

# *Using psycholinguistic techniques in a second language teaching setting*

Book or Report Section

Accepted Version

Marinis, T. and Cunnings, I. (2018) Using psycholinguistic techniques in a second language teaching setting. In: Wright, C., Piske, T. and Young-Scholten, M. (eds.) Mind matters in SLA. Multilingual Matters, Bristol, pp. 185-202. ISBN 9781788921626 Available at <http://centaur.reading.ac.uk/78737/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Publisher: Multilingual Matters

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## Chapter 10

### Using psycholinguistic techniques in a second language teaching setting

Theodoros Marinis & Ian Cunnings

#### Introduction

The curriculum in linguistics, applied linguistics, and foreign language teaching programmes includes a large range of subjects, but often it does not include grounding in psycholinguistic techniques that can be used to assess language development in a classroom setting. As a result, when language teachers want to evaluate student's learning, they have to rely on **language assessments/tests** that are part of textbooks or other material developed on the basis of the language syllabus. Many of these tests are in written form and measure the students' explicit knowledge of vocabulary or grammar. The format is often similar to that provided during instruction, and tests the students' explicit knowledge of the material taught rather than the degree to which they have internalised (or they had started to internalise) the vocabulary or grammar. Students' performance in these tests is often mixed. For example, some students score highly in written tests but have difficulties in comprehending or using language in real situations. Others are able to use language in real situations but are not particularly successful in these assessments. This can be a puzzle for both teachers and students. What this demonstrates, however, is that these forms of assessment are measuring students' explicit knowledge in the context they have learnt it, but not necessarily their ability to use this knowledge for comprehension and/or spontaneous expression. This sort of assessment also demonstrates large individual differences between learners in **learning strategies** but also outcomes of these strategies. Some learners focus on learning rules explicitly but do not internalise them; others internalise the rules but they don't have explicit knowledge of them and have difficulties to apply them in the context of a test.

Another important issue is that of **modality**. The aim in learning a language is to develop both comprehension and production skills. However, the two modalities don't develop at the same rate, with comprehension often preceding production at early stages of development. Therefore, it is

important to be able to assess comprehension separately from production. This requires careful design of the test material in order to be able to separate the two modalities and measure students' strengths and weaknesses in comprehension separately from their strengths and weaknesses in production.

The field of psycholinguistics has made impressive advances within the last twenty years in terms of developing methods to assess different modalities (comprehension, production) and levels of language representation (phonology, morphology, syntax, semantics, pragmatics) separately from each other, as well as separating **implicit** from **explicit knowledge** of language in assessment. Second language teachers and research students can use this knowledge and methods to develop their own assessments and evaluate their students' learning in a classroom setting throughout the year. This can enable them to tailor teaching and learning on the basis of student progress and provide additional activities or move faster in the syllabus when they see that students have internalised specific rules and properties of the language.

This chapter presents a selection of psycholinguistic techniques together with ideas about how these techniques can be used to assess language abilities in a classroom setting. The chapter focuses on four different psycholinguistic techniques by describing the rationale for each technique, the possible phenomena they can be used for, and the groups they are suitable for, the procedure for each technique, how to analyse the data, and their strengths and limitations. We focus on psycholinguistic techniques that can be implemented using readily available software and hardware (e.g. a standard computer setup), rather than more complex techniques that require specialist equipment and neurological training, e.g. eye-tracking, ERP (event-related potentials measuring brain activity) and fMRI scans (functional magnetic resonance imaging). The first part of this chapter will focus on production and the second part on comprehension.

## **Psycholinguistic techniques measuring language production**

A large range of production tasks has been developed within the psycholinguistic tradition that are at the disposal of language teachers and research students, for example, elicited imitation, elicited production, narrative tasks, and **syntactic priming** tasks, to mention just a few. These tasks have

advantages and disadvantages as each one assesses language production in a slightly different way, has specific task demands and may be best suited for specific learner groups depending on the area to be assessed, the age, and the proficiency level of the learners. Due to space limitations, this chapter will introduce and discuss two of these tasks, *elicited imitation* and *syntactic priming*. See other chapters in this volume, such as Kahoul *et al.* (chapter 6), Rogers *et al.* (chapter 7) and Mora and Safronova (chapter 8), for illustrations of other tasks which are also used in collecting data for studies of second language acquisition.

### *Elicited imitation*

**Elicited Imitation** (EI) or Sentence Repetition has a long tradition in first and second language acquisition research and has often been used by researchers to measure language proficiency in a foreign language teaching setting, although rarely as a formal assessment (for second language acquisition, see e.g. Lust, Flynn & Foley, 1996; Jessop *et al.*, 2007; Bley-Vroman & Chaudron, 1994). At the heart of EI tasks is the innate human capacity (and spontaneous tendency) to imitate behaviour in our environment, including language, for example sounds, words, and sentences. The rationale behind using EI tasks to measure language proficiency is that, although imitating someone else's behaviour is a spontaneous capacity, it is very hard to imitate sentences accurately if the structures in the sentence are not part of one's grammatical system. This is because to imitate a sentence, participants have to be able to analyse it at all levels of representation (phonological, morpho-syntactic, semantic), extract its meaning, store it temporarily in **short term memory** and then use the production system to repeat it. Accurate verbatim imitation of a sentence depends on all processes and levels of representation related to comprehension and production and the ability to store and retrieve language material from memory using the **episodic buffer**, a temporary storage system with limited capacity that holds integrated chunks (Baddeley, 2000). Hence, verbatim sentence imitation is not possible if someone has not yet acquired the specific structures that are part of the sentence. Exceptions are when the sentences are very short, and as a result, participants can echo the sentence passively without having to use their grammatical system to analyse and reproduce it. Therefore, sentences in EI tasks have to be of a considerable length in order to force the language learner to use

their grammatical system in order to imitate the sentence. Length is relative to a range of factors, such as age, memory capacity, and language proficiency. Therefore, it is important to try out sentences of variable length before deciding what length is suitable for specific learners. Very short sentences can lead to passive echoing, which cannot inform us about the learner's proficiency, whereas very long sentences can lead to breakdown – the learner may not be able to repeat the sentence at all.

Phenomena. EI tasks can be used to assess a variety of linguistic phenomena with the caveat that sentence length has to be fine-tuned to avoid automatic echoing and still be imitated. For example, previous research has used EI tasks to investigate learners' acquisition of sentence structure, phrase structure, word order, and anaphora across a range of languages (for a review see Lust, Flynn & Foley, 1996). Recent work from a multi-national network of researchers (COST action IS0804 '*Language Impairment in a Multilingual Society*') has used EI to investigate morphosyntactic abilities of sequential bilingual children across a range of languages by creating parallel versions of EI tasks in several languages. The EI tasks of the network included a range of different structures grouped around two themes - structures that are syntactically complex across languages, such as relative clauses (*The swan that the deer chased knocked over the plant*) and object which-questions (*Which picture did he paint at home yesterday?*), and those involving embedding and/or syntactic movement. A second theme included structures that are challenging in specific languages only, mostly related to the morphological makeup of the language, for example tense and aspect marking in English (*The policeman has been looking at us*) and passives (*She was seen by the doctor in the morning*).

Participant groups. EI tasks can be used with both adults and children, and by both first and second language learners. However, care should be taken that the vocabulary, grammar and sentence length is not only appropriate for the participants' proficiency level but also for their age. If participants do not know specific words, the grammatical structure is beyond their proficiency level or the sentences are extremely long, it will be very difficult for them to imitate the sentence. As a result, the task will not be able to measure the participants' grammatical abilities, but their ability to keep in memory and retrieve novel words, hence their phonological memory. Moreover, hearing loss may affect general

listening comprehension in auditory paradigms, like the one in EI tasks, so that participants with age- or non-age related hearing loss may perform worse than those with better hearing, even though there may be no difference between the two groups in their L2 knowledge.

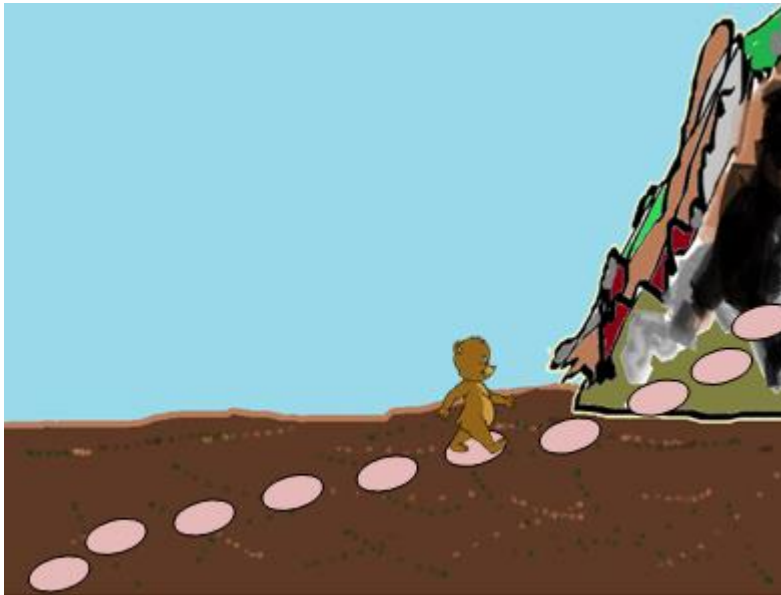
Procedure. The procedure in EI tasks is very straightforward: the sentences are read and learners have to try to repeat each sentence verbatim. To be able to analyse each grammatical structure separately, it is advisable to include 4-6 different sentences for each sentence type. For experimental purposes, it is useful to pre-record the sentences, so all learners can listen to the sentences in exactly the same way in terms of the speaker's voice, intonation, speed of presentation, and loudness. This can be done by playing the pre-recorded sentences from a CD/DVD/MP3 player/tablet or by incorporating the pre-recorded sentence into a PowerPoint presentation. The latter also enables the inclusion of visual support, so participants are aware of the length of the task, as shown in Figure 1 below.



**Figure 1** PowerPoint for EI in adults

Tasks with children can also have a story supported with pictures, as shown in Figure 2, to help ensure engagement with the task. Figure 2 comes from a set of EI tasks developed for bilingual children as part of the EU-funded COST Action ‘*Language Impairment in a Multilingual Society*’

(Marinis & Armon-Lotem, 2015). To make the EI task motivating for children, it was embedded within a treasure hunt game, in which a teddy bear was looking for a treasure in several locations.



**Figure 2** PowerPoint for EI in children

Figure 2 shows the slide, in which the teddy bear is moving towards a cave. Each stone corresponded to a sentence. When teddy jumped onto a stone, the children heard a sentence, which they had to repeat in order for the teddy bear to jump to the next stone. This motivated children to repeat all sentences in order to see where the treasure is and what it is. The sentences they had to repeat were not related in terms of meaning to the treasure hunt story.

Headphones can be used to ensure good quality listening to the sentences and enhance attention to the task for the participant. The repeated sentences should be recorded using high quality recording equipment to analyse at a later stage.

Data analysis. EI tasks can be scored in different ways depending on whether they are being used for research or for classroom assessment purposes, the focus of the study, and practical reasons, such as the amount of time available. Widely used schemes include giving a score of 1 if the sentence was repeated verbatim and 0 if there are any changes in the sentence. Scaled scoring schemes include



scores from 0-3, where a score of 3 is given if the sentence was repeated entirely verbatim, a score of 2 if there was one change, a score of 1 if there were 2 to 3 changes and a score of 0 if there were 4 or more changes in the repeated sentence. Structural scoring schemes assign 1 if the learner used the structure targeted irrespective of whether or not there were changes, e.g. in vocabulary, in other parts of the sentence. Unless pronunciation is the focus of EI assessment, non-target pronunciation is also ignored. A score of 0 is given if the learner makes an error in the sentence structure that was targeted, e.g. omission of an auxiliary in sentences targeting auxiliaries. A score of 0 is also given if the learner does not produce the structure targeted and substitutes it with another structure, e.g. substituting an object relative clause (*The swan that the deer chased knocked over the plant*) with a subject relative clause (*The swan that chased the deer knocked over the plant*).

Strengths and limitations. EI tasks have several advantages compared to other types of tasks, but also limitations. They are quick and easy to administer and they are not very demanding in terms of the procedure. They have clear target sentences, they can include a large range of sentence types, and they can be scored in several ways depending on the focus of the analysis. The most important limitation regards the challenge of optimizing the length of the stimulus sentence, that is, the relationship between the developmental stage of the participant and the sentence length. If sentences are too short for the learners' language abilities, they may be able to repeat them passively and the results will not reflect the participant's proficiency. The opposite effect may also occur. If the sentences are too long, they may exceed the participant's memory capacity. Participants may fail to repeat sentences not because they have not acquired a particular structure, such as auxiliaries in the English task, but because of the length of the sentence. This can be partially addressed through structural scoring, which identifies if participants were able to consistently (across similar test items) use the specific structure irrespective of whether or not they repeated the sentence verbatim.

### *Syntactic priming*

Syntactic Priming (SP) tasks, similarly to the EI tasks, make use of the human tendency to imitate behaviour in our environment. However, unlike EI tasks that are designed to produce a verbatim repetition of a sentence uttered by the experimenter, SP tasks measure the participant's unconscious imitation of a specific syntactic structure used by the experimenter while describing a picture. For example, the experimenter describes a picture of a giraffe pushing an elephant using a sentence in the passive (*the elephant was pushed by the giraffe*). Then the learner has to describe a picture of a similar event, for example a tiger kicking a lion. What is being measured is whether the learner will use the same structure used by the experimenter and produce a sentence in the passive (*the lion was kicked by the tiger*) or an alternative one (*the tiger was kicking the lion*). The sentence produced by the experimenter is called the *prime* whereas the sentence produced by the participant is called the *target*. The rationale behind using SP tasks is similar to the rationale in EI tasks; if a specific structure is part of the learner's grammatical system, the learner will use it to describe a picture. If, on the other hand, a structure is not part of the learner's grammatical system, the learner will not use it to describe a picture even though this structure was used previously by the experimenter. Instead, the learner will use an alternative structure that is part of their grammatical system. This alternative structure is typically less complex than the target structure.

Similarly to EI tasks, SP tasks require from the participant the ability to analyse sentences at all levels of representation (phonological, morpho-syntactic, semantic), extract the meaning and use the production system to describe a new picture. Unlike EI tasks, SP tasks do not require storage and retrieval of specific words. Instead, the *prime* is assumed to activate an abstract syntactic representation in the learner's grammatical system and as a result, the learner is more likely to use the same syntactic representation than an alternative one. If the learner mirrors the experimenter and uses the same syntactic structure to describe a new picture, this provides evidence that the specific structure is part of the learner's repertoire. Sentences do not have to be long in SP tasks because learners do not have to repeat a sentence, but describe a picture that was not already described by the experimenter.

Phenomena. SP tasks can be used with any phenomena that can be expressed using two different sentence structures. Two phenomena that have been studied extensively using SP tasks are transitive constructions that can be expressed using an active (*the tiger was kicking the lion*) or a passive sentence (*the lion was kicked by the tiger*) (e.g., Savage, et al., 2003) and double-object constructions that can be expressed either using dative in double-object constructions (*the girl gave the boy the book*) or prepositional-object constructions (*the girl gave the book to the boy*) (e.g., Thothathiri & Snedeker, 2008). For second language acquisition research it is important to take into account how the L1 may influence syntactic priming in the L2.

Participant groups. SP tasks can be used with both children (e.g., Bencini & Valian, 2008; Shimpi *et al.*, 2007) and adults (e.g., Bock, 1986; Branigan, 2007; Branigan *et al.*, 2000; Pickering & Ferreira, 2008), including second language learners. As with the EI task, the sentences should have appropriate vocabulary for the target group. SP tasks have been successfully used with children as low as 3 years of age (e.g., Savage *et al.*, 2003) and L2 learners with a range of proficiency levels (e.g., Hartsuiker *et al.*, 2004; McDonough, 2006; McDonough & Mackey, 2008).

Procedure. SP tasks typically involve the experimenter and then the participant describing series of pictures. The experimenter describes a picture first and then the participant describes the next picture. What is being measured is whether or not the participant will be *primed* by the experimenter, that is, whether or not the participant will use the same structure used by the experimenter to describe the picture or an alternative structure (e.g., passive vs. active). Therefore, it is crucial that both sentence structures are equally appropriate to describe the event in the picture. Apart from picture description, SP tasks can also involve written or oral sentence completion and dialogue with a confederate (e.g., Branigan *et al.*, 2000). To observe priming effects, it is advisable to use at least 10 different sentences for each of the alternative structures.

Priming between the two alternative structures can be tested in a randomised order or in separate blocks, for example ten trials priming passives followed by ten trials priming actives with a break in between the two blocks or the two blocks presented on different days or in two different

groups of participants. Depending on the aim of the study, there may be overlap in the words of the prime and the target or the prime and target may have different words. Overlap in the words between the prime and target (e.g. prime: *The elephant was kicked by the giraffe*; target: *the lion was kicked by the tiger*) usually leads to a stronger priming effect because in such cases there is priming not only from the structure (e.g. passive), but also from the specific lexical items used; in this case the same verb (*kicked*) is used in the prime and the target. Therefore, the clearest evidence for priming of the abstract structure is provided when there is no lexical overlap, but only structural overlap between the prime and the target (e.g. prime: *The elephant was kicked by the giraffe*; target: *the lion was pushed by the tiger*).

Unlike EI tasks, sentences in SP tasks are usually not pre-recorded; instead the experimenter and participant describe pictures in turn. As in all production tasks, high quality recording equipment should be used to record the participants' production in order to be able to conduct analyses after the task has been completed.

Data analysis. In SP tasks the data are not analysed in terms of accuracy or grammaticality, but in terms of whether the percentage of use of a specific structure, e.g. passives, is higher when it is preceded by primes with the same structure (passives) than when it is preceded by primes with a different structure (actives).

Strengths and limitations. The most important strength of the SP tasks is that, similarly to the EI tasks, they allow us to investigate structures that may not be frequent in naturalistic speech yet are part of a native speaker's linguistic competence. In addition, in the SP task we have specific target sentences that we try to elicit. A second important strength is that SP tasks provide evidence for the availability of a syntactic structure by tapping into the speakers' unconscious knowledge. Practical advantages of SP tasks are that they are easy to administer and score. The prime sentences are also relatively easy to construct, much easier than in EI tasks because the sentence length in SP tasks is not a crucial variable to control for. SP tasks usually involve a small number of sentence types, e.g. passives and actives, rather than a large number of structures, which is often the case with EI tasks.

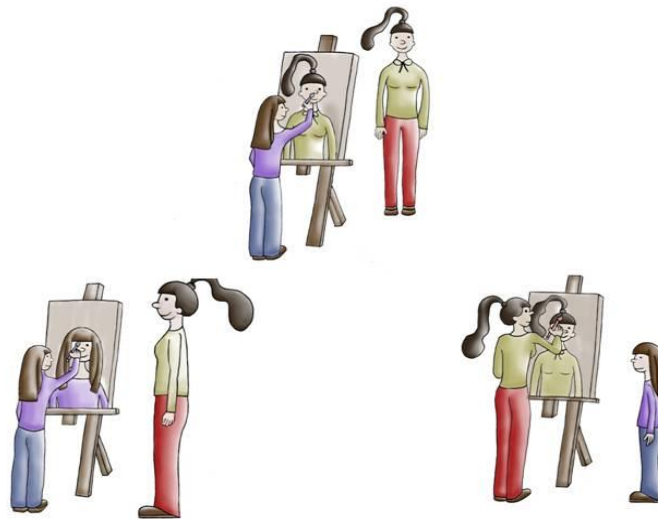
A challenging issue in SP tasks is the preparation of pictures that go with the sentences. This may require a professional illustrator to ensure that the pairs of pictures of the prime and target sentences are matched on visual features, such as size, picture type, similarity of the event. An important disadvantage of SP tasks is that priming of one structure may be so effective that participants perseverate and use the same structure in consecutive sentences. This can occur when the SP task consists of randomised lists of sentences or blocks. This can be avoided if a large number of filler sentences unrelated to the focus of the investigation/assessment are included in between the experimental sentences that include a range of other structures. These filler sentences can block effects of perseveration. A further limitation is that it cannot include a large number of structure types within the same task without resulting in a very long test, and crucially it can only address phenomena whose meanings are expressed through two types of structures.

## **Psycholinguistic techniques measuring second language comprehension**

As with production tasks, many different types of comprehension tasks have been designed within the psycholinguistics tradition that can also be used by language teachers and research students. Examples of comprehension tasks are picture selection tasks, picture verification tasks, self-paced reading and self-paced listening tasks, eye-tracking while reading, and eye-tracking while listening tasks, to mention just a few. Each has specific task demands and measures language comprehension in a slightly different way. Each task also has its advantages and disadvantages and may therefore be best suited for specific groups of learners. This chapter will introduce and discuss two of these tasks, *picture selection* and *self-paced reading*.

### *Picture selection task*

Picture selection (PS) tasks test whether learners are able to comprehend words or sentences accurately. Learners typically listen to or read a word or a sentence and look at a set of pictures, as shown in Figure 3.



**Figure 3** Sample picture selection task, comprehension of reflexive pronouns

Figure 3 comes from a PS task investigating the comprehension of reflexive pronouns (*himself/herself*) in children with autism (Terzi et al., 2014). The target sentence here was *Maria is painting herself* and the pictures show the same two participants in all pictures, Maria (the girl) and her godmother (the adult), taking part in an action of painting. Learners were expected to select the picture on the left hand side, in which Maria is painting herself while her godmother is standing next to her. Apart from the target picture, PS tasks include at least one more picture that acts as a distractor/foil. Distractors are crucial in PS tasks because they provide an alternative interpretation for the word or sentence. For example, Figure 3 has two distractors testing two different interpretations. The picture in the middle tests whether the participants will interpret the reflexive pronoun *herself* as the personal pronoun *her* (*Mary is painting her*). The picture on the right tests whether they will apply the reflexive interpretation to the other person in the picture (*The godmother is painting herself*). This

picture is included to ensure participants understand who the intended agent of the ‘painting’ action is. In this example the key competitor was the picture with the pronoun interpretation because previous research has shown that some children with autism interpret reflexive pronouns, such as herself, as personal pronouns, such as her (Perovic *et al.*, 2013). If participants have not acquired the properties of reflexives and interpret them as personal pronouns, they should select the picture in the middle.

The rationale of PS tasks is that the participants’ grammatical system will guide their picture selection. To perform accurately in PS tasks, participants have to be able to analyse the words or sentences at all levels (phonological, morpho-syntactic, semantic) and extract the meaning of the word/sentence using their grammatical system. Moreover, they have to scan the pictures, map the meaning of the word/sentence onto the pictures and select the picture that best represents the meaning they have created.

The difficulty of a particular PS task depends on the type and number of distractors. The closer the competitor to the target, the more difficult the task becomes because participants have to compare pictures corresponding to slightly different meanings. For example, in tasks investigating the comprehension of relative clauses (e.g., *The elephant that the giraffe is chasing is pushing the rhino*), it is possible to have several pictures as competitors that have meanings close to the target: target = giraffe chasing elephant, elephant pushing rhino; distractor 1: elephant chasing giraffe, giraffe pushing rhino, distractor 2: elephant chasing giraffe, elephant pushing rhino, distractor 3: giraffe chasing elephant, elephant pushing rhino. The difficulty also increases with the number of pictures in an array because the more pictures participants have to scan before making a decision, the more interpretations they have to make and the more time it takes to select a picture.

Similarly to EI tasks, PS tasks require storage of specific words/sentences and their meaning until participants select a picture. Thus, PS tasks have memory requirements. The more pictures included, the more the memory demands because it will take longer for participants to scan the pictures and select the right one. Apart from memory, PS tasks place demands on attention, especially when there is a large degree of similarity between the pictures. Participants have to scan the pictures carefully to identify the differences between them, which requires high levels of attention.

Phenomena. PS tasks can be used with any phenomena that can be depicted using words in isolation or sentences. The British Picture Vocabulary Scales III (Dunn *et al.*, 2009), and its American counterpart, the Peabody Picture Vocabulary Test IV (Dunn & Dunn, 2007) is a good example of a PS task at the word level because it includes a range of words belonging to different syntactic categories including nouns and verbs. Similarly, the Test for Reception of Grammar 2 (Bishop, 2003) is a good example of a PS task at the sentence level because it includes sentences of different types.

Participant groups. PS tasks can be used in language acquisition research with both children and adults (e.g., Friedmann & Novogrodsky, 2004; van der Lely, 1996), including second language learners (e.g., Papadopoulou *et al.*, 2011; Verhagen, 2013). PS tasks have been used with children as low as 3 years of age (e.g., Johnson *et al.*, 2005).

Procedure. PS tasks involve listening to or reading a word or sentence and selecting the picture that fits best with the word or picture from a set of at least two pictures. It is advisable to include at least 6 different sentences for each structure in order to have a representative sample of data. If the PS task targets only one structure, for example passives, it is advisable to include a number of filler items targeting other structures to avoid participants guessing the purpose of the task. The ratio between experimental items and fillers should be at least 1:1 for adult participants who are likely to be looking for a pattern in the sentences and for clues for the purpose of the study. For PS tasks with children, the number of filler items may be smaller because they are less likely to be thinking about the purpose of the study. Moreover, the length of tasks should be kept as short as possible with children to avoid fatigue.

It is good practice to pre-record the words/sentences in PS tasks when they are presented orally so all participants listen to the words/sentences in exactly the same way. If the words/sentences are to be presented visually on a computer, they can be presented at the bottom of the screen under the pictures. If the researcher/assessor wants to allow participants time to think about the words/sentences without making high memory demands on the participant, the words/sentences can stay onscreen until a picture is selected. Alternatively, the words/sentences can be presented for a limited, set amount of



time. Software, such as PowerPoint or E-Prime (Psychology Software Tools), can be used for the presentation of the words/sentences and pictures.

Participants' responses can be recorded by the experimenter or by the participant using paper and pencil. This is the simplest procedure for recording participants' responses. Responses can also be recorded using a keyboard, mouse, or button box and computer software. If there is interest in recording the participants' reaction time in selecting the picture, then experimental software, such as E-Prime (Psychology Software Tools), should be used to ensure accurate data logging. Reaction time data can be more sensitive than accuracy data because they show us how fast people make a response. Robust knowledge of a structure enables the participant to react quickly in selecting the right picture, whereas uncertainty about a structure can lead to slower reaction times.

Data analysis. In PS tasks the data are analysed in terms of accuracy. If reaction times are recorded, the data can also be analysed in terms of reaction times.

Strengths and limitations. PS tasks have several strengths. They can be used to test structures that learners are not otherwise using in their oral or written production, and thus, it is difficult to know if they have been acquired. They can include a range of different words and structures within the same task. At the practical level, PS tasks are relatively easy to set up and quick to administer because the only thing learners have to do is point to pictures or press a button to select a picture. Creating pictures may be challenging and costly if the pictures are created by a professional illustrator. An alternative is to create pictures using photographs. In this case, care should be taken that the photographs are of high quality with a neutral background and the actions and objects are clearly recognisable. An important limitation of the task is that it is not usually timed and learners can thus take their time to choose the correct picture and may develop strategies to complete the task. For example, a participant may select a picture by thinking and rejecting which pictures cannot go with the word/sentence instead of selecting a picture based on their knowledge of the word/picture tested. In other words, they can use their general knowledge of language and metalinguistic awareness to

select a picture. Alternatively, it is possible to introduce a time-limit, so participants have to make a fast response based on their intuition and implicit knowledge, to try and minimise strategic behaviour.

### *Self-paced reading task*

Self-paced reading (SPR) is a task used to measure the time taken to read a text or part thereof. The rationale behind SPR is that difficult portions of text will take longer to read, and the researcher or assessor can draw inferences regarding the cognitive processes underlying language comprehension at any given point in a sentence by recording how long a participant takes to read each portion of text. While tasks such as PS provide an explicit measure of language comprehension, SPR provides an implicit measure of the processes involved in successful comprehension (Marinis, 2010).

In a typical SPR task, participants read a series of texts one segment at a time, pushing a key or button to move through the text segment by segment. The reaction time taken to press the key or button at each segment can then be used to gauge the relative ease or difficulty of processing at different points in time in the stimulus sentence. In the SPR task, participants are first presented with a piece of text covered by a mask, usually a series of Xs, as in Figure 4a. The participant then presses a key or button to read the first segment. A segment can be either an individual word, phrase, sentence or paragraph depending on the needs of the researcher or assessor. Figure 4 exemplifies word-by-word SPR. When the participant first presses the response key, the first word of the sentence appears, as in Figure 4b. Each subsequent key-press reveals the sentence, one word at a time, covering up the previously read word. The reaction time taken to press the response key at each word provides an implicit measure of the cognitive processes underlying language comprehension. Inferences about language processing can be made by comparing reading times at critical portions of text in maximally similar sentences. For example, reaction times at or shortly after the pronoun ‘she’ in Figure 4 will likely be longer in this sentence as compared to a sentence containing a gender-matching antecedent (e.g. *Jane mentioned that she was tired*).

- a) XXXX XXXXXXXXXXXX XXXX XXX XXX XXXXX.
- b) John XXXXXXXXXXXX XXXX XXX XXX XXXXX.
- c) XXXX mentioned XXXX XXX XXX XXXXX.
- d) XXXX XXXXXXXXXXXX that XXX XXX XXXXX.
- e) XXXX XXXXXXXXXXXX XXXX she XXX XXXXX.
- f) XXXX XXXXXXXXXXXX XXXX XXX was XXXXX.
- g) XXXX XXXXXXXXXXXX XXXX XXX XXX tired.

**Figure 4** Example word-by-word self-paced reading task studying pronoun-reference resolution

To ensure that participants pay attention to what they read, a comprehension question is typically asked after each trial. This question can require a yes/no or true/false answer, or more complex responses depending on the needs of the researcher/assessor.

Phenomena. Provided adequate reading ability, SPR can be used to investigate a wide variety of linguistic phenomena. It has been used to investigate how adult native speakers and second-language learners process and interpret different types of ambiguous sentences (e.g. Felser *et al.*, 2003; Roberts & Felser, 2011). It has also been used to investigate the acquisition of different morphosyntactic features, such as gender/number agreement (e.g. Sagarra & Herschensohn, 2010). In such cases, the researcher can compare reaction times during reading of sentences containing grammatical and ungrammatical agreement (e.g. *The boy unsurprisingly was late to school* vs. *The boys unsurprisingly was late to school*) as an implicit measure of the acquisition of a particular phenomenon, without the need to require participants to make an explicit sentence judgment. SPR has also been widely used to investigate different linguistic dependencies, such as anaphora resolution and syntactic movement (e.g. Marinis *et al.*, 2005).

Participant groups. SPR can be used to investigate language comprehension in adult native speakers and different groups of adult second language learners (Marinis, 2003, 2010, 2013). It has also been used with children as young as 8 (Traxler, 2002).

Procedure. The recording of reaction time data requires specialist software to ensure that reaction times are recorded accurately. SPR tasks are thus usually conducted using experimental software, such as E-Prime (Psychology Software Tools). A free alternative to E-Prime that was developed primarily to run SPR tasks and that is widely used in psycholinguistics is Linger (Rohde, 2005).

SPR tasks involve participants reading sentences or larger pieces of text one segment at a time, pressing a response key to move from one segment to the next. Before beginning a study, the researcher will need to decide how large each segment to-be-displayed should be based on their research questions and the populations being studied. Although larger segments may be easier to process, increasingly large segments provide less fine-grained information regarding the time-course of language processing compared to smaller segments, such as word-by-word presentation.

To ensure participants do not become explicitly aware of the manipulations being studied, it is important to include a sufficient number of fillers in addition to the experimental items of main interest. In a typical SPR task, the researcher will want to include at least 6-8 sentences in each experimental condition. Fillers should include a wide variety of syntactic structures to ensure participants do not become habituated to the experimental sentences, i.e. that they pay attention to the sentences yet do not figure out which structures are being tested. Usually fillers should be included at a ratio of at least 2:1 for adults, although fewer fillers can be used if the researcher fears the study will become too long.

The post-item comprehension question can differ depending on the needs of the researcher. Often these questions are only included to ensure that participants pay attention to the sentences and the answers may not be of primary interest other than to ensure a certain threshold level of performance (e.g. 75% correct). However, depending on the research question, the researcher may want to also include questions that probe a specific aspect of comprehension. In Figure 4, for example, a question could probe interpretation of the pronoun (e.g. *Was John tired?*). Researchers

should, however, be wary that the comprehension questions do not afford too much explicit attention to the experimental manipulations.

Data analysis. SPR tasks provide two sources of data. One source of data is the accuracy to the comprehension questions. This provides an ‘offline’, explicit measure of language comprehension. The second source of data, the reaction times at each segment, provides an ‘online’ measure of implicit language comprehension. Reaction times are difficult to analyse in absolute terms, and as such relative differences in reaction times need to be compared across experimental conditions. Typically, reaction times will be averaged at each segment, and compared across two or more conditions (e.g. grammatical vs. ungrammatical) to see if the conditions differ. This illustrates the need for specialist software when analysing data from SPR tasks. Extremely short or extremely long reaction times (e.g. 3.5 standard deviations above/below a participant’s mean reaction time for a given segment) are sometimes discarded, on the assumption that such ‘noisy’ data index either a button miss-press, or conscious awareness of the experimental manipulation. Reaction times to items in which incorrect responses to the comprehension question are given are also sometimes discarded.

Strengths and limitations. As the task involves reading, it is limited to participant groups who have attained sufficient levels of reading comprehension. It may thus not be appropriate for particularly young children or for those with reading impairments. An alternative to SPR is self-paced listening, where participants press a button to *listen to*, rather than read, a sentence one segment at a time, and might be more appropriate in such cases. Self-paced listening has been used successfully with monolingual and bilingual children as young as 6, including those with language impairment (e.g. Felser *et al.*, 2003; Marinis & Saddy, 2013). Although the overt button-press in SPR does not exactly mimic normal reading, SPR studies obtain results comparable to methods using more naturalistic reading, such as eye-tracking (Ferreira & Henderson, 1990).

As numerous factors influence reading times (Rayner, 1998), SPR should only be used to investigate sentences that are as maximally similar as possible, and critical regions of comparison (e.g. the pronoun ‘she’ in Figure 4) should ideally be identical across conditions. The researcher also

needs to be aware of ‘spillover’ effects, where effects of experimental manipulations appear in the segments after a critical region of interest (e.g. *was* and *tired* in Figure 4). As such, the text immediately after the critical region should ideally be identical across conditions as well.

The length and complexity of the sentences should also be considered, based on factors such as participant age and reading ability. As typical SPR studies use non-cumulative presentation, in which earlier segments of a sentence are masked as the participant reads, it does not allow rereading of earlier portions of text, unlike normal reading, which may cause increasing problems if sentences become too long.

## **Conclusions**

Psycholinguistic techniques offer a variety of ways of investigating second language acquisition and have been used in the assessment of language ability in a range of different populations of speakers. The tasks outlined in this chapter provide second language teachers with alternative ways of investigating and testing different linguistic phenomena in both second language production and comprehension that can complement existing assessment resources. Psycholinguistic techniques are particularly useful in providing implicit measures of language ability to gauge how well a student or group of students have truly internalised knowledge of the vocabulary and grammar of the language being learnt.

The psycholinguistic tasks outlined in this chapter provide those who are still studying to become teachers with tools that they can use in their research projects when investigating second language acquisition in different learner populations. These can be complemented with existing testing and assessment batteries, to examine the extent to which explicit measures of vocabulary and grammatical knowledge correlate with implicit psycholinguistic measures of language production and comprehension.

We have not discussed in detail here best practice in conducting statistical analyses of the different types of data obtained in typical psycholinguistic studies, or how such analyses can be correlated with standardised measures of language assessment. For those interested in going beyond

descriptive statistics and using inferential statistics to generalise study findings from a sample of learners to the wider population, we direct the interested reader to Larson-Hall (2015) and Plonsky (2015) for recent introductions to inferential statistical analysis, which both focus on research in second language acquisition. Cunnings (2012) and Linck and Cunnings (2015) also provide an overview of recent advances in the analysis of psycholinguistic data, focusing in particular on its application to research in second language acquisition.

Language teachers and those studying to become language teachers already have a number of assessment tools at their disposal that they can use to gauge learner abilities to help encourage successful language learning. We hope that the psycholinguistic tasks outlined here can complement these existing batteries and provide new insight into internalised, implicit levels of language ability that can provide inspiration in devising new ways of assessing second language acquisition.

## References

- Baddeley, A. D. (2000) The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4, 417-423.
- Bencini, G. and Valian, V. (2008) Abstract sentence representation in 3-year-olds: Evidence from comprehension and production. *Journal of Memory and Language* 59, 97-113.
- Bishop, D. (2003) *Test for Reception of Grammar, Version 2*. London: The Psychological Corporation.
- Bock, J. K. (1986) Syntactic persistence in language production. *Cognitive Psychology* 18, 355–387.
- Branigan, H. (2007) Syntactic Priming. *Language and Linguistics Compass* 1, 1-16.
- Branigan, H., Pickering, M. and Cleland, A. (2000) Syntactic co-ordination in dialogue. *Cognition*, 75, B13–B25.
- Bley-Vroman, R. and Chaudron, C. (1994) Elicited imitation as a measure of second-language competence. In E. Tarone, S. Gass, and A. Cohen (Eds.), *Research Methodology in Second-Language Acquisition*. Hillsdale, N.J.: Lawrence Erlbaum.
- Cunnings, I. (2012) An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28, 369-382.
- Dunn, L.M. and Dunn, D. M. (2007) *Peabody Picture Vocabulary Test IV*. Minneapolis, MN. : Pearson Assessments
- Dunn, L. M., Dunn, D. M., Styles, B. and Sewell, J. (2009) *The British Picture Vocabulary Scale III – (3<sup>rd</sup> Edition)*. London: GL Assessment.
- Felser, C., Marinis, T. and Clahsen, H. (2003) Children’s processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition* 11, 127-163.
- Felser, C., Roberts, L., Marinis, T. and Gross, R. (2003) The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24, 453-489.
- Ferreira, F. and Henderson, J. (1990) Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 555-568.
- Friedmann, N. and Novogrodsky, R. (2004) The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language* 31, 661-681.
- Hartsuiker, R., Pickering, M. and Veltkamp, E. (2004) Is syntax shared or separate between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science* 15, 409-414.
- Jessop, L., Suzuki, W. and Tomita, Y. (2007) Elicited Imitation in Second Language Acquisition Research. *The Canadian Modern Language Review* 64, 215–238.
- Johnson, V., de Villiers, J. and Seymour, H. (2005) Agreement without understanding? The case of third person singular /s/. *First Language* 25, 317–30.



- Larson-Hall, J. (2015) *A guide to doing statistics in second language research using SPSS and R, 2<sup>nd</sup> Edition*. New York: Routledge.
- Linck, J. and Cunnings, I. (2015) The utility and application of mixed effects models in second language research. *Language Learning* 65(S1), 185-207.
- Lust, B., Flynn, S. and Foley, C. (1996) What children know about what they say: Elicited imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee and H. Smith Cairns (eds.) *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press.
- Marinis, T. (2003) Psycholinguistic techniques in second language acquisition research. *Second Language Research* 19, 144-161.
- Marinis, T. (2010) On-line sentence processing methods in typical and atypical populations. In S. Unsworth and E. Blom (eds.) *Experimental Methods in Language Acquisition Research*. Amsterdam: John Benjamins.
- Marinis, T. (2013) Online Psycholinguistic Methods in Second Language Acquisition. In: C. Chapelle (ed.) *The Encyclopaedia of Applied Linguistics*. Wiley-Blackwell.
- Marinis, T. and Armon-Lotem, S. (2015) Sentence repetition. In S. Armon-Lotem, J. de Jong and N. Meir (eds.) *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Marinis, T., Roberts, L., Felser, C. and Clahsen, H. (2005) Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27, 53-78.
- Marinis, T. and Saddy, D. (2013) Parsing the passive: Comparing children with specific language impairment to sequential bilingual children. *Language Acquisition* 20, 155-179.
- McDonough, K. (2006) Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition* 28, 179-207.
- McDonough, K. and Mackey, A. (2008) Syntactic priming and ESL question development. *Studies in Second Language Acquisition* 30, 31-47.
- Papadopoulou, D., Varlokosta, S., Spyropoulos, V., Kaili, H., Prokou S. and Revithiadou, A. (2011) Case morphology and word order in L2 Turkish: Evidence from Greek learners. *Second Language Research* 25, 173-204.
- Perovic, A., Modyanova, N. and Wexler, K. (2013) Comprehension of reflexive and personal pronouns in children with autism: A syntactic or pragmatic deficit? *Applied Psycholinguistics* 34, 813-835.
- Pickering, M. and Ferreira, V. (2008) Structural priming: A critical review. *Psychological Bulletin* 134, 427-459.
- Plonsky, L. (2015) *Advancing quantitative methods in second language research*. New York: Routledge.
- Psychology Software Tools Inc. (2012) *E-Prime 2.0*. <http://www.pstnet.com>

- Rayner, K. (1998) Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 372-422.
- Roberts, L. and Felser, C. (2011) Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics* 32, 299-331.
- Rohde, D. (2005) Linger: a flexible platform for language processing experiments. MIT.  
<http://tedlab.mit.edu/~dr/Linger/>
- Sagarra, N. and Herschensohn, J. (2010) The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120, 2022-2039.
- Savage, C., Lieven, E., Theakston, A. and Tomasello, M. (2003) Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science* 6, 557-567.
- Shimpi, P., Gamez, P., Huttenlocher, J., and Vasilyeva, M. (2007) Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology* 43, 1334-1346.
- Terzi, A., Marinis, T., Kotsopoulou, A. and Francis, K. (2014) Grammatical abilities of Greek-speaking children with autism. *Language Acquisition* 21, 4-44.
- Thothathiri, M. and Snedeker, J. (2008) Give and take: syntactic priming during spoken language comprehension. *Cognition* 108, 51-68.
- Traxler, M. (2002) Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 55, 75-96.
- van der Lely, H. (1996) Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua* 98, 243-272.
- Verhagen, J. (2013) From dummy auxiliary to auxiliary in Moroccan adult learners' production and comprehension of Dutch. In E. Blom, I. van de Craats and J. Verhagen (eds.) *Dummy auxiliaries in first and second language development*. Berlin/New York: Mouton de Gruyter.