imü dilbilimi

SELF-PACED READING TESTS WITH PSYCHOPY: A VERSATILE TOOL FOR LINGUISTIC DATA COLLECTION

Engin Evrim ÖNEM
Dr., Erciyes University, Department of Basic English, School of Foreign Languages, eonem@erciyes..edu.tr, ORCID: 0000-0002-2711-7511

Önem, E. E. (2025). Self-paced reading tests with PsychoPy: A versatile tool for linguistic data collection. In O. Çınar, F. Başbuğ & H. Aydemir (Eds.), *Contemporary studies in linguistics I: IMU Linguistics 15th anniversary commemorative volume* (pp. 289–304). Artsürem. https://doi.org/10.7816/imuling-15-2025-01X015

ABSTRACT

This chapter examines the application of self-paced reading (SPR) tests using PsychoPy software in linguistic research, highlighting the method's ability to provide precise measurements of the cognitive processes involved in language comprehension. It reviews existing studies utilizing PsychoPy to investigate phenomena such as syntactic ambiguity resolution and discourse processing. Furthermore, the chapter proposes a methodological framework for conducting SPR experiments, offering specific guidelines for task design, stimulus presentation, and data analysis. By evaluating the advantages and limitations of this approach, this work aims to contribute to the methodological toolkit of linguists and researchers interested in real-time language processing.

Keywords: Self-paced reading, PsychoPy, Linguistics

ÖZ

Bu bölümde, PsychoPy yazılımı kullanılarak uygulanan kendi hızında okuma testlerinin dilbilimsel araştırmalarda kullanımı ve bu yöntemin dil anlama sürecindeki bilişsel süreçlerin ölçülmesine olan katkıları incelenmektedir. Ayrıca bu bölüm, sözdizimsel belirsizliklerin çözümü ve söylem işlemleme gibi çeşitli olguları araştırmak için PsychoPy kullanan mevcut çalışmaları ortaya koymaktadır. Ek olarak, test tasarımı, uyaran sunumu ve veri analizi konularında rehberlik sunarak, PsychoPy ile kendi hızında okuma deneyleri yürütmek için yöntemsel bir çerçeve önerilmektedir. PscyhoPy yazılımının faydalarını ve sınırlılıklarını sunan bu çalışma, dil işlemleme ile ilgilenen dilbilimcilerin ve araştırmacıların yöntembilimsel araç setine katkıda bulunmayı amaçlamaktadır.

Anahtar Sözcükler: Kendi hızında okuma, PsychoPy, Dilbilim

1. Introduction

Language is a system reflecting various interconnected cognitive processes in the human mind. Due to the close connection and relationship between language production and comprehension in the cognitive system, "selection and inhibition of verbal or non-verbal input stimuli are comparable to the executive mechanisms involved in the activation of semantic and/or lexical representations" (Turgeon & Macoir, 2008: 10). Although some early as well as some modern psycholinguistic studies focus on language deficiencies, modern methods have been proposed to analyze these mechanisms. As Ullman (2013) summarized, these methods include, but are not limited to, a wide variety of data collection tools, from medical imaging equipment to written tests, questionnaires, and judgment calls. All these methods actually focus on reaction, which is the voluntary or involuntary response to an external stimulus requiring complex processes. The duration spent reacting to any stimulus can be used to measure complex cognitive processes. As Balakrishnan et al. (2014) stated, reaction time is the difference in time between the presentation of a stimulus and the inception of the response to that stimulus. In other words, it is the duration from stimulus onset to voluntary response.

Francis Galton is considered the first to study simple reaction time, as he investigated teenagers' reaction times to light and sound stimuli in the 19th century (Woods et al., 2015). Since Galton, psycholinguistic studies have developed various methodologies utilizing reaction time, including simple reaction time, recognition, choice, and serial reaction time studies to study language processing (Baayen & Milin, 2010; Woods et al., 2015). Early experimental psychology utilized reaction time (RT) measurements to infer cognitive processes, as it was widely accepted that longer RTs indicated more complex or demanding processing, or vice versa. This established a crucial methodological framework for measuring temporal aspects of cognition. While basic RT tasks often involved simple stimuli and responses, reading studies required more sophisticated methods to track the processing of continuous text. As Jegerski (2013) stated, along with the advances in computers in the 1970s, researchers applied the principles of RT studies to language processing, and the self-paced reading (SPR) technique emerged to measure the time course of language processing. Also, in line with the development of psychometrics in the realm of psycholinguistic research, efficient and flexible tools have become important. In this sense, PsychoPy has stood out as a powerful and open-source software suite designed to construct and execute a wide array of psychological experiments, including those crucial to understanding language processing. Its versatility has cemented its place in the academic toolkit as it offers a platform for creating and running self-paced reading experiments. In essence, PsychoPy provides a comprehensive toolkit for psycholinguists, enabling the precise presentation of linguistic stimuli and the accurate recording of reaction and reading times, empowering researchers to delve deeper into the intricacies of language processing.

This chapter aims to explore the application, advantages, and disadvantages of SPR with PsychoPy, review existing studies, and provide a methodological framework for future research. However, this chapter should not be confused with a guide to designing and running self-paced reading (SPR) experiments on PsychoPy. However, if needed, researchers can find detailed tutorials and documents on creating and running self-paced reading tests with PsychoPy on the dedicated website, https://www.psychopy.org/.

2. Self-Paced Reading: Principles and Methodology

2.1 The underlying principles of self-paced reading

The self-paced reading (SPR) technique is an effective way to deconstruct the intricate, moment-by-moment processes of language comprehension through the presentation of texts or segments in a serial manner. SPR is not just a question of what one knows but how one knows it at the moment. As Marsden et al. (2018) discuss, SPR is an online, computer-based experimental method that presents sentences divided into single words or meaningful phrases. Participants, being active controllers of the reading pace, move through the text by pressing a designated key. Notably, the system automatically records the time intervals between every key press and provides a precise measure of the cognitive effort exerted at each section.

The foundational philosophy of SPR, which involves deducing cognitive processing from reading time patterns, is straightforward yet deeply perceptive. As Jegerski (2013) puts it, "the eyes can be a window on cognition" (4), which highlights that individuals' reading activity, and more particularly the amount of time spent processing each word or sentence, has a direct correlation to the cognitive load imposed on their cognition. As participants control the presentation rate of text, SPR paradigms heavily rely on reading time data as a direct index of the cognitive processing load (Marinis, 2010). This data reveals a participant's sensitivity to and awareness of the linguistic phenomenon being researched, as well as their working memory capacity, which is responsible for holding and manipulating information during reading (Marsden et al., 2018). In fact, SPR's strongest point is its ability to detect processing issues through fluctuations in reading time. Measuring reading times for each segment marks trends in relation to language processing. Decreased reading times typically suggest processing ease, which means the linguistic input is aligned with the reader's expectations or requires minimal cognitive effort. Conversely, longer reading times reflect processing difficulty, indicating that the reader is experiencing difficulty with unexpected or complex linguistic information. It is this sensitivity that renders SPR particularly handy for investigating the impact of ambiguities, anomalies, or complicated linguistic dependencies on language processing. For example, a sudden increase in reading time on a specific word may

indicate that the word creates an unexpected syntactic structure or semantic deviation that requires more cognitive resources to process. By closely observing reading time patterns, researchers can gain a fine-grained understanding of the cognitive operations involved in syntactically parsing and interpreting language, which enables them to draw firm conclusions about the nature of online language comprehension.

2.2 Paradigms for self-paced reading

As the designation suggests, the focus of the study is to ascertain the time required to read the linguistic data displayed on the monitor. A number of paradigms of self-paced reading are available, on which researchers can rely to present and analyze linguistic stimuli. These are most often cumulative and non-cumulative. As Luke & Christianson (2013) describe, under the cumulative paradigm, parts of the linguistic stimuli (clause, phrase, or sentence) are displayed as the participant types any key, and each part is shown until the entire stimulus is displayed on the screen. Conversely, under the non-cumulative model, a new segment appears on the screen when the button is pressed, and the prior segment is hidden. Another distinction in how linguistic stimuli are provided is based on the segment's position on the screen (Jegerski, 2013). The segment(s) can be positioned at the middle of the screen (centered) or presented in a sequence format, from left to right (linear). According to Marinis (2010), when segments are centrally presented, participants are unable to estimate the length of the sentence, which provides a clearer view of the processing. Furthermore, according to Jegerski (2013), since linear presentation is similar to reading real sentences, it may be better to present stimuli in this manner. These advantages have led to the development of the moving window method, which presents segments in a linear and noncumulative manner, similar to conventional reading. Such differences in presentation and procedural considerations also underlie the flexibility inherent in self-paced reading paradigms, which enable the investigation of the temporal dynamics of language processing.

With this range of methodological choices, researchers can tailor self-paced reading paradigms by combining presentation methods and other measures to address specific research questions on language processing. For example, in their research, Luke & Christianson (2013) created a paradigm that combined self-paced reading with masked priming to enable the investigation of sentence context effects on early word recognition. They explained in the conclusion of their paper that their paradigm accurately mimicked results from masked-priming and self-paced reading literature. Gong et al. (2023) introduced a new paradigm, a forward-and-backward mode, by combining the self-paced reading technique, which typically requires reading forward, with the eye-tracking technique, which enables participants to move forward and backward as needed, resulting in successful outcomes in psycholinguistics. Ultimately, it can be affirmed that SPR paradigms are easily adaptable to the needs of

researchers, and this is the primary reason why the self-paced reading technique has been adopted in the majority of studies, as Jegerski (2013) notes.

2.3 Advantages and disadvantages of SPR

Since it is one of the online (or real-time) data gathering techniques, there are some distinct advantages of self-paced reading compared to traditional offline psycholinguistic data collection techniques, such as questionnaires or sentence completion tasks. Firstly, online tests are implicit. For instance, live collection tools are resistant to the metalinguistic abilities of participants because such tools record the unconscious and habitual responses of the participants to what they are engaged in (Marinis, 2010; Jegerski, 2013). However, participants would demonstrate controlled and conscious decisions by performing a slow deliberation over an offline or non-real-time activity at hand, utilizing their explicit knowledge of language and metalinguistic ability (Marinis, 2010; Jegerski, 2013). In fact, implicit grammar rules are limited in SPR (Jegerski, 2013), suggesting that SPR is a more accurate reflection of language processing. Secondly, SPR offers a finer measurement of processing. Marinis (2010) claims that because the reading time of each word, phrase, or clause can be independently measured, SPR easily furnishes information on processing difficulty or ease in individual segments. Greater reading time, as mentioned above, points to processing difficulty, and vice versa. Thirdly, SPR is very simple to administer and inexpensive. In particular, no elaborate participant training or ongoing researcher supervision is required, as it is pretty simple in most cases for the participants: read in chunks and press buttons to proceed. There is also no need for experimental equipment, as only a computer is needed. If other psycholinguistic measures are considered, such advantages are apparent.

Conversely, the self-paced reading technique has some drawbacks, particularly when compared to traditional reading. Firstly, the potential for exhibiting pseudo-reading patterns and their effect on reading times is an issue of ecological validity (Jegerski, 2013). As mentioned above, although items are presented linearly, as in regular reading, they are actually offered in a segmented form, which is far from the normal reading experience. Therefore, this can result in unintended processing load and can interfere with the data of reading time. Habit formation is another potential drawback. Participants in an SPR experiment must repeatedly press a button, and this physical response can lead to automatic reactions over time, which is unnatural in the context of regular reading. Jegerski (2013) also emphasizes the importance of participant fluency in reading for SPR paradigms. Unless readers are native or native-like speakers in the language being tested, reading times may be misleading as to the amount of time spent on reading.

Yet, all these drawbacks are remediable. For example, Jegerski (2013) reported that barely any of the subjects in one of her experiments

had trouble reading in chunks, and button pressing was already a habit for most young subjects. Additionally, breaking after some time and/or limiting the time for SPR experiments can solve fatigue and habit formation problems. Finally, although reading fluency is required for SPR experiments, differences in the native language and second/foreign language processing have been studied (for an extensive review of studies, see Marsden et al., 2018). In other words, non-fluency does exist in linguistic research carried out through the self-paced reading method. In conclusion, although SPR has its limitations, it remains a valuable and widely applied methodology in psycholinguistic research, as it offers high temporal resolution and provides fine-grained information about the cognitive processes underlying language comprehension.

3. PsychoPy: A Powerful Tool for SPR Implementation

PsychoPy is open-source software that allows researchers to simplify and extend it, as it is programmed in Python, a free programming language, and freely distributed software. The free computer language simplifies personalizing the software to integrate with other software and websites, supporting online data collection. The software originated as a psychophysics tool but has since expanded significantly to support complex language-based paradigms. Early releases, as seen by Peirce (2007), laid the groundwork for what it is now, providing scientists with a flexible platform that enables fine control of experimental stimuli and data acquisition. At the same time, the software offers an open space for collaboration, where scientists continuously contribute to its development, maintaining it at the forefront of experimental psychology software. That is, PsychoPy is an ever-changing software that continues to expand, providing researchers more time to ponder their experimental design rather than spending time struggling with complex programming.

Some intrinsic features of PsychoPy make it suitable for self-reading experiments. Firstly, PsychoPy offers precise timing records and flexible control of display. The time taken between the appearance of each linguistic stimulus on the screen and the target key being pressed is recorded in milliseconds. PsychoPy's accuracy in measuring reaction times is also crucial for language experiments, such as lexical decision, ambiguity resolution, or sentence judgment tasks. The software captures participant responses, whether button presses or vocal, along with the duration from stimulus onset to response. Additionally, the segment provided can be easily manipulated in PsychoPy. In the context of self-paced reading (SPR) experiments, PsychoPy is especially well-suited to present sentence fragments with precision. It can be configured to present text word by word or phrase by phrase, allowing participants to read at their own pace. This allows for reading times for each fragment to be investigated in minute detail, providing invaluable information on cognitive processes in language understanding.

Second, PsychoPy's use extends beyond its accurate control and timing, which are critical because it is a powerful tool for presenting stimuli in psycholinguistic research. According to Peirce (2007), PsychoPy enables the creation and manipulation of various types of stimuli, including text, images, and sound files. This is particularly crucial in language research, where researchers often need to present rich textual stimuli, manipulate linguistic factors, and control stimulus presentation timing to the millisecond. Peirce (2009) describes the software's capacity to generate diverse stimuli, a critical feature that allows researchers to create dynamic and interactive testing environments.

Another critical feature of PsychoPy is its convenience. The relaxed or adaptable nature of PsychoPy makes it easier for researchers to design experiments. For example, while experienced computer users or "people who know about programming" would be most advantaged by PsychoPy, there is another mode called the Builder view (Peirce et al., 2019), where a researcher can choose the type of stimulus, the input type for the response that is expected and alter preferences such as the time for any response or what kind of data to save or not save just like they would on Windows® or Mac® OS. Therefore, the simplicity of using PsychoPy to design complex experiments is apparent (Peirce et al., 2019). In fact, this feature is also discussed in detail in Peirce et al. (2022), which provides a step-by-step tutorial on designing experiments with the PsychoPy system.

Another fundamental feature of PsychoPy is the ability to export and log data. Since both the presentation of the stimulus and the recording of reaction time are critical in self-paced reading, PsychoPy takes accurate logs of data and renders them exportable across a range of formats (e.g., PDF, XLSX, CSV, etc.) so that researchers can analyze them freely with their preferred programs. As Peirce et al. (2019) continue to note, the power of PsychoPy lies in its ability to bridge the gap between high-level experimental design and implementable code that can be tailored to meet the specific needs of the researchers.

Finally, note should be taken of the observation that PsychoPy is not a single software to be used in psycholinguistic research. There are a relatively large number of computer programs available for use in psycholinguistic research, including E-Prime®, NBS Presentation®, Psychophysics Toolbox, OpenSesame, Expyriment, Gorilla, jsPsych, Lab.js, and Testable. However, when comparing the precision of visual and auditory stimulus timing and response times provided by various software, Bridges et al. (2020) asserted that PsychoPy performed much more reliably in both laboratory and online studies than other toolkits. However, they also highlighted researchers' adoption of specific timing validation procedures for their experimental design. In addition, Gallant & Libben (2019) provided the primary benefits of PsychoPy, including greater research efficacy, higher ecological validity, and the ability to recruit more representative and historically underrepresented participant samples. They considered the central aspect of millisecond timing in this regard and

offered helpful technical details and recommendations. Therefore, although numerous software packages are available for psycholinguistic experiments, the core functionality and timing accuracy demonstrated by PsychoPy make it a highly trustworthy option, albeit one that still needs to be tested on an individual basis for specific experimental paradigms.

4. Applications of SPR with PsychoPy in Linguistic Research

Many studies employ SPR with PsychoPy to investigate various phenomena in linguistics. Some studies examined syntactic ambiguity resolution and semantic interpretation. For example, Hassim & Kukona (2024) examined the effect of extended cognitive control on processing syntactically ambiguous "garden path" sentences using two experiments on PsychoPy and concluded that a high proportion of incongruent Stroop trials enhanced extended cognitive control, which in turn had a causal influence on improving sentence comprehension, whether the sentence was written or heard. In a further study, Kyriacou et al. (2023) emphasized the resolution of passivization ambiguity in passivized idioms using an eye-tracking task. In addition to the eye-tracing task, they also conducted a lexical decision task using PsychoPy with keywords and the same number of pseudowords, all of which were the same length as the keywords, to ensure that any differences found in control keyword processing were attributed to their contextual inconsistency rather than to other lexical features. In the end, they discovered that keywords were responded to significantly faster than pseudowords, and response times for the keywords did not have a substantial impact on the condition (whether they were in a figurative, literal, or control situation). Jansen et al. (2017) used PsychoPy and a self-paced reading test to study semantic priming on homonymous nouns and found that reaction time and error rate were identical and unaffected by semantic association, whether absent or present, in both ambiguous and unambiguous conditions. They argued that the non-selective access hypothesis was prejudiced in favor of both senses of the processed word being activated.

Some other studies utilized PsychoPy and SPR to examine semantic integration and anomaly detection. For instance, Berger et al. (2022) used PsychoPy to investigate how merely seeing a task cue, compared to performing the cued task, affects unconscious semantic priming based on task-set dominance. The findings revealed that the mere presence of a cue resulted in inhibition of the reversed priming task sets compared to performance on the cued task, indicating that the particular timing of activation and inhibition of the task set following cue presentation could be controlled by certain experimental conditions. In a further study, Mifka-Profozic et al. (2020) examined the influence of syntactic violation and semantic ambiguity on the processing of the modal auxiliaries "can" and "may" in agent-oriented, epistemic, and speaker-oriented contexts. Reading penalties following incongruent modal use were observed in agent-oriented and epistemic contexts, but speaker-oriented

modality processing remained uninfluenced by the inconsistency of formality. PsychoPy is also used in discourse processing research. For example, Wetzel et al. (2022) investigated the processing of native speakers of French causal and concessive sentences with appropriate or inadequate discourse connectives, as well as how connective frequency and polyfunctionality affected processing, based on three experiments. The results showed that processing was affected by incoherent elements irrespective of the connective and that reading fluency decreased with less frequent connectives, especially polyfunctional ones. Studies have shown that frequent connectives are convenient for reading; however, processing discourse is a robust process that allows readers to extract meaning despite infrequent connectives. Crible et al. (2021) studied whether native French speakers and French learners processed contrastive relations indicated by syntactic parallelism with and without discourse connectives in the same way in three experiments using PsychoPy. They examined the impact of parallelism with or without a connective, as well as with often versus rarely and unclear versus clear connectives. The findings stressed that while parallelism was essential to both groups, it was a more potent cue for non-native speakers and was influenced by task difficulty in native speakers.

It can be said that studies identify PsychoPy's capacity to facilitate rich explorations of language processing online, providing valuable insights into the cognitive processes underlying various linguistic phenomena in both native and non-native speakers, across different experimental settings, and even in second language studies. As stated by Hamrick (2022), since second language studies require extremely sensitive, reliable chronometric measures, PsychoPy and SPR are also relevant to second-language studies. Among many studies, Berghoff (2020) explored the processing of object-subject ambiguities by early second-language acquirers in a self-paced reading test administered on PsychoPy. The study reported that early childhood second-language acquirers in the L2 group relied more heavily on online reanalysis to process object-subject ambiguities. Banerjee et al. (2025) used PsychoPy and SPR to examine the effect of word priming on the response time of Bengali-speaking bilingual dyslexic children in identifying the target word in another study, reporting a positive impact of word priming on processing. Patterson & Nicklin (2023) analyzed the reliability and quality of online reaction time data in L2 studies by comparing the self-paced reading responses of crowdsourced workers with those of online students, using an in-person measure as a baseline. The outcome was that strong L1 effects were reproduced across both online populations; however, the data quality of online students was lower compared to that of crowdsourced participants. In conclusion, the various applications outlined above, along with many more, demonstrate PsychoPy's central role as a versatile and reliable tool for collecting self-paced reading data across a broad spectrum of linguistic research, ranging from sentence comprehension and lexical

ambiguity resolution to semantic processing, discourse analysis, and even second language acquisition.

5. Methodological Framework for SPR Experiments with PsychoPy

This section outlines the crucial steps in designing an SPR experiment with PsychoPy, including selecting an appropriate paradigm, presenting the stimuli, and discussing data analysis, interpretation, and potential methodological challenges.

First and foremost, selecting an appropriate SPR paradigm is crucial for eliciting the target cognitive processes and obtaining interpretable reaction times. The variations, as noted earlier, involve the moving window or cumulative paradigm. A researcher needs to select an appropriate paradigm for the linguistic structures with which they are working. For example, Ferreira & Clifton (1986) note that the moving window paradigm allows for high temporal resolution in experiments testing the fine-grained temporal properties of processing, such as the short-term influence of syntactic or semantic variation. Another critical feature of SPR design is controlling the amount and complexity of the text segments. Because reading time within every segment is being measured, sentences must be of equal length and makeup to avoid introducing inequalities in terms of complexity and length (Marinis, 2010). Target sentences and control sentences must be constructed in a way that they are syntactically similar, with similar word or character lengths per segment, so that variation in processing load due to variation in grammaticality is reduced. For example, longer sections of course take longer to read, and greater lexical frequency or syntactic complexity also influence reading times, independent of experimental manipulation (Rayner, 1998). Therefore, target and control items must also be equated on word frequency measures and psycholinguistic properties such as the number of syllables. Third, filler item development is also necessary for any SPR experiment with PsychoPy, as it prevents participants from becoming overly sensitive to the experimental manipulation, reduces the predictability of the target stimuli, and allows for an equal task demand throughout the experiment. According to Goodall (2021), fillers and the experimental items are also required to be counterbalanced on a ratio of 1:1. The filler items, being efficacious, need to be cleverly designed and semantically equatable, need to have an array of semantic content and syntactic structures to avoid allowing participants to develop target-specific processing procedures and need to be of identical length and difficulty as the target items.

Second, planning is required with care about stimulus presentation in SPR using PsychoPy. As noted above, the fundamental method of SPR is to present text section by section, typically after a participant presses a key. Researchers should ensure that the presentation remains uniform across trials and participants, which involves maintaining a consistent location for the text on the screen and providing distinct instructions for

participants to follow when executing the task (Jegerski, 2013). Another aspect to mention is the display of comprehension question(s) suited to linguistic stimuli. As Marinis (2010) argues, comprehension questions should be placed at the end of a segment or sentence so that participants do not form habitual or automatic answers by maintaining attention and readiness. Simultaneously, trials on the SPR should also be of only 15-20 minutes, such that fatigue cannot accumulate in the participants, with at least a week maintained between trials to prevent long-term memory recalls (Marinis, 2010; Jegerski, 2013). Another element that requires consideration is adjusting the text display. According to research, font characteristics such as font size, typeface, and spacing can have a significant impact on reading speed and comprehension. In an experiment, Banerjee et al. (2011) investigated the effect of font size and style on reading performance on computer screens and recommended using font sizes of 14-point for on-screen reading. They suggested Courier New for optimal speed and Verdana for visual quality and reduced mental load. Therefore, for SPR tasks, sans-serif fonts such as Arial, Verdana, Courier New, or Calibri, with a suitable font size (e.g., 12-14 points), are typically recommended for optimal readability on computer screens. A further consideration here is minimizing visual distraction for the SPR task. This includes having a plain background color (e.g., white or gray), excluding unnecessary items on the screen, and ensuring the test environment is free from external distractions (Jegerski, 2013). Finally, researchers should ensure adequate timing and synchronization by running a pilot test and factoring in the monitor's frame rate.

Data analysis and interpretation in SPR with PsychoPy involves several crucial steps: initial data cleaning, preprocessing, applying appropriate statistical methods, and interpreting reading time patterns in the context of linguistic hypotheses. Raw data exported from PsychoPy often require preprocessing and cleaning before they are ready for statistical analysis. This pre-processing involves the identification and exclusion of trials with technical errors (e.g., anticipatory key presses) or participant errors (e.g., incorrect answers to comprehension questions), as defined by researchers before data collection (Ferreira & Clifton, 1986). Outliers in the data, which could result from fatigue, transient lapses, or technical errors, must be excluded. Standard methods include visual inspection of the data, application of standard deviation-based cut-offs (e.g., truncating data points more than 2.5 or 3 standard deviations from a subject's mean under a condition), or using more robust methods, such as the median absolute deviation (Baayen & Milin, 2010). The final step in preprocessing data is the optional aggregation process, which depends on the question being asked. For instance, reading times to specific regions of interest within clauses or sentences might be summed up (e.g., adding reading times across several words within a significant area) to provide a more accurate overall estimate of processing effort for the area. SPR experiment reading time data tends to be investigated using statistical

methods that can account for the repeated measures nature of the data (multiple data points per participant and item). Linear mixed-effects models (LMMs), as Baayen et al. (2008) and Jegerski (2013) explain, have become increasingly popular and are well-suited to this type of data. LMMs enable the estimation of the effect of experimental manipulations (fixed effects) while accounting for variation between items and participants (random effects). Jaeger (2008) explains that fixed effects are experimental manipulations under investigation (e.g., syntactic ambiguity, semantic anomaly, or presence/absence of a discourse connective) and LMMs estimate the average effect of such manipulations on reading times. Random effects, in contrast, reflect non-independence at both the participant and item levels through random participant and item intercepts and slopes, allowing the model to adjust for individual differences in baseline reading speed and for sensitivity to differential amounts of experimental treatment across items (Jaeger, 2008). Parallel with aggregation, Jegerski (2013) states that if LMMs are utilized, computing aggregate means and conducting individual item analyses will no longer be required, and data trimming may be reduced or eliminated. On the other hand, when standard ANOVAs and t-tests are employed, the data need to be reduced and trimmed into aggregate means to account for subject and item variance, as necessitated by the need for item analysis in psycholinguistics.

Knowing the pattern of reading times is crucial for concluding cognitive processes and the linguistic phenomenon under study. First, the variation in reading times for particular segments can be attributed to the ease or difficulty of processing. As mentioned previously, longer reading times or a slowing down of reading speed for a specific segment indicate increased difficulty in processing, and vice versa. Additionally, processing difficulty at a point within the sentence may not be immediately apparent. Still, it can instead be detected in the reading times of subsequent words or phrases, known as spillover. This could indicate that the cognitive processes elicited by a manipulation continue to influence processing in following areas (Rayner, 1998). Furthermore, the effect of a linguistic manipulation could be confounded with other factors, such as between-participant differences in working memory capacity (Just & Carpenter, 1992) or usage frequency of a particular linguistic construction (Chater & Manning, 2006), and can be controlled using LMMs. Significantly, failing to find noticeable differences in reading times across conditions can be too revealing, as these findings could be informative, suggesting that the manipulated linguistic factor has no significant impact on online processing, as measured by reading times.

6. Addressing Potential Methodological Challenges

There are several key considerations to take into account when designing stable self-paced reading (SPR) experiments in PsychoPy, including addressing individual variation, controlling for practice effects

and fatigue, and ensuring adherence to ethical practices in participant handling. First, SPR data are marked by extreme individual variation in reading speed, which is highly expected. Nevertheless, as mentioned above, linear mixed-effects models can also be utilized to handle this variability by incorporating random effects for participants (Baayen et al., 2008), allowing researchers to model experimental condition manipulations while holding individual differences in baseline reading times constant. Another potential SPR difficulty is the impact of practice effects and fatigue. Following Jegerski's (2013) suggestion, practice trials should be included at the start of the experiment to allow participants to adapt to the task, and these steps from practice trials are typically excluded from analysis. In addition, as time elapses from the experiment's commencement, participants may start to tire or lose concentration, resulting in progressively slower and less consistent reading times. To minimize such effects, short breaks between blocks of trials (Marinis, 2010), maintaining the experiment duration between 15 and 20 minutes (Marinis, 2010; Jegerski, 2013), and using a sufficient number of filler items (Goodall, 2021) can be employed. Finally, as ethical considerations are essential in any study involving human subjects, researchers must take note of key ethical requirements, including participants' informed consent, confidentiality, anonymity, and clearance from an ethics committee.

7. Conclusion

The self-paced reading (SPR) technique is an advantageous method for examining the time course of language processing, providing information about cognitive processes that are not easily achieved with other methods. PsychoPy, with its precise timing, flexible stimulus presentation, and easy-to-use interface, provides a solid foundation for conducting SPR experiments. This chapter has discussed the SPR method, as implemented in the PsychoPy program, with an emphasis on its utility as a functional tool for linguistic studies. This chapter outlines the basic principles of SPR, explaining its various paradigms and emphasizing the importance of factors such as stimulus control, data analysis, and the investigation of potential artifacts. A survey of experiments conducted with SPR in PsychoPy to demonstrate the wide range of the technique's potential contributions to different areas of linguistic research, such as syntactic ambiguity resolution, semantic interpretation, discourse processing, and second language acquisition. Finally, the chapter's methodological framework offers researchers a blueprint for designing successful SPR experiments in PsychoPy. By paying close attention to paradigm choice, stimulus design, data analysis techniques, and control of potential confounds, researchers can leverage the strengths of SPR and PsychoPy to inform the complexities of language processing. In short, SPR, as facilitated by PsychoPy, is a valuable and growing tool for linguists seeking to unravel the mystery of how people process language in real-time.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28. https://doi.org/10.21500/20112084.807

Balakrishnan, G., Uppinakudru, G., Singh, G. G., Bangera, S., Raghavendra, A. D., & Thangavel, D. (2014). A comparative study on visual choice reaction time for different colors in females. *Neurology Research International, 2014*, Article 301473. https://doi.org/10.1155/2014/301473

Banerjee, J., Majumdar, D., Pal, M. S., & Majumdar, D. (2011). Readability, subjective preference and mental workload studies on young Indian adults for selection of optimum font type and size during onscreen reading. *The American Journal of the Medical Sciences, 4*(2), 131–143.

Banerjee, J., Nithin, K. S., Sai, U. H. K., Chakraborty, B., & Basu, A. (2025). Exploration of semantic priming studies in native language of children: Study on Bengali L1 and English L2 bilingual children. *SN Computer Science, 6*, Article 119. https://doi.org/10.1007/s42979-025-03664-4

Berger, A., Kunde, W., & Kiefer, M. (2022). Task cue influences on lexical decision performance and masked semantic priming effects: The role of cue–task compatibility. *Attention, Perception, & Psychophysics, 84*, 2684–2701. https://doi.org/10.3758/s13414-022-02568-2

Berghoff, R. (2020). The processing of object–subject ambiguities in early second-language acquirers. *Applied Psycholinguistics, 41*(4), 963–992. https://doi.org/10.1017/S0142716420000314

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414. https://doi.org/10.7717/peerj.9414

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335–344. https://doi.org/10.1016/j.tics.2006.05.006

Crible, L., Wetzel, M., & Zufferey, S. (2021). Lexical and structural cues to discourse processing in first and second language. *Frontiers in Psychology, 12*, Article 685491. https://doi.org/10.3389/fpsyg.2021.685491

Ferreira, F., & Clifton Jr., C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Gallant, J., & Libben, G. (2019). No lab, no problem: Designing lexical comprehension and production experiments using PsychoPy3. *The Mental Lexicon, 14*(1), 152–168. https://doi.org/10.1075/ml.00002.gal

Gong, T., Gao, X., & Jiang, T. (2023). A "dummy's" program for self-paced forward and backward reading. *Behavior Research Methods, 55*, 4419–4436. https://doi.org/10.3758/s13428-022-02025-w

Goodall, G. (2021). Sentence acceptability experiments: What, how, and why. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 7–38). Cambridge University Press.

Hamrick, P. (2022). Conducting reaction time research in second language psycholinguistics. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics*. https://doi.org/10.4324/9781003018872

Hassim, N., & Kukona, A. (2024). Linking cognitive control to language comprehension: Proportion congruency effects in syntactic ambiguity resolution. *Language, Cognition and Neuroscience, 39*(4), 431–447. https://doi.org/10.1080/23273798.2024.2314027

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jansen, M. T., Jansen, N. C., Weber, A., Gadea, G. H., Ansari, E., & Scheren, P. (2017). Semantic priming with homonymous nouns: Hints of clarifying the issue of selective vs. non-selective priming. *Journal of European Psychology Students, 8*(1), 15–29. https://doi.org/10.5334/jeps.408

Jegerski, J. (2013). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Kyriacou, M., Conklin, K., & Thompson, D. (2023). Ambiguity resolution in passivized idioms: Is there a shift in the most likely interpretation? *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 77*(3), 212–226. https://doi.org/10.1037/cep0000300

Luke, S. G., & Christianson, K. (2013). SPaM: A combined self-paced reading and masked-priming paradigm. *Behavior Research Methods, 45*, 143–150. https://doi.org/10.3758/s13428-012-0239-4

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). John Benjamins. https://doi.org/10.1075/lllt.27.09mar

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics, 39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Mifka-Profozic, N., O'Reilly, D., & Guo, J. (2020). Sensitivity to syntactic violation and semantic ambiguity in English modal verbs: A self-paced reading study. *Applied Psycholinguistics, 41*(5), 1017–1043. https://doi.org/10.1017/S0142716420000338

Patterson, A. S., & Nicklin, C. (2023). L2 self-paced reading data collection across three contexts: In-person, online, and crowdsourcing. *Research Methods in Applied Linguistics, 2*(1), 100045. https://doi.org/10.1016/j.rmal.2023.100045

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, Article 10. https://doi.org/10.3389/neuro.11.010.2008

Peirce, J. W., Hirst, R., & MacAskill, M. (2022). *Building experiments in PsychoPy* (2nd ed.). Sage.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Turgeon, Y., & Macoir, J. (2008). Classical and contemporary assessment of aphasia and acquired disorders of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 3–11). Elsevier.

Ullman, M. T. (2013). Language and the brain. In R. Fasold & J. Connor-Linton (Eds.), *An introduction to language and linguistics* (6th ed., pp. 235–274). Cambridge University Press.

Wetzel, M., Zufferey, S., & Gygax, P. (2022). How robust is discourse processing for native readers? The role of connectives and the coherence relations they convey. *Frontiers in Psychology, 13*, Article 822151. https://doi.org/10.3389/fpsyg.2022.822151

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience, 9*, Article 131. https://doi.org/10.3389/fnhum.2015.00131