Please cite as：

Suzuki, Y., & Elgort, I. (2023). Measuring automaticity in a second language: A methodological synthesis of experimental tasks over three decades (1990-2021). In Y. Suzuki (Ed.), Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology (pp. 206-234). New York, NY: Routledge.

https://www.taylorfrancis.com/chapters/edit/10.4324/9781003414643-12/measuring-automaticity-second-language-comprehension-yuichi-suzuki-irina-elgort

Chapter 9 Measuring automaticity in second language comprehension: A methodological synthesis of experimental tasks over three decades (1990-2021)

Yuichi Suzuki
Irina Elgort

Abstract

This chapter reports a methodological synthesis of experimental tasks used in research of automaticity in second language (L2) comprehension. Our survey yielded 34 lexical and 46 grammatical tasks (e.g., primed/non-primed lexical decision, semantic/acceptability judgement, picture-sentence matching, comprehension with eye-tracking, self-paced reading, word-monitoring tasks), which we classified task based on the types of processes being investigated. We synthesized key outcome measures (accuracy, reaction time, and/or coefficient of variability) and the software (e.g., E-prime, DMDX) used in different task types. Although this work identified many psycholinguistic tasks available for L2 researchers to investigate automaticity, it also revealed several gaps in the L2 research into automaticity, most notably the scarcity of tasks for assessing automaticity in auditory lexical processing. We also present methodological guidelines on how to select experimental tasks for assessing automaticity in L2 studies.

Introduction

Second language (L2) researchers commonly investigate target language proficiency in terms of receptive (reading and listening) and productive skills (speaking and writing). A key indicator of proficiency in L2 comprehension and production is automaticity of processing. While adult first-language (L1) speakers have developed highly efficient language processing mechanisms through intensive and extensive experience from birth, L2 learners' language processing is less efficient, especially at the earlier developmental stages. Yet, achieving a higher degree of automaticity in visual and auditory input processing is needed to free up cognitive resources and engage in meaningfocused L2 comprehension and production. Because automatic language processing requires efficient access to relevant aspects of knowledge, which cannot be measured directly, we conceptualize automaticity as one of the key properties of processing that can inform us about the quality of knowledge (e.g., DeKeyser, 2009).

Recently, methodological issues have received focal attention in the field of applied linguistics (AL) and second language acquisition (SLA) (Gass et al., 2021; Marsden et al., 2018; Plonsky et al., 2020). In order to select methodologies that are aligned with research goals, it is essential for SLA researchers to understand which experimental tasks are used to investigate different L2 component processes and knowledge, and how these tasks are designed. Given that automaticity has high theoretical and practical relevance for L2 acquisition and use (for review, see e.g., DeKeyser, 2001; Segalowitz, 2003; Segalowitz & Hulstijn, 2005), a variety of experimental tasks have been proposed to evaluate automaticity over the decades. However, to our knowledge, no comprehensive systematic review of these tasks is available to date. To this end, this synthesis aims to identify and describe the domain of research on automaticity and automatization in SLA from a methodological perspective.

The goal of our systematic review is to synthesize the kinds of experimental tasks available for L2 researchers who wish to study automaticity. We have surveyed AL and SLA journal articles that report measures of automatization and automaticity. The chapter's focus is on lexical and grammatical processing skills in L2 comprehension (see production measures in Suzuki & Révész, this volume). We also present methodological guidelines for selecting experimental tasks to assess automaticity in L2 processing and skills as well as exercises on interpreting experimental design.

L2 Comprehension Model: Processes, Knowledge, and Automaticity

L2 comprehension requires an orchestration of multiple processing components. Figure 9.1 illustrates the construct of automaticity in relation to the L2 comprehension processes and the knowledge these processes draw on. The model is based on several major models of reading comprehension (e.g., Grabe & Stoller, 2011; Perfetti & Stafura, 2014) and listening comprehension (e.g., Field, 2013; Vandergrift & Goh, 2012). In this model, L2 comprehension relies on the ability to decode visual or auditory signal, recognize words and access their meanings, parse morpho-syntactic structures, and interpret and infer meaning in relation to the reader/listener's representation of the text (discourse) and general (including non-linguistic) knowledge. This division may be somewhat artificial, but it is useful for developing experimental tasks for evaluating quality of reading and listening comprehension and diagnosing issues that arise in L2 processing and learning.

There are three integral elements in the proposed model (Figure 9.1): (a) knowledge; (b) processes; and (c) automaticity. Knowledge includes orthographic, phonological, lexical (including single words and multi-word expressions), grammatical (linguistic regularities such as inflections and morpho-syntax), pragmatic, and discourse knowledge (as well as general knowledge about the world). These various types of knowledge underlie comprehension processes; in turn, comprehension of input contributes to knowledge development.

Processes in comprehension may be grouped into lower-level (decoding, word-level processing, and sentence-level processing) and higher-level processes (comprehension of a given text/discourse). In the model, the locus of the decoding stage of comprehension is sublexical. At the decoding stage in reading, for example, visual input (i.e., printed letters/characters) is processedfirst, activating orthographic representations and leading to the activation of related phonological information and auditory representations. In listening, spoken input is decoded first, with the physical acoustic information being matched to listeners' representational knowledge, leading to the activation of orthographic representations in literate individuals. While either the grapheme decoding or phoneme decoding process is prioritized in reading and listening, respectively, at the decoding stages, both orthographic and phonological knowledge is

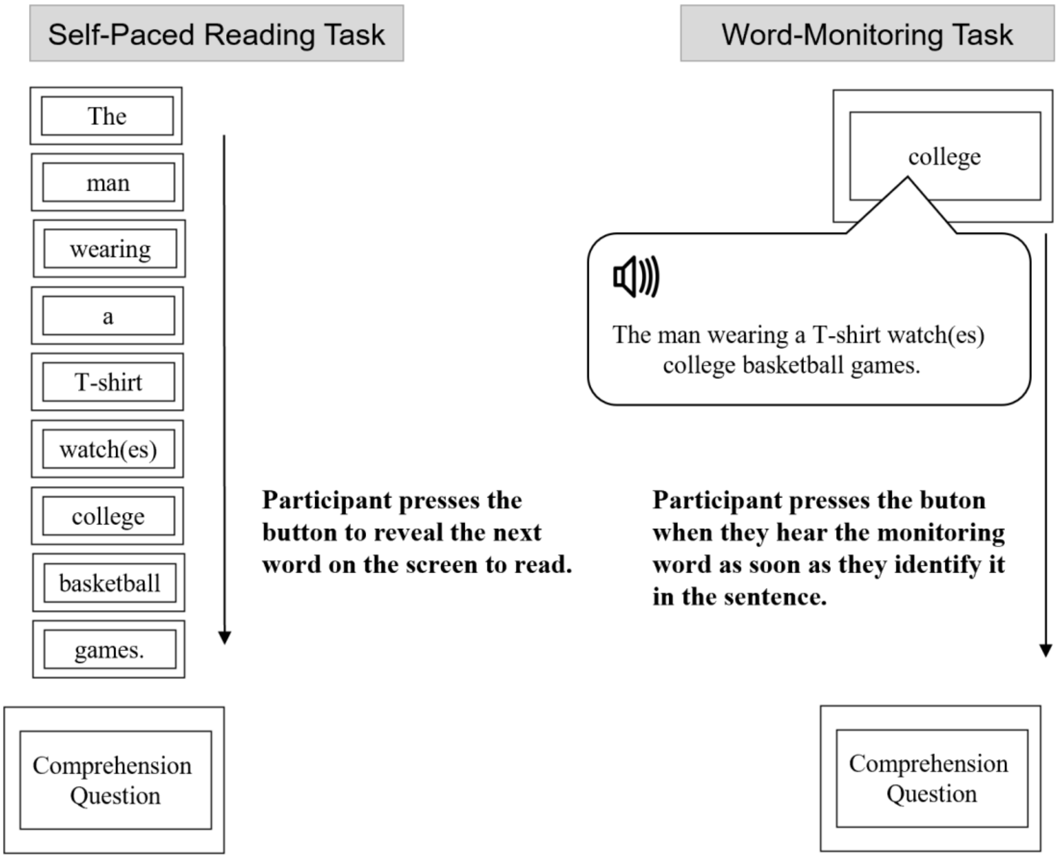

Figure 9.1. A blueprint of L2 comprehension.

activated (Perfetti & Bell, 1991; Taft et al., 2008). Readers and listeners also engage in grapheme-phoneme mapping as they process input (Castles et al., 2018). The locus of the next word-level processing stage of comprehension is lexical, involving recognition of word forms and activation of meanings. At this stage, lower-level grapheme and phoneme representations and their combinations activate lexical representations (and the activation from the lexical level flows back to the letter and phoneme representations). The lexical level processes include word identification, segmentation, and morphological and semantic processing (e.g., Perfetti & Stafura, 2014). It is possible that some formulaic sequences beyond single words that are stored and/ or processed holistically (e.g., lexical bundles, idioms) may also be accessed at this processing stage of comprehension (Conklin & Schmitt, 2012). Next, text and speech sequences containing lexical information are analyzed and parsed into propositions during the sentence-level processing stage. Grammatical knowledge underlies morphosyntactic processing in which the comprehender integrates the information from word-level processing to a larger unit of sentence(s) incrementally, as processing unfolds in real time (e.g., Clahsen & Felser, 2006). The comprehender also anticipates or predicts what linguistic (and non-linguistic) information comes next in all processing stages (e.g., Kaan & Grüter, 2021; Kuperberg & Jaeger, 2016). Abstract propositions generated in the course of the lower-level processing (decoding, word-level processing, and sentence-level processing) are interpreted at the higher levels of discourse comprehension processes (comprehension phase). In comprehension, readers and listeners construct a situational model of the text that reflects the subject matter and sets parameters, such as causation, intention, time, space, and protagonists (in narrative texts). This situational model is constantly being updated during comprehension, and discontinuity in even one of these parameters slows down comprehension and results in generating new inferences (Zwaan & Radvansky, 1998). The writer's/speaker's (illocutionary) intention also needs to be interpreted based on pragmatic knowledge as well as more general world knowledge (see Field, 2013, for more details).

In our model, automaticity is a middleware connecting processing and knowledge in L2 comprehension. SLA researchers have qualified both processing and knowledge as automatized, e.g., "automatized processing" (e.g., Hulstijn, 2007) and "automatized knowledge" (e.g., DeKeyser, 2017). Automaticity is a gradual concept. The degree of automaticity depends on how quickly, reliably, effortlessly, unconsciously linguistic knowledge can be accessed, and it is inferred from or assessed through the performance on a certain task (e.g., how quickly linguistic information is processed). Similarly, automatization refers to the gradual improvement of task performance indicated by the aforementioned criteria of automaticity (e.g., speed, stability, effortfulness, consciousness).

Automaticity of processing has major consequences for L2 comprehension. Lower-level processes (e.g., decoding, word identification and segmentation) can be automatized to a greater extent than higher-level comprehension processes, as the latter are highly context dependent, involving self-monitoring and integration with background knowledge (Lim & Godfroid, 2015; Perfetti, 2007). Because comprehension processes rely on limited working memory capacity (e.g., Baddeley, 2012), more efficient lower-level processing can free up cognitive resources in working memory needed for higher-order comprehension processes.

We further propose that automaticity is enabled by the quality of knowledge (e.g., how precise, integrated, and flexible linguistic representations are) and/ or by the type of knowledge (e.g., procedural versus declarative). Perfetti and Hart (2001), for example, argue for a causal relationship between efficiency of lexical processes, underpinned by the quality of lexical representations (aka lexical quality), and comprehension variability. They define lexical quality as detailed knowledge about word forms (orthography and phonology) and meanings. Perfetti (2007) explains that high lexical quality affords rapid and reliable meaning retrieval needed in reading comprehension. In the domain of grammar, the qualitative distinction is made between procedural and declarative knowledge. According to skill acquisition theory (DeKeyser, 2017), for instance, declarative knowledge (e.g., facts and rules) itself cannot be automatized; what can be automatized is the procedural knowledge that underpins comprehension processes and skills.

While declarative knowledge (e.g., metalinguistic rules) can be used flexibly for different language processing (e.g., reading and listening), its use consumes considerable mental resource. Hence, procedural knowledge, which is efficient and expends little resource, gradually replaces declarative knowledge through practice. Procedural knowledge can be further strengthened and automatized, enabling efficient skill execution (Suzuki, this volume).

In sum, our model specifies bidirectional relationships between knowledge and processing components in achieving automaticity. Automaticity builds upon knowledge and processing and serves as the foundation for fluent reading and listening comprehension. As indicated by the vertical and horizontal bidirectional arrows in Figure 9.1, the relationship between every processing and knowledge component in the model is interactive and mediated by automaticity. More than one type of knowledge can contribute to one component process in comprehension. For instance, although orthographic and phonological knowledge are linked to the decoding stage in Figure 9.1, these knowledge components are also utilized in word identification processes in word-level processing (Jacobs et al., 1998; Perfetti, 2007; Rastle & Brysbaert, 2006).

Tasks and Measures of Automaticity of Lexical and Grammatical Processing

Given limited space, in this chapter, we focus on two components of L2 comprehension, most amenable to automatization: word- and sentence-level processing, associated with lexical and syntactic knowledge (for pronunciation skills, see Saito & Plonsky, 2019). As illustrated in Figure 9.1, fluent access to lexical and syntactic knowledge has a cascading effect on higher-order processes, with disfluency at lower levels creating a bottleneck in comprehension.

Researchers studying automatization typically use online tasks that involve real-time language processing. Whereas offline or untimed tasks (e.g., multiple choice, fill-in-the-blank, and translation) provide accuracy scores indicative of the product (result) of language processing, online tasks are usually delivered using the computer software that offers precise control over the timing of visual and auditory input and accurate recording of response times (RTs) and/ or physiological responses (e.g., eye movements, event-related brain potentials). These temporal measurements allow us to investigate how L2 knowledge is deployed during real-time language processing, yielding useful information about the automaticity of access to L2 knowledge. Online tasks are usually performed under time constraints or pressure, that is, under the conditions that significantly reduce the involvement of strategically controlled processes and declarative/explicit knowledge.

Although RT on a given task is the most common (but rather simplistic) measurement of automaticity, fast processing does not necessarily mean automatic processing (see Suzuki, this volume). For instance, fast processing is not necessarily stable (i.e., showing little variability) or ballistic (i.e., impossible to stop once started). Using multiple criteria (e.g., RT as well as stability, use of strategies, and consciousness) for evaluating automaticity of processing is a useful strategy for identifying potential success in L2 reading and listening comprehension.

A key characteristic of automaticity is stability (or RT consistency) in a given task. According to Segalowitz and Segalowitz (1993), automatization of L2 skills is characterized by a qualitative shift (i.e., restructuring) in processing, rather than a mere speed-up of the existing routine. For instance, language learners may initially engage in a costly process of linking L2 forms with L1 translation equivalents during L2 comprehension, and this L1-translation route could become more efficient with practice. However, this practiced L1-translation route will still be less efficient than a direct L2 lexical semantic route (without the mediation of L1 translation), which can be established with the kind of practice that strengthens within-L2 meaning connection and supports automatization (Jiang, 2000; Kroll & Stewart, 1994). Similarly, while faster grammar processing may involve speeding up of declarative knowledge use (e.g., applying metalinguistic knowledge of rules), automatized (more stable) grammar processing would primarily rely on procedural knowledge with diminished access to declarative knowledge.

One method used to assess the change in the stability and efficiency of processing that indexes restructuring is to calculate the coefficient of variability/ variation of a person's RT (CV or CVRT), which is computed by dividing a person's standard deviation (SD) of RT by that person's mean RT (Segalowitz & Segalowitz, 1993). This measure has been utilized to study the extent to which learners automatize their L2 lexical (e.g., Elgort, 2011; Elgort & Warren, 2014; Hui, 2021; Hulstijn et al., 2009) and grammatical and syntactic processing skills (e.g., Hui & Godfroid, 2020; Hulstijn et al., 2009; Lim & Godfroid, 2015; Suzuki & Sunada, 2018).

There are several types of behavioral online tasks used to measure automaticity and automatization in L2 studies. Next, we present four such online task types: (a) judgment task (with or without priming); (b) matching task; (c) self-paced and word-monitoring task; and (d) reading and listening comprehension task (with eye-tracking).

Judgement task. Judgment tasks require learners to make a judgment/ decision about a spoken or written linguistic stimulus; for example, lexical decision (word/non-word), semantic decision (animate/inanimate), and grammaticality judgment (acceptable/unacceptable). The lexical decision task is probably the most commonly used task in assessing automaticity of access to lexical knowledge. In this task, participants read a sequence of letters or listen to a sequence of sounds and decide whether what they see or hear is a word (e.g., violin) or not a word (e.g., somer). They are instructed to make decisions as accurately and quickly as possible. This task requires participants to access lexical representations in the language (or languages) specified by the task. RT data is collected from many trials for each learner, and from many learners, because individual responses in such tasks (especially those of L2 participants) tend to vary. In addition, because the accuracy threshold for inclusion in the analysis is usually set high (e.g., 90–95% accuracy), L2 studies tend to require more participants than L1 studies. These studies are usually comparative, with RTs compared for different groups of participants and/or different types of stimuli. Shorter RT (i.e., faster decisions) may indicate more automatic L2 lexical processing and higher-quality lexical knowledge (e.g., Xu et al., 2014).

In L2 grammar research, acceptability (grammaticality) judgment tasks are commonly used to assess grammatical knowledge. In this task, participants make a judgment whether a presented sentence is grammatical or ungrammatical. While RT is the main outcome variable of interest when assessing automaticity of lexicality decisions, accuracy rates of acceptability judgments are often taken as an indicator of quality of L2 grammatical knowledge representations. However, because learners can use their metalinguistic knowledge and strategies to provide correct responses if sufficient time is available, acceptability judgment tasks that measure automaticity must be performed under time pressure. In a recent methodological review of acceptability judgment task, however, Plonsky et al. (2020) found that the quality of knowledge (e.g., automaticity, knowledge type: explicit/implicit, declarative/procedural) was not explicitly discussed in the majority of studies employing acceptability judgment task (76% out of 302 studies). On the other hand, some researchers argue that when an acceptability judgment task imposes time pressure on each item response, its accuracy score may indicate automatized processing (or implicit knowledge; see Godfroid et al., 2015). In addition to time pressure, using auditory (as opposed to visual) stimuli is more likely to render an acceptability judgment task as a measure of automatization (and use of implicit knowledge).

Priming. Priming manipulations are used in conjunction with different tasks, most commonly the lexical decision task. In a primed lexical decision task (Figure 9.2), the target (i.e., a stimulus to which decisions are made) is presented in the context of another stimulus (prime), either related or unrelated to the target. In word-level processing research, the relationship between the target (e.g., violin) and the prime could be semantic (e.g., piano or play), or form-based (e.g., violent, viobin, violin), or morphological (e.g., violinist). Priming manipulations target specific knowledge components; form-priming, for example, is typically used to examine formal-lexical representations, whereas semantic priming is used to assess quality of the lexical semantic representations. In priming experiments, high lexical quality (such as that in normal L1 lexical processing) may result in inhibition, that is, slower responses in the priming condition compared with the control condition (e.g., in form-priming; e.g., Davis & Lupker, 2006), or in facilitation, that is, faster responses in the priming condition compared with the control condition (e.g., in semantic priming; e.g., McRae & Boisvert, 1998). Thus, L2 researchers can use primed lexical decisions to probe different aspects of lexical representations and test which instructional and learning activities lead to better quality of L2 knowledge (Elgort, 2011). Experiments can be designed to minimize participants' awareness of the prime, targeting implicit processing without awareness, assumed to be automatic. For example, in masked priming, participants may not even be aware of the presence of the prime presented for a very short time (e.g., 50–60 milliseconds) and preceded and followed by a mask (such as

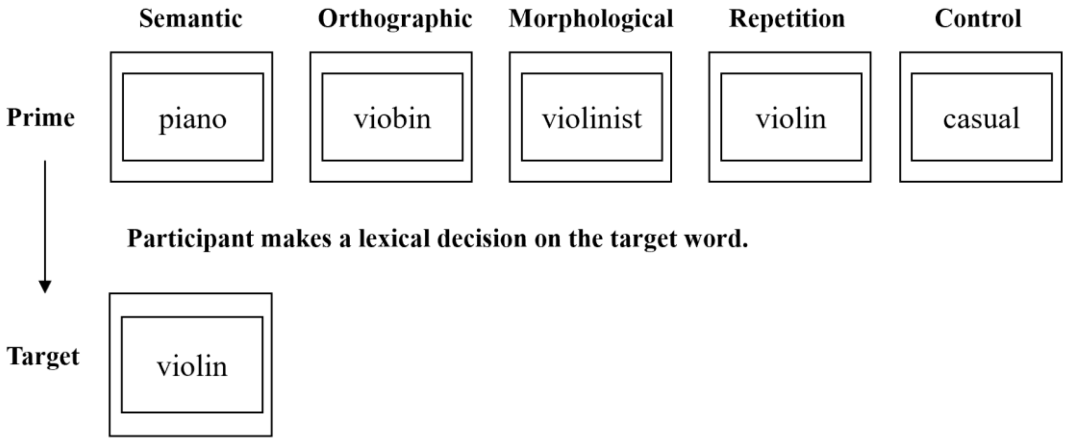

Figure 9.2. Illustration of Primed Lexical Decision Task.

hash-signs or random letters, in visual experiments).

Matching task. Although accuracy data from acceptability judgment tasks have been commonly used in the research of grammar knowledge representation, in part due to the influence from formal or generative approaches to SLA piano Participant makes a lexical decision on the target word. violin Prime Target Semantic Orthographic Repetition viobin Morphological violinist violin Control casual FIGURE 9.2 Illustration of primed lexical decision task. (e.g., Plonsky et al., 2020), RT has also been used to investigate L2 grammar knowledge (e.g., Jiang, 2011). For instance, a matching task is a useful task in which a participant is presented with a picture and chooses the correct sentence that matches the event described in the picture. Unlike acceptability judgment tasks, only grammatical sentences are typically used in matching tasks. RT and CV are used to gauge the speed and stability of grammatical processing at the sentence level.

Self-paced and Word-Monitoring Tasks. Sensitivity to certain linguistic phenomena (e.g., grammatical errors, pronoun resolution) has been studied using self-paced reading and word-monitoring tasks, which examine automaticity of grammar knowledge use in real-time sentence processing. In these tasks, slower RT is expected when participants process a part of the sentence containing an anomaly, such as linguistic (e.g., missing third-person s), syntactic (e.g., relative clause ambiguities) or semantic (e.g., meaning incongruencies) anomalies, relative to non-anomalous sentences. In this sense, these tasks are similar to priming tasks because they compare RT on manipulated and baseline (e.g., non-anomalous) trials.

In the self-paced reading task, participants read sentences, one word at a time, pressing a keyboard key or response button to show the next word (Figure 9.3). After each sentence, participants answer a comprehension question, to direct their attention to meaning rather than form (i.e., reducing deliberate attention to linguistic errors). For instance, participants read either (a) grammatical or (b) ungrammatical sentence containing third-person s manipulations:

(a) The man wearing a T-shirt **watch** college basketball games.
(b) The man wearing a T-shirt **watches** college basketball games.
Comprehension question: Does the man watch baseball?

Participants are expected to read the critical word (and one or two subsequent words) slower in Sentence (a) relative to Sentence (b), if they can detect the grammatical error (i.e., missing third-person s). Because this online error detection (linked to slower RT) is presumably enabled by robust morphosyntactic processing without voluntary control or awareness, the online sensitivity to errors in this task may be considered to indicate ballistic and/or implicit processing. The difference in RT to the critical word(s) between Sentence (a) and Sentence (b) reflects whether grammatical errors are automatically detected.

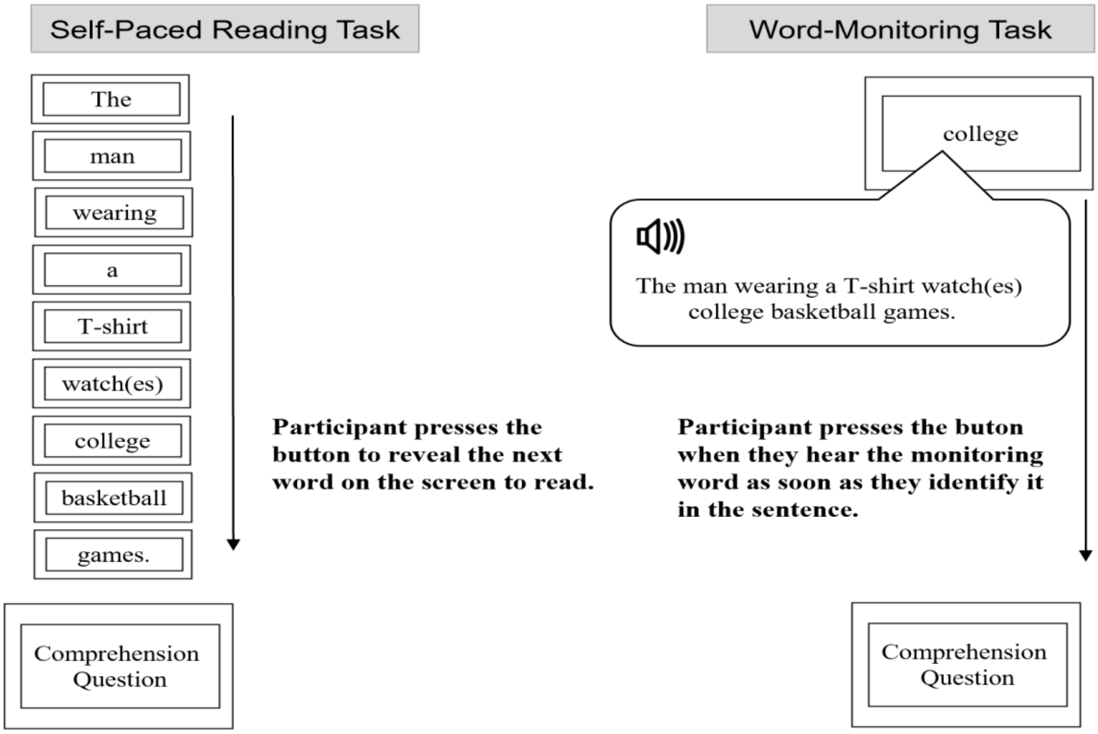

Figure 9.3. Illustration of Self-Paced Reading Task and Word-Monitoring Task.

A similar rational applies to a word-monitoring task (Figure 9.3) where participants listen to a sentence and press the keyboard button as soon as they hear the word they are instructed to monitor (e.g., “college”). RT difference to the monitored word (“college”) between (a) grammatical and (b) ungrammatical sentences is used as an index of error sensitivity. Self-paced reading and word-monitoring tasks are used in different strands of L2 research, including studies of automaticity (Marsden et al., 2018), and some researchers have argued that these tasks may be used to assess automatized implicit knowledge, minimizing the influence of explicit knowledge (e.g., Godfroid, 2016; Suzuki & DeKeyser, 2017).

Reading and Listening Comprehension Tasks. One of the approaches used to study comprehension in a natural way is recording eye movements during reading or listening. Eye-tracking can be used to examine real-time lexical and morphosyntactic processing. The use of such comprehension tasks with eye-tracking is relatively new in L2 automaticity research (e.g., Ling & Grüter, 2020; Suzuki & DeKeyser, 2017). Given the space limitation, although reading and listening comprehension studies with eye-tracking were included in the current synthesis, their detailed analysis is outside the scope of this chapter (see Godfroid, 2019a, for a comprehensive review).

Methodological Synthesis

Given the prominent importance of improving methodological rigor in the field of AL and SLA (e.g., Gass et al., 2021), this study contributes to our understanding of research on automaticity and automatization from a methodological perspective. While previous methodological syntheses examined the use of single tasks measuring grammatical knowledge and processing, such as acceptability judgment task (Plonsky et al., 2020) and selfpaced reading task (Marsden et al., 2018), our synthesis focused on various types of experimental tasks available for assessing automaticity and automatization. We conducted a methodological synthesis and surveyed the measurements of automaticity that are utilized in AL/SLA research. The following research questions (RQs) were addressed:

1. How much L2 research, focusing on automaticity and automatization, has been published in AL and SLA journals?
2. What methods have been used to measure automaticity and automatization in accessing L2 lexical and grammatical processing, in AL and SLA?
    (a) What tasks and experimental paradigms have been used to study automaticity and automatization?
    (b) What behavioral measures of automaticity and automatization have been used (e.g., accuracy, RT, CV, eye-movement data)?
    (c) What software has been used to program and deliver the experimental tasks?

The first RQ aimed to identify the research domain of automatization and automaticity. The second RQ focused on the tasks and approaches used to study automaticity and automatization. Although empirical research on automaticity goes beyond lexical and grammatical knowledge (RQ1), given the limited space, RQ2 focused on these two linguistic domains, in which automaticity is most commonly studied (i.e., 80 out of 99 coded tasks, or 81%; with the remaining 19% covering other domains and skills such as pragmatics and general reading speed).

Literature Search and Eligibility Criteria

We identified the body of primary research using two databases (Linguistics and

Language Behavior Abstracts [LLBA] and Education Resources Information Center [ERIC]) and the following 11 AL/SLA journals: *Studies in Second Language Acquisition*, *Language Learning*, *Second Language Research*, *Language Teaching Researc*h, *The Modern Language Journa*l, *System*, *TESOL Quarterly*, *International Journal of Applied Linguistics*, *International Review of Applied Linguistics in Language Teaching*, *Annual Review of Applied Linguistics*, and *Language Teaching*. Although empirical research on automaticity has also been published in psycholinguistic journals, such as *Applied Psycholinguistics* and *Bilingualism: Language and Cognition*, we focused on the AL and SLA databases and the journals in the analysis presented in this chapter. Other sources such as book chapters, theses, and conference papers were excluded. The data collection started in June 2021 and was completed in September 2021.

We used the following keywords (full text search) to locate target research articles:

("second language" OR "foreign language" OR L2 OR FL) AND (Automatization OR Automaticity OR Automatisation OR Automatized OR Automatised OR "coefficient of variation" OR "coefficient of variance" OR "coefficient of variability" OR CVRT) OR ("reaction time" OR RT OR latenc OR "reading time" OR "response time") NOT ("speaking fluency") NOT ("writing fluency")

These searches resulted in 2167 hits. After eliminating duplicates, the following five inclusion criteria were applied:

1. Only empirical studies (i.e., cross-sectional, longitudinal, and intervention research) were included; theoretical and meta-analysis articles were excluded.
2. Only studies investigating receptive knowledge and skills were included; studies examining productive knowledge and skills were excluded.
3. Studies investigating automaticity in the linguistic knowledge domain(s) of phonology, lexis, grammar, and pragmatics, and in relation to reading or listening skills were included.
4. Studies using time-based measures of language processing (e.g., RT, eye-movement) were included. Tasks were also included when accuracy was used as the main measure of automaticity of knowledge and processing. As the focus of this synthesis is on behavioral measurements, EEG and fMRI studies were excluded.
5. Only studies that contained an explicit claim to investigating automatization or automaticity in a second language were included. Specifically; (a) the title or abstract

indicates the motivation of research is to examine automatization; (b) a claim about studying automatization or automaticity is made in the literature review and/or research questions or goals, or (c) the measurements are explicitly tied to the concept of automatization (e.g., in the method section). When one of these criteria (a) – (c) is met, the discussion section was checked to ensure that the study findings were interpreted in relation to automatization/automaticity. Studies presenting an ad-hoc interpretation related to automaticity in the discussion section only were excluded.

The original sample was reduced to a total of 115 articles that met the first four criteria. These articles were further checked using criterion 5 to establish whether or not they explicitly identified automaticity or automatization as a dependent/outcome variable. A total of 69 out of 115 articles explicitly indicated that they measured automatization.[1] Studies that claimed to investigate the development of procedural knowledge in the literature review section, but did not explicitly claim to measure automatization or automaticity in the subsequent sections were excluded (e.g., Li & DeKeyser, 2017). We also excluded several cases where automaticity was only mentioned in the discussion section, as a way of interpreting the findings, but not used as one of the main constructs investigated in the study. For instance, we did not include Hopp (2013), which attributed differences in the findings (i.e., patterns of online processing by L2 and L1 speakers) to less automatized processing of L2 learners in the discussion section. These strict selection criteria allowed us to zoom in on the target research domain, i.e., L2 tasks that are specifically tailored to examine and measure automatization and automaticity—the target construct in the article.

Coding

For the selected studies, we coded the following characteristics: (a) target linguistic domain (lexical [including the processing of formal-lexical and lexical semantic representations], grammatical [including morphological and syntactic processing], others[2] [pragmatics, reading, listening skills]); (b) task; (c) task modality (auditory, visual, bimodal); (d) dependent measures (RT, CV, eye-movement measures); and (e) software. The coding scheme was developed through an iterative coding and discussion process

---

[1] The coding result for Criterion 5 initially diverged between the two authors. The first author flagged 46 articles that were difficult to provide a clear-cut code. The second author checked all those 46 articles. Any discrepancies that arose were discussed and resolved in the end.

[2] We found no studies that focused on phonology at the pre-lexical processing level.

between the authors, with documented additions and refinement of original definitions.

Findings and Discussion from the Methodological Synthesis

Research Domain on Automatization and Automaticity (RQ1)

Of the 69 articles, 55 targeted lexical or grammatical knowledge domains (see supplementary materials for the list of all studies). This suggests that the majority (80%) of L2 studies on automatization focused on some aspects of lexical and/or grammatical knowledge. In the 55 articles, 80 experimental tasks were reported.3 These tasks were used to tap a wide range of aspects in lexical knowledge that are used in word-level processing, that is, knowledge of form and meaning (e.g., Elgort, 2011; Ling & Grüter, 2020; Solovyeva & DeKeyser, 2018), morphology (e.g., Li et al., 2017), formulaic sequences, such as collocations (e.g., Sonbul & Schmitt, 2013) and idioms (e.g., Carrol et al., 2016). Grammatical structures targeted in the articles ranged from verbal inflections (e.g., Rodgers, 2011), morphosyntactic structures, such as case-marking, gender marking, tense-aspect-mood system (e.g., Roberts & Liszka, 2013; Suzuki & DeKeyser, 2017; Vafaee et al., 2017), and syntactic structures such as wh-movement and filler-gap dependencies (e.g., Dekydtspotter & Miller, 2013).

Notably, in SLA research, automaticity was often tied to a specific type of knowledge, such as “automatic competence” (Jiang, 2007), “automatized explicit knowledge” (Suzuki & DeKeyser, 2017), “non-declarative knowledge” (Obermeier & Elgort, 2021), “tacit knowledge” (Elgort & Warren, 2014), and “implicit knowledge” (Godfroid et al., 2015; Sonbul & Schmitt, 2013). Notwithstanding these varied constructs stemming from theoretical orientations of the researchers and study domains, automaticity was often considered in association with aspects in L2 knowledge in SLA research, rather than exclusively in terms of processing or skill.

Table 9.1 presents the number of articles in AL/SLA journals. More than half of the articles were published in Language Learning and Studies in Second Language Acquisition (32 out of 55). The absolute number of articles on automatization has increased over the three decades: 1990–2000 (n = 9), 2001–2010 (n = 12), and 2011–2021 (n = 34). Since the number of articles published in these journals also increased during the last three decades, we looked at the proportion of articles investigating automaticity and automatization by decade: 1990–2000 (0.17%), 2001–2010 (0.19%), and 2011–2021 (0.50%); this confirmed the observed increase.

Table 9.1. The number of articles in AL/SLA journals (N = 55)

Journal	Number of articles
Language Learning	16
Studies in Second Language Acquisition	16
Second Language Research	7
The Modern Language Journal	7
International Review of Applied Linguistics in Language Teaching	3
System	3
Language Teaching Research	2
TESOL Quarterly	1

Tasks for Measuring Automaticity of Lexical and Grammatical Processing (RQ2a)

Before providing an in-depth discussion of the tasks used to test automaticity of lexical and grammatical processing, we highlight the most notable imbalance for the modality of the tasks. Visual tasks represented 80% of all tasks. Out of 34 lexical knowledge tasks, 30 tasks were visual (88%), with only two auditory and two bimodal tasks. Out of 46 grammatical knowledge tasks, 34 tasks were visual task (78%). This bias indicates a gap in existing research on automatization in L2. This is surprising since automatic access to L2 knowledge is even morecritical in listening than in reading, due to the fleeting nature of connected speech. There is a clear need to address this gap by using auditory processing tasks in research on automaticity.

Table 9.2 presents tasks used to measure automaticity in lexical (k = 34) and grammatical knowledge (k = 46). For lexical knowledge, almost all tasks (30/34 = 88%) were categorized as judgment4 tasks, either primed or unprimed, and were, in most cases, either lexical decision or semantic judgment (e.g., animate/inanimate, L1 translation accuracy).

Table 2. Tasks used to Assess Automaticity in Lexical and Grammatical Knowledge (*k = 82*)

Lexical Knowledge & Word-level processing	***k***	**Grammar & Sentence level processing**	***k***
Judgement	16	Judgement	14
- Lexicality (*k* = 8)		- Acceptability (*k* = 13)*	
- Semantic (*k* = 7)		- Semantic (*k* = 1)	
- Spoken – Written Word-form			

Mapping (*k* = 1)			
Primed Judgement - Lexicality (*k* = 14)	14	Primed Judgement - Semantic (*k* = 2)	2
Self-paced Reading	1	Self-paced reading	13
Reading Comprehension with Eye-Tracking	2	Reading Comprehension with Eye-Tracking	1
Listening Comprehension with Eye-Tracking	1	Listening Comprehension with Eye-Tracking	1
		Matching (Picture – Sentence)	7
		Word-monitoring task	6
		Fill-in-the-blank	1
		Self-paced listening	1

*Two studies combined acceptability judgement tasks and eye-tracking technique (Clahsen, Balkhair, Schutter, & Cunnings, 2013; Godfroid et al., 2015).

A spoken–written word-form mapping task (k = 1) stands out among the judgment tasks. In this task, a visual presentation of a Chinese character (word) was followed by a visual presentation of a pinyin, accompanied by its sound. Participants were instructed to decide whether the pinyin and sound represented the correct pronunciation of the visually presented character (Xu et al., 2014). The goal of the task was to evaluate phonological representations at the word processing level. This example shows how researchers can be creative in devising experimental tasks to tap into aspects of lexical knowledge in a fine-grained manner.

Different types of priming were combined with the lexical decision task to investigate automaticity: semantic priming (k=5), morphological priming (k=4), form-priming (k =3), repetition-priming (k=1), and collocation priming (k=1). Interestingly, we did not find studies that used primed semantic judgment tasks in our analysis, yet, in psycholinguistic research primed semantic judgment tasks (such as semantic relatedness, categorization, and sense judgments) are relatively common (e.g., Finkbeiner et al., 2004).

Furthermore, we did not find studies testing sublexical processing (e.g., grapheme or phoneme decoding, or grapheme-phoneme mapping). This is somewhat surprising, because automatization of decoding is highly desirable and the quality of phonological representations can be probed by speech perception tasks (e.g., AX discrimination task). Instead, in SLA, researchers investigating automaticity are primarily interested in directly measuring the word-level and sentence-level processing. However, it may be useful to isolate the decoding processing as a sublexical process and measure its automaticity, because inefficiencies at the sublexical decoding stage may be the cause of disfluent word-level or/and sentence-level processing. This gap may be specific to the AL and SLA automatization research, as automatic sublexical processing and decoding are investigated in L1 reading studies (e.g., Hasenäcker & Schroeder, 2022).

For assessment of grammatical knowledge, acceptability judgment task and self-paced reading task were used frequently (see the detailed discussion of acceptability judgment task in the next section). The self-paced reading task has been the most common tool to study automaticity, which is consistent with a recent methodological synthesis of self-paced reading task by Marsden et al. (2018). Their synthesis revealed that the most common rationale for using self-paced reading in L2 research was measuring automatic knowledge or automaticity. Furthermore, the popularity of self-paced reading mirrors the skewed usage of visual modality in this synthesis (80%).

We identified four kinds of tasks exclusively used for grammar knowledge assessment. A matching task is a useful procedure where RT to select the right picture (and CV in some cases) is used as an index of automaticity. This task is versatile, as the modality of sentence presentation could be either auditory (k = 3) or visual (k = 4), or possibly bimodal; it is straightforward to compute RT (as well as CV), requiring no subtraction of RT in one condition from that in other condition.

The word-monitoring task is a promising task proposed as a measure of implicit knowledge in SLA (e.g., see Suzuki et al., 2023 for neural evidence), and it was less commonly used than the self-paced reading task but more often than self-paced listening task. Given the aural modality of word-monitoring task, it a useful tool to examine automaticity in aural skills. Although the selfpaced listening task was also used as a measure of implicit knowledge in de Jong's (2005) research, this task has not been used in relation to automaticity since 2005. The infrequent use of self-paced listening is also reported in Jiang's (2011) review in L2 psycholinguistic research. One possible reason is the difficulty and labor-intensive nature of preparing stimuli (e.g., each word needs to be carefully edited out from a sentence).

The fill-in-the-blank task, conducted under time-pressure, was used by Suzuki and DeKeyser (2017) as a measure of automatized explicit knowledge. In this task, participants were asked to fill in the target grammatical structure in a gapped sentence as quickly as possible.

Our analysis also revealed asymmetries in the use of tasks for measuring lexical and grammar knowledge. Three out of the first five tasks (semantic judgment task, primed judgment task, and self-paced reading task) showed the skewed frequency of usage in lexical versus grammatical knowledge studies. While semantic judgment task was primarily used to study lexical knowledge and processing, it was used once as a grammar test (Paciorek & Williams, 2015). Paciorek and Williams (2015) asked learners to classify sentences by type of change (increase versus decrease), but indirectly assessed their sensitivity to the semantic preferences of novel verbs (whether the verb took abstract or concrete collocates). This indirect elicitation of the target knowledge is similar to the approach used in self-paced reading and word-monitoring tasks where the researcher asks L2 learners to focus on meaning but are interested in assessing their sensitivity to grammatical anomaly, measured by RT difference (see literature review). Paciorek and William's task is a useful addition to the researchers' tool box of grammar tests.

Priming paradigms are more likely to be used in studies of lexical knowledge, whereas self-paced reading task is frequently used to assess grammatical knowledge. Self-paced reading was used only once in the domain of lexical knowledge (Obermeier & Elgort, 2021) to measure participants' ability to access figurative meanings of newly learned L2 collocations (e.g., throw in the towel) in sentence reading, offering a more ecologically valid measure of lexical quality. A priming paradigm was deployed to examine the processing of L2 English wh-dependencies in Dekydtspotter and Miller (2013). In this task, participants read a sentence word by word, while classifying a picture as animate or not. These rare cases show some creative ideas for tailoring tasks to assess automaticity in L2 processing, associated with different knowledge domains.

Dependent Measures (RQ2b)

Table 9.3 summarizes the dependent variables used in the studies identified in this synthesis. Accuracy rate was used as the sole dependent variable (with no other measures) in four acceptability judgment tasks and one fill-in-the-blank task conducted under time pressure.

Table 9.3. Dependent Measure of the Tasks

	Lexical/Word-Level	Grammar/Sentence-Level
Accuracy	0	5
RT	31 (RT difference = 15)	37 (RT difference = 23)
CV	16	2

When the accuracy rate was used as an indicator of automaticity in acceptability judgement tasks, imposing time pressure (i.e., learners were instructed to make a judgement as quickly as possible) or setting a time limit per item was used in 64% of the tasks (7 out of 11 tasks, excluding one task with eye-tracking). Only two tasks (18%) utilized auditory stimuli. This disproportionate use of the visual judgement mode is consistent with a recent comprehensive methodological synthesis of acceptability judgement task (Plonsky et al., 2020).

Table 9.4. Task Parameter and Dependent Variable in Acceptability (Grammaticality) Judgement Task

Dependent Measure	Task Parameters	k
Accuracy Only	Auditory, Time-pressured	2
	Visual, Timed	2
Accuracy and RT	Visual, Time-pressured	3
	Visual, Untimed	2
	Visual, Not reported	2

Although, theoretically, it would be difficult to justify the use of the visual acceptability judgment task without the time pressure as a measure of automaticity (k = 2), RT has been interpreted alongside the accuracy data (n = 7), as illustrated in Table 9.4. In the earlier work (Robinson & Ha, 1993; Robinson, 1997), RTs from acceptability judgment tasks were compared between different conditions (e.g., trained versus new items). Robinson and his colleague used RT as an indicator of processing speed of grammatical rules for familiar (trained) and novel (untrained) sentences, which may be an interesting avenue to pursue to study the nature of automaticity. Although this approach has rarely been used in the field since, RT in acceptability judgments was utilized in more recent studies to examine “solidity of the knowledge” (Jung, 2015) and “monitored processing” (Lado et al., 2014). Nonetheless, because there is individual variability in RT that is not solely due to the processing of the target grammatical structure (e.g., due to individual reading speed or quality of lexical knowledge), RT may only partially reflect automatic processing of target grammatical structure. In order to directly measure RT of target grammatical processing, for instance, Andringa et al. (2012) developed an acceptability judgment task in which the start of each sentence was short (three to four words), reducing the influence of (general) sentence reading speed.

Our analysis showed that RT was widely used in both lexical and grammatical knowledge domains, but CV was primarily used in lexical decision tasks, both unprimed (k = 8) and primed (k = 4). CV as a measure of automaticity was under-utilized in assessing grammatical knowledge: it was computed only in studies that used matching task as a test of grammar knowledge (Ammar, 2008; Rodgers, 2011). The CV analyses corroborated corrective feedback advantage in Ammar’s (2008) study and development of grammar knowledge from beginner to advanced levels in Rodgers’ (2011) study. Echoing Godfroid’s (2019b) call for utilizing CV in vocabulary research, we emphasize the importance of examining and reporting CV in grammar knowledge tests. This would not necessitate any additional data collection, because “once researchers have obtained RT data, they basically get the CVRT measure for free” (Godfroid, 2019b, p. 448).

About half of the tasks (k = 38 out of 68) used RT difference (rather than absolute RTs) to index automaticity (see literature review section). This approach was used in 15 primed lexical decision tasks and 23 tasks assessment of grammar knowledge, including self-paced reading (e.g., Roberts & Liszka, 2013), self-paced listening (de Jong, 2005), word-monitoring (e.g., Suzuki & DeKeyser, 2017), semantic judgment (Paciorek & Williams, 2015), and matching (de Jong, 2005).

Experiment Software (RQ2c)

Of 80 tasks, the top three programs used to code and present experiments were DMDX (20), E-prime (17), and SuperLab (3). Experiments probing lexical knowledge were programmed more frequently with E-prime than DMDX (k = 12 versus 7, respectively), whereas the opposite was observed for grammatical knowledge (k = 5 versus 13). The rest of programs/software (e.g., PsychoPy, Linger, Ibex Farm) were used only once. Some programs, such as Hypercard, used until early 2000, have been discontinued. No software information was reported in 13 articles. In eye-movement experiments, EyeLink (5), SMI RED eye-tracker (1), and Tobii (1) were used.

Methodological Guidelines

In this section, we present guidelines on how to select experimental tasks for assessing automaticity in studies of L2 lexical and grammatical processing and knowledge (see, e.g., Jiang, 2011, for a technical guide for programming and implementing the computerized tasks that were identified in this survey).

For tasks measuring automaticity, researchers should first give careful consideration to the target construct. Automaticity is multifaceted; researchers should devise a task that can capture specific aspect(s) of automaticity (speed, stability, and/or consciousness). A direct method of assessing automaticity in studies of grammar is to use picture–sentence matching tasks and compute RT and CV, which correspond to speed and stability of processing, respectively. Certainly, using both RT and CV (reflecting speed and efficiency of sentencelevel, morphosyntactic processing) is recommended for assessing automatized grammatical knowledge more comprehensively. A caveat is that the validity of CV as a measure of automaticity is still under debate (e.g., Hui, 2020; Hulstijn et al., 2009; Lim & Godfroid, 2015; Solovyeva & DeKeyser, 2018). More empirical studies should be conducted to scrutinize the utility of CV for capturing automaticity in L2 processing.

Exemplary Study for Vocabulary
Hui, B. (2020). Processing variability in intentional and incidental word learning: An extension of Solovyeva and Dekeyser (2018*). Studies in Second Language Acquisition, 42*(2), 327-357.

An important contribution to research on automatization of lexical knowledge (as measured by CV) has been made by Hui's 2020 article. Hui's research went beyond measuring automatization at a single point in time in the course of learning by examining the trajectory of CV changes during initial word learning stages. Another important contribution of this paper was its attempt to extend the use of the CV measure of automatization beyond decontextualized deliberate learning and beyond response-based behavioral experiments.

In study one, Hui conducted a deliberate word learning experiment and obtained CV measures on multiple testing sessions (blocks) over time, aiming to build a longitudinal picture of changes from the no-knowledge stage, through acquisition of declarative and procedural knowledge, to automatization. The CV was calculated using RT data on correct responses in a sematic judgment task, where participants categorized target (Swahili) words as animate or inanimate. A clear item inclusion criterion was applied prior to the analysis (i.e., 80% accuracy on the final test). The researcher also made an important empirically motivated decision to fit statistical models with the test-block, as a primary-interest predictor, without assuming that the CV would reduce in a linear manner. This made it possible to establish that the CV changes across participants followed an inverted U-shaped trajectory, with an initial increase in variability followed by a decrease, signaling increased automatization toward the end of the experiment.

Hui also conducted a re-analysis of the eye-movement data from a previously published reading study (Elgort et al., 2018), calculating CV on the first 12 occurrences of low-frequency words. In eye-tracking studies, an important methodological decision is which measures to include in the analysis. In this study, the selection of two early processing measures for the CV analysis (i.e., first-fixation duration and gaze duration) was motivated by the need to reduce the amount of controlled, strategic processing, bringing them in line with the RT data obtained in judgment tasks performed under time pressure.

Automatized grammatical knowledge often comes with different theoretical labels such as implicit knowledge, procedural knowledge, and automatized explicit knowledge. However, out of seven matching tasks identified in this survey, their measures (RT and/or CV) were not explicitly linked to any of these theoretical constructs (e.g., Rodgers, 2011; DeKeyser, 1997; cf. de Jong, 2005, which expressed a more nuanced stance). A different type of task is usually employed to scrutinize different types of grammatical knowledge. For instance, if researchers are interested in assessing implicit knowledge, the task should be designed to limit opportunities to access and use explicit knowledge. In order to minimize access to explicit knowledge, tasks need to meet two criteria (see Suzuki & DeKeyser, 2017): (a) focus on meaning and (b) real-time sentence processing. The tasks in this synthesis that meet these two requirements are real-time grammar comprehension tasks such as self-paced reading and word-monitoring tasks, as well as semantic judgment tasks devised by Paciorek and Williams (2015), and reading/ listening comprehension with eye-tracking task. RT difference was used in these tasks to capture the sensitivity to ungrammaticality, and the tasks are accompanied by the comprehension questions and direct participants' attention away from linguistic forms (hence, minimizing the conscious application of explicit knowledge). It is also useful to assess the awareness of participants via retrospective verbal report if automaticity is assessed with regard to "lack of awareness" (see Godfroid, 2016, in Exemplary Study).

Exemplary Study for Grammar
Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition, 38*(2), 177-215.

This study investigated to what extent listening to auditory input containing target grammatical structure leads to the acquisition of automatized (implicit) knowledge. Thirty-eight upper-intermediate L2 German learners in the U.S. processed 144 spoken sentences containing a difficult morphological structure (vowel-changing verbs) and matched them to the correct pictures. No rule explanation or instruction to search rules was provided during the treatment or testing. Two types of outcomes tests were used: (a) word-monitoring task as a measure of automatized knowledge and (b) controlled oral production as a measure of non-automatized, productive knowledge. Most of the learners (33 out of 38) could not report the ungrammatical verbs in the input flood, suggesting that their learning took place without awareness. Regardless of awareness status, they developed sensitivity to ungrammatical sentences in the word-monitoring task, as indicated by significantly slower RT for ungrammatical over grammatical sentences. While their receptive skills were automatized, as evidenced by the word-monitoring task, the development of productive knowledge was limited. Only learners with some prior knowledge of target structure tended to show improvement of productive knowledge. Input-rich listening treatment facilitated automatization of grammatical structures in receptive mode.

If the strict operationalization of "lack of awareness" is unnecessary for the purpose of the study, an acceptability judgment task may suffice to assess automaticity in many situations. Although acceptability judgment tasks under time pressure and auditory modality may draw on explicit knowledge to some extent, it can measure automatized explicit knowledge (e.g., Vafaee et al., 2017; cf. Godfroid et al., 2015). While time pressure may be less stringent to limit the use of explicit knowledge, setting the time limit (e.g., the average native speakers' RT plus 20%) may be too stringent. A recent study suggests that imposing a time limit can disrupt the basic language processing before making any judgment by not only L2 but also L1 speakers (Maie & Godfroid, 2022). It may thus be most judicious to use time pressure (rather than time limit) in acceptability judgment tasks (preferably auditory modality) to assess the degree of automaticity in grammatical knowledge, while avoiding the discussion of (non)consciousness of language processing.

When examining L2 morphosyntactic processing, researchers should be mindful of the efficiency of lower-level processing. For instance, lexical knowledge (deployed for decoding and word-level processing) influences the performance of grammatical knowledge tasks. In order to check the validity of the tasks (i.e., to what extent task performance stems from participants' grammatical knowledge, relatively independently from lexical knowledge), it is useful to administer a lexical task to assess the knowledge of lexical items from the grammar task (see Maie & Godfroid, 2022, for further discussion).

In this methodological synthesis, two types of the judgment task were prevalent in studies of lexical knowledge and processing—a lexical decision (primed and unprimed) and semantic judgment. The choice of the task depends on the domain of interest. Both lexical and semantic judgment tasks require access to formal-lexical representations (i.e., recognition of wordforms, either spoken or written) and lexical semantic representations (i.e., retrieval of meanings). However, the lexical decision task (e.g., "is hive a word?") does not necessarily require deep processing of meaning (Grainger & Jacobs, 1996). Therefore, researchers interested in automaticity of semantic processing need to either combine semantic priming with lexical decisions, or use semantic judgment tasks (e.g., semantic categorization or semantic relatedness tasks).

Because L2 speakers are more strategic in their approach to experimental tasks and often prioritize accuracy over fluency, studying automatization is more challenging in L2 than in L1. Therefore, experimental instructions need to explicitly emphasize response speed, cautioning against overthinking, and participants should be given ample pre-experiment practice opportunities. Strategic processing, characteristic of L2 participants' task behavior, can also be reduced by using priming; particularly, masked priming, where the prime is masked and presented very briefly. Thus, when primes are presented subliminally, participants are completely unaware of the relationship between primes and 9781032539904_pi-282.indd 227 16-Jun-23 4.32.25 PM 228 Yuichi Suzuki and Irina Elgort targets, making it possible to attribute the priming effect to automatic activation of specific knowledge components. Using priming in studies of automaticity also reduces unwanted RT variability, compared with unprimed tasks. This is because, in priming studies, researchers compare differences in RTs to the same target, under primed and unprimed (control) conditions. For instance, RTs to "violin" (target) preceded by "piano" (related prime) are compared with RTs to "violin" preceded by "casual" (unrelated prime). This priming paradigm thus allows researchers to test automatization in L2 processing, while minimizing unwanted variability caused by the L2 participants' knowledge of individual words. However, when designing priming experiments, researchers should carefully consider the characteristics of their experiential stimuli that affect lexical processing (e.g., word-form frequency, bigram frequency, orthographic and semantic neighborhood, pronounceability, word length, concreteness, imageability; see Balota et al., 2006).

Conclusion

In order to understand the body of AL and SLA research on automaticity, we conducted an initial methodological synthesis of tasks and measurements that have been used in the last three decades (1990–2021). This synthesis identified 34 lexical tasks and 46 grammatical tasks developed for assessing automaticity, which we categorized further into nine task types and coded for their measurements (accuracy, RT, CV). The findings indicated a paucity of research in several domains. First and foremost, despite the tight connection between processing automaticity and auditory input, visual modality was predominant, particularly in assessing automaticity in the domain of lexical knowledge. Another important finding is the imbalance in the frequency of task use in studies of lexical and grammatical knowledge. While self-paced reading was primarily used to measure grammar knowledge, a semantic judgment task and priming were primarily used to measure lexical knowledge. These tasks have the potential to be used to examine both lexical and grammatical processing; future empirical research should explore cross-domain application of different experimental tasks, especially when automaticity in both linguistic domains is considered to support L2 processing. Although most tasks identified in the synthesis originate from L1 psycholinguistic research, they were adopted creatively by L2 researchers to investigate automaticity in L2 knowledge and processing. The current study focused on tasks in the lexical and grammatical studies published in AL/SLA journals; nevertheless, the current synthesis could be used as a basis for future syntheses that could extend the scope to tasks used to assess automaticity in other knowledge domains and in studies published in psycholinguistic journals.

Exercises on developing experimental tasks (see Supplementary Materials for

Answers)

1. Elgort (2011) investigated the quality of lexical knowledge of L2 vocabulary items learned using flashcards. After the learning phase, participants completed three primed lexical decision tasks (form-priming, repetition-priming, and semantic priming), in which the studied vocabulary items were used as primes. Explain why these items were used as primes (not as targets) in these tasks.
2. Self-paced reading is rarely used to examine automaticity in lexical processing. One exception is Obermeier and Elgort (2021). Explain why the researchers chose this task to assess the processing of recently learned figurative phrases.
3. In order to assess automaticity in grammatical knowledge using an acceptability judgment task, what experimental parameters do you need to pay attention to?
4. There are different types of comprehension-based tasks for assessing automaticity in grammatical knowledge such as semantic judgement, self-paced reading, and word-monitoring tasks. It may be difficult to test all kinds of morphosyntactic structures in a language. Think of several grammatical features that can be tested using these comprehension-based tasks.

Supplementary Materials

To view the supplementary materials, visit the following link: https://osf.io/2zmsv/

References

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research, 12*, 183–210.

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*, 49–78.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of Psycholinguistics* (2nd ed., pp. 285–375). Elsevier.

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition, 38*, 403–443.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*, 5–51.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.

Clahsen, H., Balkhair, L., Schutter, J.-S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research, 29*, 7–31.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61.

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 668–687.

de Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition, 27*, 205–234.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195–221.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). Cambridge University Press.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In H. M. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 119–138). Wiley-Blackwell.

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 15–32). Routledge.

Dekydtspotter, L., & Miller, A. K. (2013). Inhibitive and facilitative priming induced by traces in the processing of wh-dependencies in a second language. *Second Language Research, 29*, 345–372.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*, 367–413.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning, 64*, 365–414.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition, 40*, 341–366.

Field, J. (2013). Cognitive validity. In A. Geranpaye & L. Taylor (Eds.), *Examining*

Listening: Research and Practice in Assessing Second Language Listening (pp. 77–151). Cambridge University Press.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*, 1–22.

Gass, S., Loewen, S., & Plonsky, L. (2021). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching, 54*, 245–258.

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition, 38*, 177–215.

Godfroid, A. (2019a). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. Routledge.

Godfroid, A. (2019b). Sensitive measures of vocabulary knowledge and processing: Expanding nation's framework. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 433–453). Routledge.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgements measure distinct types of knowledge. *Studies in Second Language Acquisition, 37*, 269–297.

Grabe, W., & Stoller, F. L. (2011). *Teaching and Researching Reading* (2nd ed.). Pearson Education.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518–565.

Hasenäcker, J., & Schroeder, S. (2022). Specific predictors of length and frequency effects in German beginning readers: Testing component processes of sublexical and lexical reading in the DRC. *Reading and Writing, 35*, 1627–1650.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*, 33–56.

Hui, B. (2020). Processing variability in intentional and incidental word learning: An extension of Solovyeva and DeKeyser (2018). *Studies in Second Language Acquisition, 42*, 237–357.

Hui, B., & Godfroid, A. (2021). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics, 42*, 1089–1115.

Hulstijn, J. H. (2007). Psycholinguistic perspectives on language and its acquisition. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 783–795). Springer US.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied*

Psycholinguistics, 30, 555–582.

Jacobs, A. M., Ray, A., & Ziegler, J. C. (1998). Mrom-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition* (pp. 147–188).

Erlbaum. Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*, 47–77.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*, 1–33.

Jiang, N. (2011). *Conducting Reaction Time Research in Second Language Studies*. Routledge.

Jung, J. (2015). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research, 20*, 92–112.

Kaan, E., & Grüter, T. (2021). *Prediction in Second Language Processing and Learning* (Vol. 12). John Benjamins.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*, 149–174.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32–59.

Lado, B., Bowden, H. W., Stafford, C. A., & Sanz, C. (2014). A fine-grained analysis of the effects of negative evidence with and without metalinguistic information in language development. *Language Teaching Research, 18*, 320–344.

Li, J., Taft, M., & Xu, J. (2017). The processing of English derived words by Chinese-English bilinguals. *Language Learning, 67*, 858–884.

Li, M., & DeKeyser, R. M. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition, 39*, 593–620.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics, 36*, 1247–1282.

Ling, W., & Grüter, T. (2020). From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research, 38*, 289–313.

Maie, R., & Godfroid, A. (2022). Controlled and automatic processing in the acceptability judgment task: An eye-tracking study. *Language Learning, 72*, 158–197.

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of selfpaced reading in second language research. *Applied Psycholinguistics, 39*, 861–904.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 24*, 558–572.

Obermeier, A., & Elgort, I. (2021). Deliberate and contextual learning of L2 idioms: The effect of learning conditions on online processing. *System, 97*, 102428.

Paciorek, A., & Williams, J. N. (2015). Implicit learning of semantic preferences of verbs. *Studies in Second Language Acquisition, 37*, 359–382.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383.

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30*, 473–485.

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity (pp. 67–86). American Psychological Association.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22–37.

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research, 36*, 583–621.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology, 53*, 97–145.

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research, 29*, 413–439.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–247.

Robinson, P. J., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition, 15*, 413–438.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *International Review of Applied Linguistics in Language Teaching, 49*, 295–319.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*, 652–708.

Segalowitz, N. S. (2003). Automaticity and second languages. In C. J. Doughty & H. M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382–408). Blackwell.

Segalowitz, N. S., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. de Groot (Eds.), Handbook of Bilingualism: *Psycholinguistic Approaches* (pp. 371–388). Oxford University Press.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369–369.

Solovyeva, K., & DeKeyser, R. (2018). Response time variability signatures of novel word learning. *Studies in Second Language Acquisition, 40*, 225–239.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*, 121–159.

Suzuki, S., & Révész, A. (this volume). Measuring speaking and writing fluency: A methodological synthesis focusing on automaticity. In Y. Suzuki (Ed.), *Practice and Automatization in Second Language Research: Perspectives from Skill Acquisition Theory and Cognitive Psychology* (pp. 235–264). Routledge.

Suzuki, Y. (this volume). Introduction: Practice and automatization in a second language. In Y. Suzuki (Ed.), *Practice and Automatization in Second Language Research: Perspectives from Skill Acquisition Theory and Cognitive Psychology* (pp. 1–36). Routledge.

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*, 747–790.

Suzuki, Y., & Sunada, M. (2018). Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. Bilingualism: *Language and Cognition, 21*, 32–46.

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing. *Studies in Second Language Acquisition, 45*, 109–136.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming.

Journal of Memory and Language, 58, 366–379.

Vafaee, P., Suzuki, Y., & Kachinske, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition, 39*, 59–95.

Vandergrift, L., & Goh, C. M. (2012). *Teaching and Learning Second Language Listening: Metacognition in Action*. Routledge.

Xu, Y., Chang, L. Y., & Perfetti, C. A. (2014). The effect of radical-based grouping in character learning in Chinese as a foreign language. *The Modern Language Journal, 98*, 773–793.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.