard? In two experiments, L1 and L2 adults learned novel nouns in Spanish with no instruction to attend to gender. On teaching trials, one of four novel objects (i.e., objects with no conventional name) appeared on a screen as participants heard a prerecorded sentence with a novel noun (e.g., *Mira, es la catela*, 'Look, it's the$_{\text{[f/def.]}}$ catela$_{\text{[f]}}$'). On test trials, participants viewed pairs of novel objects with names of either the same (*catela*, *pifa*) or different grammatical gender (*catela, tebo*), as they heard a sentence referring to one picture (e.g., *Encuentra la catela*, 'Find the$_{\text{[fem./def.]}}$ catela'). If participants demonstrated incidental learning of gender, we predicted they would use the definite article as a cue to the subsequent noun, shifting their eyes to the target more rapidly on different-gender trials than on same-gender trials. Reaction times revealed that both L1 and L2 participants succeeded in taking advantage of gender-marked articles to more rapidly establish reference. However, this experiment ignored a salient component of real grammatical gender: that no noun is uniquely associated with a single article; indefinite articles and other determiners also precede nouns regularly. Thus, this experiment with novel nouns may not have reflected the same processes as those used to process familiar article-noun pairs. In a follow-up experiment, L1 and L2 adult Spanish speakers participated in a near-replication of this experiment, but with one subtle change: On teaching trials, only indefinite articles were used in sentences, and participants never heard definite articles in sequence with the novel nouns (e.g., *Mira, es una catela*, 'Look, it's the$_{\text{[f/indef.]}}$ catela$_{\text{[f]}}$'). At test, participants heard definite articles for the first time in the experiment. Interestingly, L1 adults effortlessly generalized between the different articles forms, responding significantly faster on different-gender than on same-gender trials. But L2 adults waited for the noun, just like they did when processing familiar article-noun sequences. Both groups learned the nouns with equivalent accuracy—evident in the reliability with which they (eventually) fixated the target referents. But critically, fluent Spanish-speaking adults flexibly accessed and integrated gender cues in early stages of word learning and then generalized this learning to a different article on test trials, while L2-learners did not demonstrate such automaticity.

In an investigation concerning the influence of L2 proficiency on real-time processing of grammatical gender, Grüter, Lew-Williams, and Fernald (2012) recruited highly proficient, late L2 learners of Spanish who had learned Spanish at age 10 or later and demonstrated native-speaker range performance on various oral and written proficiency measures. Participants included professional Spanish-English translators who work in medical facilities, individuals whose spouses were native Spanish speakers and had decided to speak only Spanish in their households, and parents who were *not* native Spanish speakers but had decided to speak only Spanish to their children. Participants were tested on their ability to exploit gender-marked articles as cues to familiar nouns and to novel nouns (in the paradigm requiring generalization between the indefinite and definite article). With this highly proficient L2 population, we found an interesting dissociation: (1) L2 participants failed to process familiar article-noun phrases as efficiently as L1 participants, which is not only surprising because of the child-friendly, simple design of the six-minute experiment, but because these L2 participants were excellent speakers of Spanish who used the language regularly in their daily lives. Yet, they failed to achieve a processing capacity demonstrated robustly by 2- and 3-year-old Spanish-learning children growing up in low-income family contexts. (2) The same participants succeeded in exploiting articles in the novel noun study requiring generalization between article forms, indicating that native-like processing of grammatical gender cues is not beyond their reach.

Collectively, these studies point to a fascinating interaction between language experience (e.g., in childhood vs. adulthood), age of exposure (e.g., birth vs. adolescence), and the nature of specific referential contexts (e.g., objects that contrast in gender vs. number). To further examine this interaction, I recruited a group of participants who offer an interesting test case for the role of language input in shaping language processing: children enrolled in elementary school Spanish immersion programs. This increasingly popular format for early education is intended to foster bilingualism in both native and non-native speakers, incorporating some of the benefits of home language environments, such as immersive language and cultural experience. By devoting hours each day to Spanish and English, language immersion is intended to confer advantages over other formats of language instruction, such as typical high school language classrooms. Participants were L1 and L2 Spanish learners in either the first two years of elementary school (kindergarten and 1st grade) or the last two years of elementary school (4th and 5th grade). In the grammatical gender processing study, the native English-speaking children learning Spanish could resemble native Spanish speakers because they were receiving more daily exposure to Spanish relative to L2 adults. But in other ways, they could resemble adult L2 learners of Spanish because they were nonetheless learning Spanish against the backdrop of English. While the L2 children succeeded in achieving native-like processing of cues in articles to biological gender and number, they failed—in both age groups—to take advantage of *grammatical* gender-marked articles. L1 Spanish-speaking children showed faster word recognition on different-gender trials in all three experiments.

Taken together, these experiments reveal consistent efficiency in processing Spanish gender-marked articles among L1 toddlers, L1 kindergarteners, L1 5th graders, L1 undergraduates, and L1 parents, and consistently *less* efficient processing among L2 kindergarteners, L2 5th graders, L2 undergraduates, and highly proficient L2 adults. There are several possible interpretations for this seemingly categorical divide between L1 and L2 learners. First, a language input explanation: Classrooms and immersion programs do not offer the same quantity/quality of language input that parents offer to infants and toddlers. Second, a maturational explanation: Exposure to language must occur early in life in order for learners to achieve native-like competence. And third, an explanation pertaining to the influence of L1: Proficiency in English (a language with no grammatical gender system) could constrain the ability to achieve native-like mastery in processing grammatical gender in Spanish. Future LWL studies (in tandem with other measures of real-time language processing) will ideally disentangle these competing explanations by testing diverse populations of L2 learners in diverse locations in diverse languages in diverse language learning contexts.

Conclusion

There is tremendous room in the field of second language learning to use the looking-while-listening procedure as a tool for uncovering how languages are learned and processed, particularly in terms of how the first language shapes efficiency in processing in the second language, how qualitative and quantitative differences in language experiences shape efficiency in processing, and how individual differences in cognitive, attitudinal, and motivational variables influence the ultimate attainment of 'native-like' processing. Measuring L2 abilities can happen in a variety of ways using a variety of effective tools, and this chapter is not intended to promote LWL as the superior option. This chapter has conveyed that the LWL procedure (1) achieves both ease of implementation and precision of data and (2) has great potential for generating advances in research on second language learning.

References

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

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*, 527–565.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*, 578–582.

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

Dussias, P.E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English bilinguals. *International Journal of Bilingualism*, *8*, 355–371.

Fernald, A. (2009). Developing fluency in understanding: How it matters. Keynote address at the 34th Annual Boston University Conference on Language Development.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science, 9*, F33-F40.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228–231.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology*, *60*, 190–217.

Fernald, A., Zangl, R., Portillo, A.L., & Marchman, V.A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I.A. Sekerina, E.M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Amsterdam: John Benjamins.

Grieco-Calub, T.M., Saffran, J.R., & Litovsky, R.Y. (2009). Spoken word recognition in toddlers who use cochlear implants. *Journal of Speech, Language, and Hearing Research*, *52*, 1390–1400.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*, 191–215.

Haith, M.M., Wentworth, N., & Canfield, R.L. (1993). The formation of expectations in early infancy. *Advances in Infancy Research, 8*, 251–297.

Hurtado, N., Marchman, V.A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, *34*, 227–249.

Kidd, C., White, K.S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental science*, *14*, 925–934.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 33*, 193–198.

Lew-Williams, C., & Fernald, A. (2009). Fluency in using morphosyntactic cues to establish reference: How do native and non-native speakers differ? *Proceedings of the 33rd Annual Boston University Conference on Language Development* (pp. 290–301). Somerville, MA: Cascadilla.

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.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–187.

Lukyanenko, C., & Fisher, C. (2014). 30-month-olds use verb agreement features in online sentence processing. *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 292–305). Somerville, MA: Cascadilla.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, F9–16.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, *15*, 576.

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

Saslow, M.G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America, 57*, 1024–1029.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Swingley, D. (2011). The looking-while-listening procedure. In E. Hoff (Ed.), *Research methods in child language* (pp. 29–42). Malden, MA: Wiley-Blackwell.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73–108.

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition, 100*, 389–433.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 20,* 1–10.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, 77, 54–59.