Studies in Second Language Acquisition, 2013, **35**, 389–415.
doi:10.1017/S0272263112000927

ECOLOGICAL VALIDITY IN EYE-TRACKING

An Empirical Study

Patti Spinner and Susan M. Gass
Michigan State University

Jennifer Behney
Youngstown State University

Eye-trackers are becoming increasingly widespread as a tool to investigate second language (L2) acquisition. Unfortunately, clear standards for methodology—including font size, font type, and placement of interest areas—are not yet available. Although many researchers stress the need for ecological validity—that is, the simulation of natural reading conditions—it may not be prudent to use such a design to investigate new directions in eye-tracking research, and particularly in research involving small lexical items such as articles. In this study, we examine whether two different screen layouts can lead to different results in an eye-tracking study on the L2 acquisition of Italian gender. The results of an experiment with an ecologically valid design are strikingly different than the results of an experiment with a design tailored to track eye movements to articles. We conclude that differences in screen layout can have significant effects on results and that it is crucial that researchers report screen layout information.

We thank Aline Godfroid for early discussions on research design and Sehoon Jung and Russ Werner for assistance with the eye-tracker. We also thank Zhen Zhang for assistance with the statistical analysis.

Versions of this article were presented at SLRF 2010 and Eurosla 2010 and 2011.

Correspondence concerning this article should be addressed to Patti Spinner, SLS Program, B-258 Wells Hall, Michigan State University, East Lansing, MI 48824; E-mail: spinnerp@msu.edu

© Cambridge University Press 2013

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Over the past 30–40 years, there has been considerable interest in issues of methodology in the field of SLA as researchers seek to find the most accurate ways to determine how languages are learned. Regardless of the particular goals of each subdiscipline of SLA (whether it be focused on understanding, for example, linguistic knowledge, the influence of context, or the role of interaction), the question of how to obtain that information remains foremost in researchers' minds. To aid researchers in data gathering and analysis, numerous books have appeared over the years that incorporate the gamut of concerns facing scholars in SLA (e.g., research questions, design, protocols, and analysis; see Brown, 1998; Dörnyei, 2007; Mackey & Gass, 2005, 2012; Nunan, 1992; Seliger & Shohamy, 1990). There have also been other treatises dealing with issues related to specific areas of research (e.g., think-aloud research, Bowles, 2010; stimulated recall, Gass & Mackey, 2000; priming, McDonough & Trofimovich, 2009; classroom research, Schachter & Gass, 1996).

To date, however, one area of research that has not been addressed by this methodological work is eye-tracking. The only book of which we are aware that deals specifically with methodological issues in eye-tracking research is Gass and Mackey (2007); however, it dedicates only three pages to a discussion of eye-movement studies. To some extent, the lack of work in this area may be due to the fact that, in general, SLA researchers have adopted the technology and methodology for eye-tracking from the field of psychology, addressing the same research questions that psychologists have asked regarding native speakers (NSs). However, at times, second language (L2) research addresses questions that are not dealt with in first language (L1) research—for example, questions relating to crosslinguistic influence, ultimate attainment, and so forth. As SLA researchers become more sophisticated in their use of eye-trackers, they will need to move away from reliance on methodologies established by researchers in other fields, given the different foci of the research questions. It is precisely in these newer research areas that there is little guidance. This article problematizes these issues by presenting the findings of two experiments and discussing the different results obtained by merely manipulating screen presentation. In particular, we pay attention to issues of ecological validity and the reasonableness of adhering strictly to such a design.

RESEARCH DESIGN IN SLA

All good research studies begin with research questions accompanied by designs and tools that can successfully answer those questions. In the history of SLA, there have been a variety of tools used for investigation and analysis. Some have been drawn from other disciplines, as is the case with acceptability judgments and with many, if not most, psycholinguistic

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

methods, and some have been unique to the field of SLA. In fact, Mackey and Gass (2005) and Gass and Mackey (2007) point out that data elicitation techniques are often limited only by a researcher's imagination. That is not to say that we should set up experiments without careful thought. As Gass and Mackey (2007) say, "there are intricate relationships between research questions and data elicitation techniques, and with so many possibilities, it can be hard to keep all of the necessary considerations straight" (p. ix). They point out that behind most research methods is a "body of practical knowledge" (p. ix) about that method that can guide researchers in their work and help them avoid potential errors that will influence the results of the study.

Even with this body of practical knowledge—namely, the numerous books and articles concerned with research design in SLA—there is not always agreement on how to appropriately address constructs, how to use tools, or how to perform analyses. For example, Leeser and Sunderman (2012) examined working memory tasks (i.e., reading span and operation span) and found that studies differ in terms of task type, scoring method, and type of data analysis, and that the researchers' results differed on the basis of these factors. Similarly, Gutiérrez (2013) questioned the way that grammaticality judgment tasks have been used to measure implicit and explicit knowledge, looking, in particular, at how researchers using grammaticality judgment tests need to take into account both time pressure and task stimulus or risk confounding their results (see also Loewen, 2009).

These efforts to refine and standardize research methodologies are a positive step for the field of SLA. In the L1 literature, discussions have taken place with regard to data analysis (see, for example, Barr, 2008; Mirman, Dixon, & Magnuson, 2008). Unfortunately, however, there has not been as much work on eye-tracking methodology. Few guidelines exist as to how to conduct eye-tracking studies in L2 research, and the guidelines that are available mainly pertain to two rather narrow areas of research focus—namely, research on reading behavior and research using the visual-world paradigm (see Roberts, 2012), for which the methodologies have been adapted from L1 research (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). As eye-tracking research moves into new areas of inquiry, many questions arise. In an effort to begin the process of clarifying and standardizing L2 eye-tracking research for a variety of research questions, we present in this article a survey of L2 eye-tracking research, providing—when available—guidelines presented in research reports. Additionally, in the second part of this article, we problematize issues by showing that different screen layouts yield different research results. We focus specifically on the concept of ecological validity.

In general, there is an emphasis placed on ecological validity in reading research in such a way that natural reading conditions are highly valued (P. Dussias & N. Sagarra, personal communication, October 17, 2010; Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, this issue). For such

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

studies, what is highly valued is (a) a small font size (i.e., 12–18 pt., as is the case in normal reading); (b) text broken only by the edge of the screen, as is the case in a book; and (c) one space between words. Although these requirements are the general consensus of experts in this area, there is little information in published research studies regarding such basics as font size, font type, location of fixation points, layout, and interest area (i.e., screen areas that measure eye movements) dimensions. This lack of information leaves researchers with relatively few clear and detailed guidelines for setting up experiments. The fact that there has not been a careful examination of methodological issues in the use of eye-tracking in L2 research is particularly significant when one breaks into a new territory of investigation, as was the case with the current study, which used eye-tracking to investigate the acquisition of morphosyntax with text rather than pictures.

Table 1 lists studies that have used eye-tracking methodology in L2 research.[1] As can be seen, many used only pictures to track eye movements. However, in those instances in which text was used, with a few exceptions, the only indication of screen layout and information about fonts was a mention of whether the fonts were upper- or lowercase.[2] As we demonstrate in this study, information about layout and size are of great significance when attempting to interpret and understand the results of eye-tracking studies.

BACKGROUND FOR THE CURRENT STUDY

As previously noted, L2 eye-tracking research has primarily either (a) observed learners' eye movements as they read a text to determine patterns of reading behavior or (b) used the visual-world paradigm, in which learners' eye movements are tracked to a limited number of pictures or objects. In this study, we take a different approach, investigating a morphosyntactic issue by examining eye fixations on various grammatical elements in the target language. In particular, we examine L2 learners' use of morphosyntactic and morphophonological cues to grammatical gender in Italian.

In Italian, morphosyntactic cues to grammatical gender involve agreement marking on determiners and adjectives, as illustrated in (1) and (2).

(1) *il* *gatto* *nero*
the [MASC] cat [MASC] black [MASC]
"the black cat"

(2) *la* *stella* *rossa*
the [FEM] star [FEM] red [FEM]
"the red star"

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Table 1. Previous eye-tracking studies

Author(s)	Year	Font	Screen display
Altarriba, Kroll, Scholl, & Rayner	1996	Not mentioned	1–2 lines, up to 72 characters per line
Blumenfeld & Marian	2007	Pictures only	Pictures only
Blumenfeld & Marian	2011	Pictures only	Pictures only
Bolger & Zapata	2011	Not mentioned	Multiline
Dussias	2004	Not mentioned	1 line
Dussias & Sagarra	2007	Lowercase letters	Not mentioned
Duyck, Van Assche, Drieghe, & Hartsuiker	2007	Not mentioned	Not mentioned
Felser, Sato, & Bertenshaw	2009	Fixed-width font (Courier New), 21 pt.	Several lines
Flecken	2011	Videos only	Videos only
Ju & Luce	2004	Pictures only	Pictures only
Kaushanskaya & Marian	2007	Times New Roman	Not mentioned
Keating	2009	Upper- and lowercase letters	1 line
Marian, Blumenfeld, & Boukrina	2008	Pictures only	Pictures only
Marian & Spivey	2003	Real objects only	Real objects only
Oller & Tullius	1973	Not mentioned	Not mentioned
O'Rourke	2008	Not mentioned	Not mentioned
Roberts, Gullberg, & Indefrey	2008	Not mentioned	3 lines, double-spaced
Schmiedtová	2011	Videos only	Videos only
Siyanova-Chanturia, Conklin, & Schmitt	2011	Not mentioned	Not mentioned
Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker	2011	Lowercase (except when capitals were appropriate), mono-spaced Courier font	Up to 2 lines, up to 85 characters per line
Weber & Cutler	2004	Pictures only	Pictures only

In example (1), the article *il* "the" and the adjective *nero* "black" are both marked for agreement with the masculine noun *gatto* "cat." In example (2) *la* is the feminine article and *rossa* is the feminine form of the adjective "red"; they agree with the feminine noun *stella* "star." If a hypothetical learner were not sure of the gender of the noun *gatto*, she could determine that it is masculine on the basis of the morphosyntactic cues on the article and the agreeing adjective. Likewise, the morphosyntactic cues on *la* and *rossa* in example (2) confirm that the noun *stella* is feminine. Gender agreement behavior for plurals is similar, although it will not be dealt with in this article.

On the other hand, morphophonological cues are noun endings that signal which gender (masculine or feminine) the noun belongs to. In Italian, nouns that end in *-o* have a strong tendency to be masculine, whereas nouns that end in *-a* have a strong tendency to be feminine. For example, *tavolo* "table" and *gatto* "cat" are masculine, and *sedia* "chair" and *casa* "house" are feminine. Masculine nouns ending with *-o* and feminine nouns ending with *-a* are referred to as transparent nouns because their morphophonological cues to gender are clear.[3] However, not all nouns follow this pattern. Many nouns are opaque—they end in neither *-o* nor *-a* and, thus, do not provide any morphophonological cue to gender. These nouns may be masculine or feminine. For example, *nome* "name" and *pesce* "fish" are masculine, and *notte* "night" and *carne* "meat" are feminine.[4]

In terms of L2 acquisition, there is robust evidence that L2 learners are sensitive to morphophonological cues to gender on nouns (e.g., Bordag, Opitz, & Pechmann, 2006; Bordag & Pechmann, 2007, 2008; Carroll, 1999; Franceschina, 2005). Learners are more successful at tasks that involve nouns with clear morphophonological patterns than those that involve opaque nouns. This finding holds with a variety of methodologies, including grammaticality judgments (Franceschina, 2005), oral production data (Fernández-García, 1999; Finnemann, 1992), picture-selection tasks (Cain, Weber-Olsen, & Smith, 1987; Franceschina, 2005; White, Valenzuela, Kozlowska-MacGregor, & Leung, 2004), written production data (Schlig, 2003), sentence completion (Alarcón, 2005), and cloze tasks (Franceschina, 2005).

It has, however, been suggested that some of L2 learners' difficulty in correctly using gender marking may stem from the fact that they are insensitive to morphosyntactic agreement markers, including gender markers (e.g., Guillelmon & Grosjean, 2001; Lew-Williams & Fernald, 2010). For instance, Lew-Williams and Fernald (2010) investigated whether NSs and L2 learners of Spanish would use grammatical gender agreement to orient to a target picture more quickly. When NSs heard a gender-marked article *la* (feminine) or *el* (masculine), they used that information to find a picture with the corresponding gender more quickly. However, nonnative speakers (NNSs) of Spanish did not.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Thus, to a large extent, research shows that learners are sensitive to morphophonological cues but insensitive to morphosyntactic ones. However, there are a few limitations to the current research. Generally, the use of morphosyntactic cues has been investigated with preferential-looking tasks or measurements of speed and accuracy of production of lexical items. These tasks, unfortunately, cannot give us enough data to do a fine-grained analysis of learners' online processing of gender. For instance, using pictures, as has been done in many studies (e.g., in the L1: Dahan, Swingley, Tanenhaus, & Magnuson, 2000; in the L2: Grüter, Lew-Williams, & Fernald, 2011; Lew-Williams & Fernald, 2010), does not provide a means to examine learners' use of morphophonological cues as opposed to morphosyntactic cues. Additionally, these measures deal with extremely subtle differences in processing speed and compare whether learners' processing of agreement information is as fast as that of NSs. The use of a different methodology allows the examination of these same issues under a different time frame and with a different focus.

There are also some limitations to the research on the use of morphophonological cues in the L2 acquisition of gender. This issue has generally been investigated by examining learners' accuracy on agreement marking in production, which can be affected by a variety of factors. An eye-tracker can provide more detailed information about learners' processing of these gender cues online and allows us to better understand the focus of learner attention through eye movements.

We, then, have one main research question: Do learners use morphosyntactic cues, morphophonological cues, or both to determine gender when marking agreement under various conditions? As previously noted, we investigate this question with an eye-tracker. We, specifically, ask the following two subquestions:

1. Do learners' eye movements indicate more reading time on and regressions (i.e., looks back) to morphosyntactic or morphophonological cues to gender?
2. Do learners' eye movements vary according to the noun's gender (i.e., masculine or feminine), transparency (i.e., transparent or opaque), or familiarity (i.e., familiar or novel)?

To address these research questions, we measure learners' eye movements to morphosyntactic cues via articles and morphophonological cues via noun endings. On the basis of previous research, we predicted that learners would demonstrate longer gaze times on and more regressions to nouns than articles. Because transparent nouns have clear cues to gender in the noun endings, we predicted that this effect would be stronger for transparent than for opaque nouns. We also predicted that overall there would be longer looking times and more

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

regressions to unfamiliar nouns and their associated articles than to familiar nouns and their associated articles because learners may have already acquired the gender for many familiar nouns. Finally, because masculine is often thought to be a default gender in many Romance languages (e.g., White et al., 2004), we predicted that looking time and regressions may be lower for masculine nouns than for feminine nouns.

We conducted two experiments (Experiment 1 and Experiment 2) to examine these research questions. In each experiment, we also manipulated the eye-tracker display (i.e., font size and layout) to examine the effect on our results. Thus, we also have a methodological research question: Does manipulating the eye-tracker display lead to differing results for our research questions regarding the L2 acquisition of Italian gender?

METHOD

Participants

Twenty-five participants took part in Experiment 1, and 20 of them were kept for analysis. One NS of Portuguese and 4 participants from fourth-semester Italian were removed from the participant pool, leaving only English-speaking learners of Italian enrolled in second-semester university courses. All of these speakers had studied L2s (e.g., Spanish, French, Italian) in elementary, middle, or high school. Five NSs also served as a control group. The 4 participants from fourth-semester Italian were not considered in this study to ensure comparability of participants between the two experiments. The results from the NSs are not presented in this article.

In Experiment 2, there were 20 learners of Italian, all NSs of English enrolled in second-semester Italian classes. As in Experiment 1, all had studied L2s previously in elementary, middle, or high school. In Table 2, information regarding the participants from both experiments is summarized.

Materials

Included in these studies were eight masculine transparent nouns, eight masculine opaque nouns, eight feminine transparent nouns, and eight feminine opaque nouns. To ensure that the students would be familiar with the nouns in the familiar category, a list of nouns was selected from the students' textbook (Merlonghi, Merlonghi, Tursi, & O'Connor, 2006). This list was given to a group of students similar to those used in the current study who were asked to rank the vocabulary in terms of familiarity from 1 to 3, 3 being the most

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Table 2. Participants

Experiment	*N*	*M* age	*N* males	Previously studied languages
Experiment 1	20	20.15	8	Spanish, 12 participants French, 6 participants Italian, 4 participants Latin, 1 participant German, 1 participant
Experiment 2	20	19.7	9	Spanish, 15 participants French, 5 participants Latin, 3 participants German, 2 participants Chinese, 2 participants Arabic, 1 participant

familiar. The most familiar nouns (M = 2.73) were then selected and controlled across conditions for familiarity, $F(7, 63) = 1.49$, $p = .19$. Nouns were also controlled across gender and transparency conditions for number of syllables (two or three), $F(7, 63) = 0.095$, $p = .998$. In addition to these 32 familiar nouns, 32 novel nouns were included. Loosely following Stark and McClelland (2000), the novel nouns were created by placing syllables from different nouns together (e.g., *fiume* "river" and *paese* "country" were used to form the novel word *fiuse*). These 64 nouns were presented in both singular and plural (only singular results are reported here), for a total of 128 slides. The nouns used in the studies can be found in Appendix A. The adjectives that learners marked for gender were taken from the first eight chapters in the first-year Italian textbook. All adjectives had transparent endings; no adjectives with opaque endings (e.g., *verde* "green") were included. The adjectives included in the studies can be found in Appendix B.

Design

Apparatus. Both experiments were conducted on an Eye Link 1000 eye-tracker with a desktop mount. Viewing was binocular, but only movements from the right eye were recorded. The eye-tracker was interfaced with a Hewlett-Packard computer with a 20-in. high-resolution color monitor. Participants were seated approximately 80 cm from the screen. The experiments were written with Experiment Builder software (SR Research Ltd.). Sentences appeared in black upper- and lowercase letters on a light gray background.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Experiment 1 Design. Research demonstrates that readers frequently skip over short (i.e., two- to three-letter) words (Frenck-Mestre, 2005). For this reason, in Experiment 1, we decided to print the words in a large font (Calibri 44 pt.) on three separate lines to capture the eye movements of the participants to articles, which for singulars are only two letters long. The sentences appeared on the left side of the screen across three lines. The article appeared on the first line and the noun appeared on the second line so that we could clearly distinguish whether participants fixated on the article or the noun. The adjective choices appeared on the right side of the screen (see Figure 1). Learners were instructed to read each sentence and then to click on the appropriate adjective choice; the choices varied as to number and gender.

We had anticipated that the participants might favor the top-left side of the screen when reading, so we added adverbs of varying length to the beginnings of the sentences so that our items would not appear at the top left. Additionally, to prevent participants from becoming accustomed to seeing each lexical item in a certain place, which might lead to strategies for eye movements, we added the prenominal adjective *grande* "big" to half of the slides to make the placement of items less predictable. Figure 2 provides an example of a slide in which *grande* was used.

In Experiment 1, there were three interest areas. One was around the article on line 1. It extended both to the left and to the right by approximately 1 cm to capture eye movements to that area that did not fall directly on the article. A second interest area was around the noun on line 2 (with a similar extension), and a third was around the adjective choices on the right. The interest areas are shown in Figure 3.

Experiment 2 Design. The screen layout and font in Experiment 2 was designed to be more ecologically valid than in Experiment 1. First, the

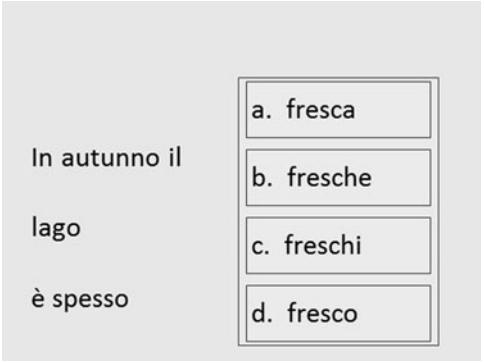


Figure 1. Exact screen layout (including font and font size) for Experiment 1 ("In autumn, the lake is often refreshing").

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

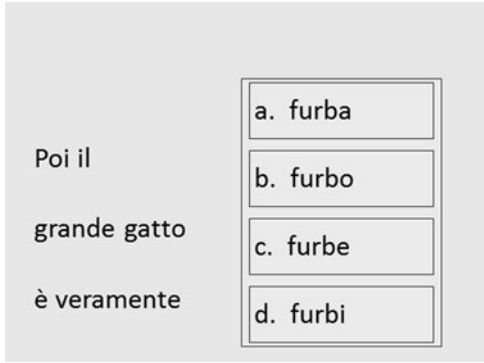


Figure 2. Exact screen layout (including font and font size) for Experiment 1 using the adjective *grande* ("Then, the big cat is really sly").

font size was reduced to Calibri 24 pt. To make the text more closely aligned to normal reading material, entire sentences appeared across only one line at the top of the screen, rather than being broken up over several lines. The adjective choices appeared below the sentence, across the bottom of the screen. See Figure 4 for the display in Experiment 2. Note that the three interest areas, which were drawn similarly to those in Experiment 1, are also shown.

Procedure

After completing a background questionnaire, the participants were seated in front of the eye-tracker. Participants were instructed to place their

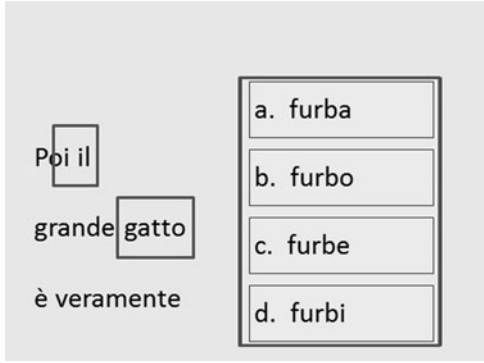


Figure 3. Exact screen layout (including font and font size) showing interest area placement for Experiment 1 ("Then, the big cat is really sly").

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

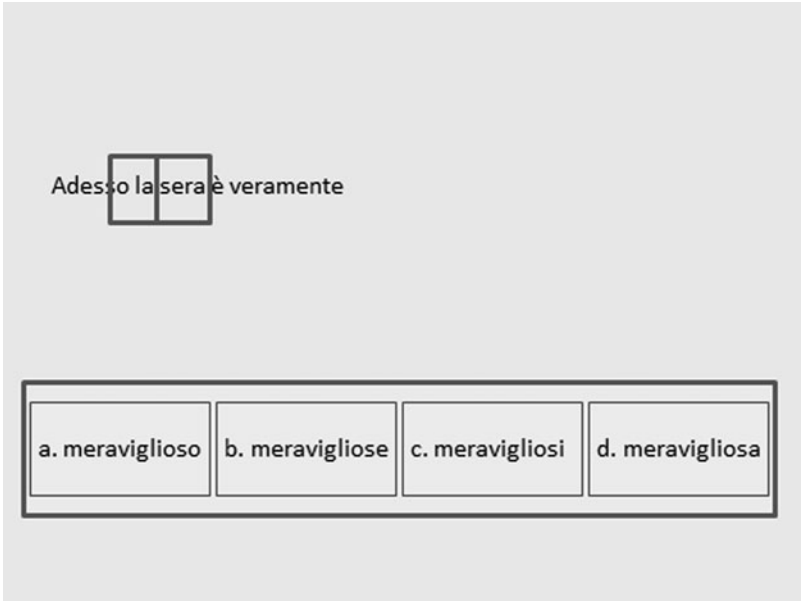


Figure 4. Exact screen layout (including font and font size) showing interest area placement for Experiment 2 ("Now, the evening is really wonderful.").

foreheads and chins against the desktop mount and to focus on the fixation target when it appeared on the screen. They were also instructed to click with the mouse on the word (i.e., adjective choice) that they felt best completed each sentence. The 128 slides were divided into three blocks, and the order of the three blocks and of the slides within the blocks was randomized for each participant. Breaks were given after each block. A 9-point calibration and a 9-point calibration accuracy test were given at the beginning of the experiment and after each break. Participants read the sentences silently at their own pace; the program advanced to the next slide when the participants clicked on their adjective choice with the mouse provided. If participants did not click on one of the adjective choices, the item timed out after 10 s and moved on to the next slide. The entire study for most participants took about 15–18 min.

Data Analysis

We calculated the fixations on and regressions to two interest areas on each slide: the interest area around the article and the interest area around the noun. Only accurate responses to items containing singular

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

nouns were included in these data analyses. Learners were 86.9% accurate (*SD* = 18.0) on singular nouns in Experiment 1 and 90.4% accurate (*SD* = 29.5) on singular nouns in Experiment 2.

RESULTS

Because the assumption of independence of variables required for ANOVA was not met and to account for the variance that was introduced by the varying length of the nouns and articles used in the study, we used a linear mixed model. We were therefore able to ascertain the amount of time participants spent looking at the noun as compared to the article, in terms of total dwell time.

In Experiment 1, whether a fixation occurred on a noun or an article was not a significant factor in the total dwell time, $\beta = -365.86$, $t(4.05) = -1.53$, $p = .199$. That is, participants did not spend significantly more time looking at nouns than at articles. This finding held for all conditions and is based on an analysis with marginal means comparisons with Sidak adjustments for multiple comparisons. Specifically, participants did not look significantly longer at masculine nouns than at masculine articles (*M* difference = 280.44 ms, *SE* = 233.48, $p = .301$) nor significantly longer at feminine nouns than at feminine articles (*M* difference = 234.85 ms, *SE* = 234.24, $p = .376$). Participants also did not look longer at transparent nouns than at articles appearing with transparent nouns (*M* difference = 247.37 ms, *SE* = 233.49, $p = .353$) nor at opaque nouns than at articles appearing with opaque nouns (*M* difference = 267.91 ms, *SE* = 234.09, $p = .320$). Finally, they did not look longer at novel nouns than at articles appearing with novel nouns (*M* difference = 378.38 ms, *SE* = 233.73, $p = .186$) nor at familiar nouns than at articles appearing with familiar nouns (*M* difference = 136.91 ms, *SE* = 233.79, $p = .592$).

However, if one considers the overall amount of time learners looked at nouns and articles, there were differences between the conditions. Noun transparency was a significant factor, in that learners looked longer at articles and nouns in the opaque condition than at those in the transparent condition, $\beta = -121.51$, $t(2{,}169.43) = -2.71$, $p = .007$ (articles: *M* difference = 100.98 ms, *SE* = 43.76, $p = .021$; nouns: *M* difference = 121.52 ms, *SE* = 44.79, $p = .007$). Noun familiarity was also a significant factor for nouns, in that learners looked longer at nouns in the novel condition than in the familiar condition, $\beta = -224.54$, $t(2{,}165.72) = -5.12$, $p < .001$ (*M* difference = 224.54 ms, *SE* = 43.87, $p < .001$). They did not look longer at articles in the novel condition than in the familiar condition (*M* difference = −16.93 ms, *SE* = 43.02, $p = .694$). The gender of the noun did not affect gaze times, $\beta = 14.37$, $t(1{,}993.38) = 0.31$, $p = .755$ (articles: *M* difference = 31.21 ms, *SE* = 43.11, $p = .469$; nouns: *M* difference = 14.37 ms, *SE* = 46.01, $p = .755$). See Figures 5 and 6.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

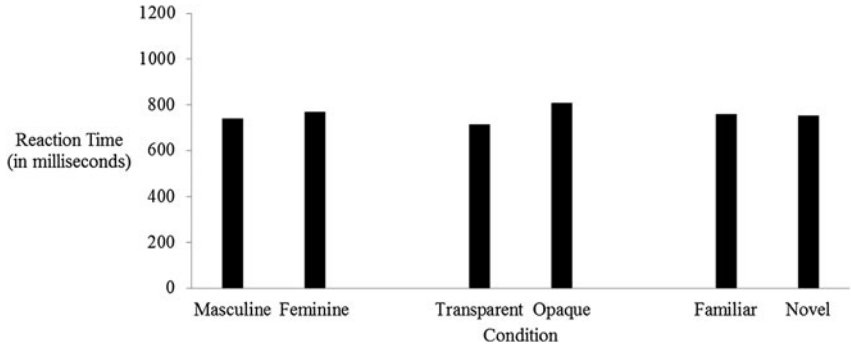


Figure 5. Experiment 1 dwell time means for articles.

Unlike in Experiment 1, in Experiment 2, whether a fixation occurred on a noun or on an article predicted a significant amount of the variation in total dwell time, $\beta = -784.68$, $t(5.31) = -4.36$, $p = .006$. That is, learners looked at nouns longer than they looked at articles. This finding held throughout all conditions, as measured by marginal means comparisons with Sidak adjustments for multiple comparisons. Specifically, participants looked significantly longer at masculine nouns than at masculine articles (*M* difference = 692.47 ms, *SE* = 175.04, $p = .012$) and significantly longer at feminine nouns than at feminine articles (*M* difference = 606.71 ms, *SE* = 175.92, $p = .019$). Participants also looked longer at transparent nouns than at articles appearing with transparent nouns (*M* difference = 546.25 ms, *SE* = 175.16, $p = .028$) and at opaque nouns than at articles appearing with opaque nouns (*M* difference = 752.94 ms, *SE* = 175.63, $p = .008$). Finally, they looked longer at novel nouns than at articles appearing with novel nouns (*M* difference = 724.22 ms, *SE* = 175.23, $p = .010$) and at familiar nouns than at articles appearing with familiar nouns (*M* difference = 574.97 ms, *SE* = 175.45, $p = .024$).

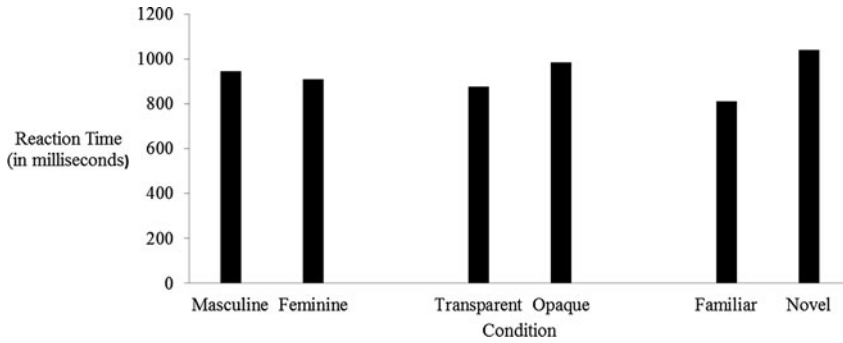


Figure 6. Experiment 1 dwell time means for nouns.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

As in Experiment 1, there were differences between the conditions when one considers overall gaze times on both nouns and articles. These differences were similar to those found in Experiment 1. Noun transparency was a significant factor in that learners looked longer at nouns, but not articles, in the opaque condition than in the transparent condition, β = –251.47, $t(2,263.26) = -6.56$, $p < .001$ (articles: *M* difference = 44.78 ms, *SE* = 37.50, $p = .233$; nouns: *M* difference = 251.47 ms, *SE* = 38.34, $p < .001$). Noun familiarity was also a significant factor in that learners looked longer at nouns, but not articles, in the novel condition than in the familiar condition, β = –171.63, $t(2,267.98) = -4.53$, $p < .001$ (articles: *M* difference = 22.38 ms, *SE* = 37.25, $p = .548$; nouns: *M* difference = 171.63 ms, *SE* = 37.89, $p < .001$). Unlike in Experiment 1, the gender of the noun did affect gaze times, but only the gaze time of nouns, with participants looking at masculine nouns longer than at feminine nouns, β = 82.94, $t(2,014.14) = -4.53$, $p < .001$ (articles: *M* difference = 2.82 ms, *SE* = 37.26, $p = .940$; nouns: *M* difference = 82.94 ms, *SE* = 39.80, $p = .037$). See Figures 7 and 8.

The second measure was the number of regressions to nouns and articles after the first pass through the text. In Experiment 1, there were approximately the same number of regressions to nouns and articles: 401 regressions to nouns and 378 regressions to articles. In contrast, in Experiment 2, there were markedly fewer regressions to articles recorded: 340 regressions to nouns and 65 regressions to articles. These results are displayed in Figures 9 and 10.

In Table 3 we present a summary of the results from the two experiments.

DISCUSSION

Research Questions on the L2 Acquisition of Gender

The primary concern of this article is to determine whether altering display conditions significantly affects the results and conclusions of an

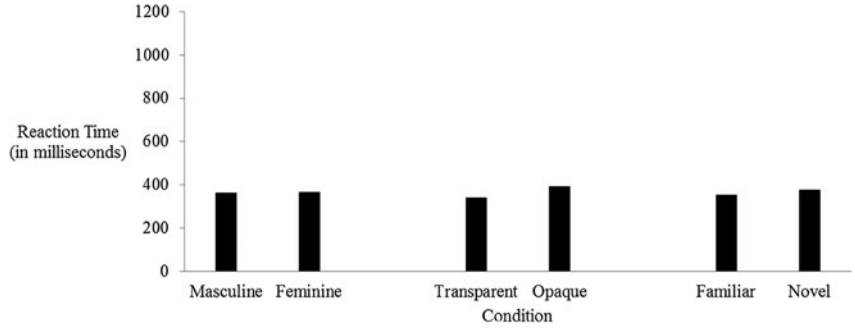


Figure 7. Experiment 2 dwell time means for articles.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

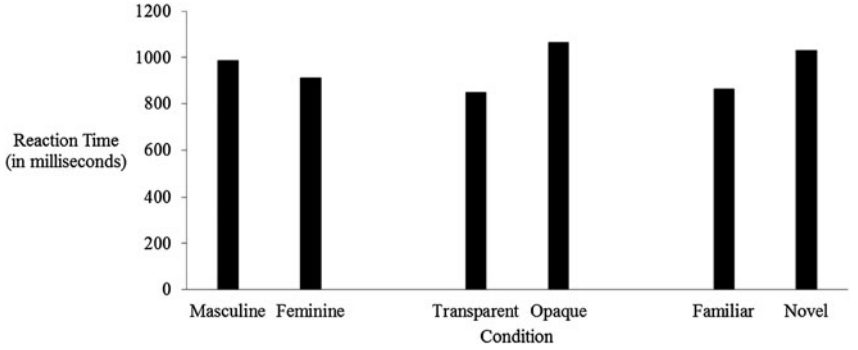


Figure 8. Experiment 2 dwell time means for nouns.

eye-tracking investigation into the L2 acquisition of gender. To this end, an examination of the conclusions that can be drawn from each experiment is now provided.

The first research question asked whether learners' eye movements indicate more reading time on and regressions (i.e., looks back) to morphosyntactic or morphophonological cues to gender. In Experiment 1, it was found that an equal amount of time was spent on nouns and articles during the gender-assignment task, and regressions to nouns and articles were roughly equal. This finding suggests that learners employ both morphophonological cues (i.e., word endings) and morphosyntactic cues (i.e., articles) when marking gender agreement in the online task. At first, these results may seem to run counter to previous studies that indicate that learners are relatively insensitive to gender agreement but are sensitive to morphophonological markers on nouns. However, the current methodology addresses a different aspect of L2 gender acquisition than has been the case in previous work. Recall that research demonstrating learners' insensitivity to gender agreement has generally relied on subtle differences in reaction times on picture-selection tasks or word-production tasks. These methods indicate that learners do not

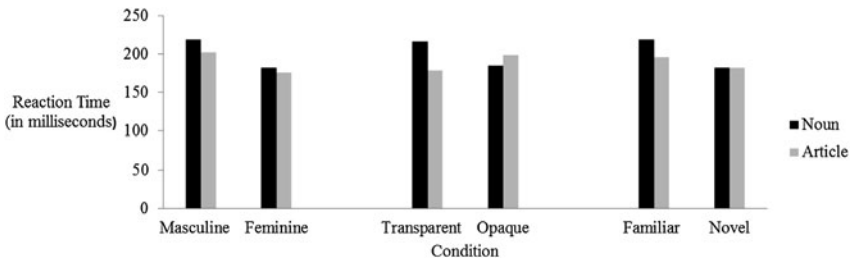


Figure 9. Regressions to nouns and articles in various conditions in Experiment 1.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

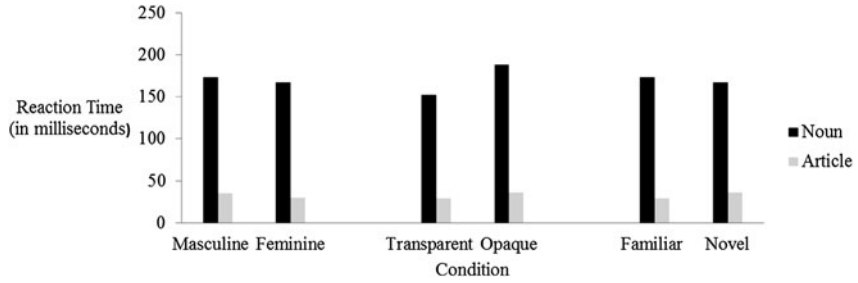


Figure 10. Regressions to nouns and articles in various conditions in Experiment 2.

process gender agreement quickly enough to employ it in these particular tasks. On the other hand, the current design allows learners to complete the task at their own pace, which gives them the opportunity to make use of all available linguistic information. That is, the current results address the question of whether or not learners rely on morphosyntactic cues to gender under conditions that are less time sensitive. It is not surprising to find that learners are able to make use of both articles and noun endings to mark gender agreement accurately under these conditions.

In contrast, an analysis of the results of Experiment 2 leads to different conclusions. Learners spend much less time looking at articles than at

Table 3. Summary of results

Measure	Experiment 1	Experiment 2
Time looking at nouns vs. articles	**Nonsignificant**	**Significant**
Masculine nouns vs. masculine articles	**Nonsignificant**	**Significant**
Feminine nouns vs. feminine articles	**Nonsignificant**	**Significant**
Transparent nouns vs. transparent articles	**Nonsignificant**	**Significant**
Opaque nouns vs. opaque articles	**Nonsignificant**	**Significant**
Familiar nouns vs. familiar articles	**Nonsignificant**	**Significant**
Novel nouns vs. novel articles	**Nonsignificant**	**Significant**
Conditions		
Masculine articles vs. feminine articles	Nonsignificant	Nonsignificant
Masculine nouns vs. feminine nouns	**Nonsignificant**	**Significant**
Transparent articles vs. opaque articles	**Significant**	**Nonsignificant**
Transparent nouns vs. opaque nouns	Significant	Significant
Familiar articles vs. novel articles	Nonsignificant	Nonsignificant
Familiar nouns vs. novel nouns	Significant	Significant

Note. The bold indicates different results between Experiment 1 and Experiment 2.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

nouns. Their eye movements also regress to articles far less often than to nouns. These results support the conclusion that, even when given ample time to use any available linguistic information, learners often fail to use morphosyntactic cues to gender, relying more heavily on morphophonological cues such as noun endings or attempting to activate their lexical knowledge regarding each noun's gender. Thus, the results of Experiments 1 and 2 lead to far different answers to our first research question. This issue is addressed further in the subsequent discussion.

Our second research question asked whether learners' eye movements vary according to the noun's gender (i.e., masculine or feminine), transparency (i.e., transparent or opaque), or familiarity (i.e., familiar or novel). In contrast to the results regarding the first research question, the results regarding the second research question are generally consistent with expectations. In both studies, more time was spent on novel nouns than on familiar nouns, a finding that could be attributed to the fact that learners spend more time searching their lexicons for a match to a novel word—a match they, of course, cannot find. In Experiment 2, results showed that learners also spent more time looking at articles with novel nouns, presumably because learners are unable to access any stored knowledge about the gender of the noun and, thus, are more reliant on the syntactic cue. Additionally, more time is spent looking at both articles and nouns when the noun is opaque. The reason that learners spend more time on opaque nouns than on transparent nouns may be similar to the reason that they spend more time looking at novel nouns; they may be searching for a cue to gender but may fail to find one. This is consistent with research by Dussias et al. (this issue) and Hernandez et al. (2004), who propose that the processing of opaque nouns requires more effort and processing depth than the processing of transparent nouns. Because the noun does not have a cue to gender, the only source for a gender cue is the article. The fact that learners must look at the article for a cue to gender with opaque nouns may explain why they look at articles longer in these cases, as well. The only finding somewhat contrary to expectations was in Experiment 2, in which the gender of the noun was a significant factor. Specifically, learners looked longer at masculine than at feminine nouns. This finding is harder to explain due to the fact that masculine gender is often considered a default gender in languages such as Spanish and Italian (e.g., White et al. 2004), and, thus, learners might be expected to spend less time viewing those nouns.

The methodology research question asked whether manipulating the eye-tracker display would lead to differing results for our research questions regarding the L2 acquisition of Italian gender. The main difference in results between Experiments 1 and 2 is in the reading times on and regressions to nouns versus articles. Recall that Experiment 1 was specially designed to capture eye movements on small items—specifically

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

articles—that are immediately followed by the noun they modify. To prevent having two adjacent interest areas, we separated the article from the noun; to capture movements to the article, which are very small lexical items, we used a large font. In Experiment 2 we used a more ecologically valid design that involved (a) having the noun and article on the same line (and thus adjacent) and (b) using a smaller font. We captured roughly the same amount of reading time on and regressions to the nouns in both studies, a finding that indicates that under certain circumstances, an ecologically valid design may be just as effective as our specially designed format. However, with the ecologically valid design, we captured markedly fewer looks to articles, as measured both by reading time and by regressions. There are at least two possible interpretations of these results: Either the findings from Experiment 1 are more accurate, or the findings from Experiment 2 are more accurate.

We argue that the findings of our specially designed layout (Experiment 1) are more accurate than the findings from Experiment 2. In particular, we believe that the ecologically valid design fails to capture many looks to the article because it is possible to see the article while viewing nearby words (such as the noun). In fact, previous research shows that small words such as articles are often skipped during reading, even by NSs (Dussias, 2010). Indeed, an examination of the recordings of eye movements of individual readers indicated that eye gazes often fell in between nouns and articles, or slightly ahead, above, or below the articles. For example, in Figure 11 from Experiment 2, the eye fixation (represented by a dot) is on the beginning of *quaderno* ("notebook") and is recorded as a fixation on the noun. Nonetheless, it is clear that both the article and the noun are captured within that one gaze. Thus, learners were likely able to perceive the article and to use the morphosyntactic cue even though the eye movement was not recorded on the article. On the other hand, in Experiment 1, the conditions were designed to maximize the measurability of looks to the article. With this design, we were able to capture far more looks to the article, better representing the actual behavior of the participants (see Figure 12). In other words, in Experiment 1, we were able to obtain a microscopic view of learners' use of available material as they determine noun gender, a necessary factor in selecting the appropriate agreeing adjective.

We also argue that the design of Experiment 1 is more revealing for this study because the use of the article was necessary under certain circumstances to mark the correct response. Both of the current experiments made use of unfamiliar nouns that make it impossible to rely on stored knowledge of gender and opaque nouns that make it impossible to rely on noun ending alone to determine gender. For the nouns that are both opaque and unfamiliar, learners have no way of knowing the

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

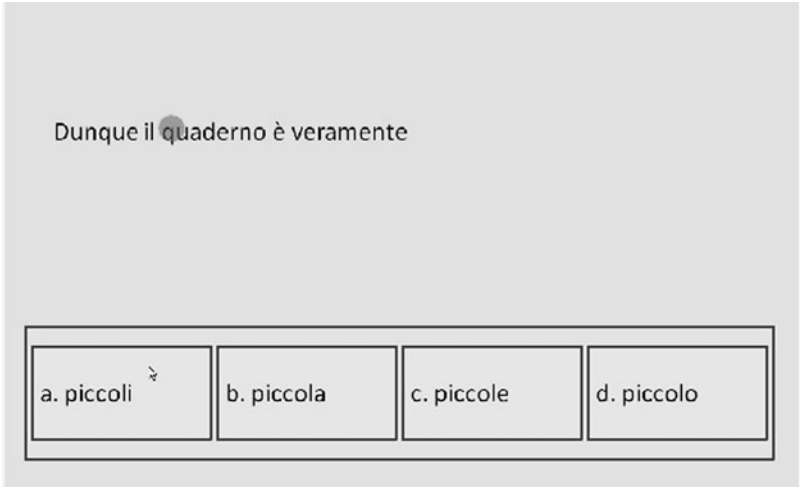


Figure 11. Screenshot of eye fixation in Experiment 2 ("So, the notebook is really small.").

gender of the noun except by using the agreement of the article. On slides that contained opaque, unfamiliar nouns, learners were 76.9% accurate in Experiment 1 (*SD* = 42.2) and 84.7% accurate in Experiment 2 (*SD* = 36.1). In Experiment 1, the eye-tracker captured an average of 794 ms (*SD* = 691 ms) of reading time on the article in each sentence with an opaque unfamiliar noun, but in Experiment 2, the eye-tracker captured an average of only 399 ms on each article in these conditions (*SD* = 501 ms). In both cases, only accurate responses have been analyzed because, presumably, to mark the correct answer, learners would have needed to observe the article, but the eye-tracker in Experiment 2 does not capture as many of these eye movements. In fact, in many cases,

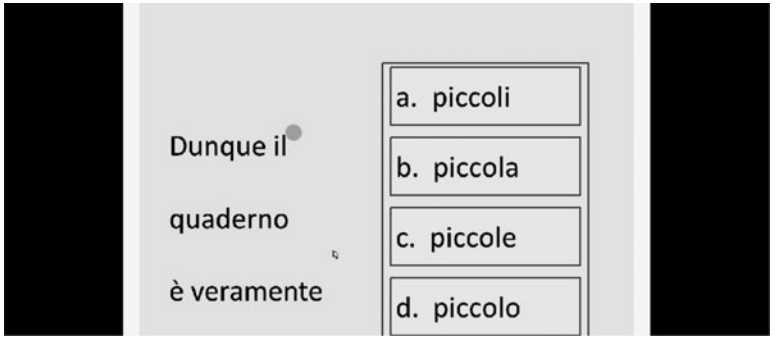


Figure 12. Screenshot of eye fixation in Experiment 1 ("So, the notebook is really small.").

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

participants marked the correct answer without registering any eye movements to the article at all.

Eye-Tracking Methodology

Based on these findings, we make two main points. First, methodology matters. Small changes in textual arrangement, font size, and other display features can lead to large changes in results. In fact, researchers using the two different designs we examined would come to opposite conclusions regarding learners' use of morphosyntactic and morphophonological cues to gender. For this reason, it is crucial that information regarding details of screen presentation be carefully presented in published articles when using eye-trackers as a tool for elicitation (see Polio & Gass, 1997, for a discussion of the important role of reporting and Brown, 2012, and Mackey, 2012, for discussions of insufficiency in reporting).

Second, we argue that, for some research questions, it may be necessary to design studies that are not ecologically valid in the sense of attempting to mimic natural reading conditions. In Experiment 1, for example, it was necessary to break the text up over several lines and use a large font to capture eye movements to small grammatical functors. However, it was not the intent of this study to examine natural reading conditions of L2 learners. Rather, the goal was to examine learners' use of gender cues in an online task that is intended not to replicate communicative behavior but rather to reveal learners' sensitivity to and use of morphosyntactic and morphophonological information. No conclusions are drawn regarding L2 reading behavior.

One argument against these conclusions is that Experiment 1 might have artificially inflated the number of looks to articles. That is, it has been found that under normal conditions, readers' eyes skip over articles. By forcing learners to look directly at the article to obtain the morphosyntactic gender cue, we have created an unnatural situation. For this reason, it is important to restrict the interpretation of the results. That is, one should not argue that under normal reading conditions learners will focus directly on articles. However, the point is that under normal reading conditions learners will be able to pick up the gender cue on articles while viewing nearby words. The first study design was created to capture those views that cannot ordinarily be observed.

LIMITATIONS AND CONCLUSIONS

Of course, there are limitations to this methodological study. First, as we have noted, the task is a relatively artificial one. Clicking on an

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

agreeing adjective from a bank of four choices is a metalinguistic endeavor, and learners might have been able to accommodate to the task over time, becoming more efficient at achieving accurate results. Second, only two possible study designs were examined out of a nearly limitless variety of possible designs. Finally, and most importantly, we cannot truly be sure which of the designs produced the "correct" result. Although we argue that the design in Experiment 1 more accurately reflects learner behavior, there is no way to conclusively determine which design is more appropriate.

At the very least, however, this article has demonstrated the importance of attention to detail and careful reporting of conditions in eye-tracking studies. Fortunately, use of this methodology is becoming more widespread, and, in time, there will be a larger body of work from which to draw study designs. In the meantime, it is imperative that this work provide enough detail for future researchers to replicate and expand upon.

NOTES

1. Our search was limited to empirical studies that were published in refereed journals. We used the Linguistics and Language Behavior Abstracts (LLBA) database, Google Scholar, and two review articles (Frenck-Mestre, 2005; Dussias, 2010) as our sources.

2. A cursory search for eye-tracking studies in L1 research yielded similar findings in terms of detail provided. Notable exceptions were studies by Braze, Shankweiler, Ni, and Palumbo (2002); Carreiras and Clifton (1999); Ni, Fodor, Crain, and Shankweiler (1988); Rayner and Duffy (1986); and Ehrlich and Rayner (1981), in which mention was made of either characters per line or words per text (and number of lines).

3. There are exceptions, such as *mano* "hand," which is feminine despite an *-o* ending, and *problema* "problem," which is masculine despite the *-a* ending. These are relatively infrequent and are not dealt with in this article.

4. Franck, Vigliocco, Antón-Méndez, Collina, and Frauenfelder (2008) state that 80% of nouns in Italian are transparent, 17% of masculine nouns are opaque, and 19% of feminine nouns are opaque.

REFERENCES

Alarcón, I. (2005). *The second language acquisition of Spanish gender agreement: The effects of linguistic variables on accuracy* (Unpublished doctoral dissertation). Indiana University, Bloomington.

Altarriba, J., Kroll, J. F., Scholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye-fixation and naming times. *Memory & Cognition*, *24*, 477–492.

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. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*, 633–660.

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, *118*, 245–257.

Bolger, P., & Zapata, G. (2011). Semantic categories and context in L2 vocabulary learning. *Language Learning*, *61*, 614–646.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1090–1101.
Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*, 299–314.
Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research*, *24*, 139–166.
Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York: Routledge.
Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Reader's eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31*, 25–44.
Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. New York: Cambridge University Press.
Brown, J. D. (2012). Writing up a replication report. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 173–197). New York: Cambridge University Press.
Cain, J., Weber-Olsen, M., & Smith, R. (1987). Acquisition strategies in a first and second language: Are they the same? *Journal of Child Language*, *14*, 333–352.
Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, *27*, 826–833.
Carroll, S. E. (1999). Input and SLA: Adults' sensitivity to different sorts of cues to French gender. *Language Learning*, *49*, 37–92.
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.
Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford: Oxford University Press.
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.
Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149–166.
Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116.
Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (this issue). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*.
Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 663–679.
Erlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Reading Research Quarterly*, *16*, 227–235.
Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, *12*, 485–502.
Fernández-García, M. (1999). Patterns of gender agreement in the speech of second language learners. In J. Gutiérrez-Rexach & F. Martínez-Gil (Eds.), *Advances in Hispanic linguistics: Papers from the 2nd Hispanic Linguistics Symposium* (pp. 3–15). Somerville, MA: Cascadilla Press.
Finnemann, M. D. (1992). Learning agreement in the noun phrase: The strategies of three first-year Spanish students. *International Review of Applied Linguistics*, *30*, 121–136.
Flecken, M. (2011). Event conceptualization by early Dutch-German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism: Language and Cognition*, *14*, 61–77.
Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: Benjamins.
Franck, J., Vigliocco, G., Antón-Mendéz, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*, 329–374.
Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research, 21*, 175–198.
Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Erlbaum.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Gass, S. M., & Mackey, A. (2007). *Data elicitation for second and foreign language research.* Mahwah, NJ: Erlbaum.
Grüter, T., Lew-Williams, C., & Fernald, A. (2011). Grammatical gender in L2: Where is the problem? In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 246–258). Somerville, MA: Cascadilla Press.
Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, *29*, 503–511.
Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*.
Hernandez, A. E., Kotz, S. A., Hoffman, J., Valentin, V. V., Dapretto, M., & Bookheimer, S. Y. (2004). The neural correlates of grammatical gender decisions in Spanish. *Brain Imaging*, *1*, 863–866.
Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, *15*, 314–318.
Kaushanskaya, M., & Marian, V. (2007). Bilingual language processing and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning*, *57*, 119–163.
Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*, 503–535.
Leeser, M., & Sunderman, D. (2012). *Methodological issues of working memory tasks for L2 processing research*. Manuscript submitted for publication.
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.
Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Bristol, UK: Multilingual Matters.
Mackey, A. (2012). Why (or why not), when and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 34–69). New York: Cambridge University Press.
Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Mahwah, NJ: Erlbaum.
Mackey, A., & Gass, S. M. (2012). *Research methods in second language acquisition: A practical guide*. Oxford: Blackwell.
Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages: A native/non-native asymmetry in bilinguals. *Journal of Psycholinguistic Research*, *37*, 141–170.
Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115.
McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. London: Routledge.
Merlonghi, F., Merlonghi, F., Tursi, F., & O'Connor, B. (2006). *Oggi in Italia: A first course in Italian* (8th ed.). Boston, MA: Wadsworth, Cengage Learning.
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.
Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515–539.
Nunan, D. (1992). *Research methods in language learning*. New York: Cambridge University Press.
Oller, J., & Tullius, J. (1973). Reading skills of non-native speakers of English. *International Review of Applied Linguistics in Language Teaching*, *11*, 69–79.
O'Rourke, B. (2008). The other C in CMC: What alternative data sources can tell us about text-based synchronous computer mediated communication and language learning. *Computer Assisted Language Learning*, *21*, 227–251.
Polio, C., & Gass, S. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition*, *19*, 499–508.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*, 113–127.
Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*, 333–357.
Schachter, J., & Gass, S. (Eds.). (1996). *Second language classroom research: Issues and opportunities.* Mahwah, NJ: Erlbaum.
Schlig, C. (2003). Analysis of agreement errors made by third-year students. *Hispania*, *86*, 312–319.
Schmiedtová, B. (2011). Do L2 speakers think in the L1 when speaking in the L2? *Vigo International Journal of Applied Linguistics*, *8*, 139–179.
Seliger, H. W., & Shohamy, E. (1990). *Second language research methods*. Oxford: Oxford University Press.
Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers. *Second Language Research*, *27*, 251–272.
Stark, C. E., & McClelland, J. L. (2000). Repetition priming for words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 945–972.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.
Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*, 88–107.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.
White, L., Valenzuela, E., Kozlowska-MacGregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*, 105–133.

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press

APPENDIX A

NOUNS USED IN EXPERIMENTS 1 AND 2

Familiar				Novel			
Masculine		Feminine		Masculine		Feminine	
Trans.	Opaq.	Trans.	Opaq.	Trans.	Opaq.	Trans.	Opaq.
quaderno	giornale	festa	vocale	giolo	nore	cuvira	pante
libro	bicchiere	sera	carne	vorno	pecale	vercita	catente
volo	nome	spiaggia	legge	merceggio	biale	cadia	nonte
viaggio	paese	casa	gente	vicato	macchiesce	riduna	stacale
formaggio	mare	mattina	notte	gago	fiuse	seltina	lorne
mercato	pesce	rivista	stagione	forderno	runase	spitira	racane
lago	fiume	chiesa	ragione	quamaro	giormose	fiesa	vogione
gatto	rumore	cucina	canzone	liggio	mame	sesa	gete

Note. Trans. = transparent; opaq. = opaque.

APPENDIX B

ADJECTIVES USED IN EXPERIMENTS 1 AND 2

Masculine				Feminine			
piccolo	buono	straniero	alto	favolosa	fredda	brutta	vicina
rosso	povero	vecchio	magro	bella	grassa	prossima	meravigliosa
pieno	fresco	diverso	basso	affollata	vera	cattiva	impegnata
sicuro	furbo	ricco	noioso	gialla	intera	brava	antipatica

https://doi.org/10.1017/S0272263112000927 Published online by Cambridge University Press