



## Primary Location of Difficulty in the Processing of Thai and English Relative Clauses

Pornsiri Singhapreecha<sup>a,\*</sup>, A. J. Benjamin Clarke<sup>b</sup>, Naparat Meechanyakul<sup>c</sup>, Yuki Hirose<sup>d</sup>

<sup>a</sup> [pornsiri.s@litu.tu.ac.th](mailto:pornsiri.s@litu.tu.ac.th), Language Institute, Thammasat University, Thailand

<sup>b</sup> [ajb.clarke@gmail.com](mailto:ajb.clarke@gmail.com), Department of Linguistics, Faculty of Liberal Arts, Thammasat University, Thailand

<sup>c</sup> [napat.m@litu.tu.ac.th](mailto:napat.m@litu.tu.ac.th), Language Institute, Thammasat University, Thailand

<sup>d</sup> [hirose@boz.c.u-tokyo.ac.jp](mailto:hirose@boz.c.u-tokyo.ac.jp), Graduate School of Arts and Sciences, University of Tokyo, Japan

\* Corresponding author, [pornsiri.s@litu.tu.ac.th](mailto:pornsiri.s@litu.tu.ac.th)

### APA Citation:

Singhapreecha, P., Clarke, A. J. B., Meechanyakul, N., & Hirose, Y. (2026). Primary location of difficulty in the processing of Thai and English relative clauses. *LEARN Journal: Language Education and Acquisition Research Network*, 19(1). 597-623. <https://doi.org/10.70730/JLDJ2047>

Received  
12/06/2025

Received in  
revised form  
28/11/2025

Accepted  
02/01/2026

### ABSTRACT

We investigated the processing of Thai and English subject- and object-extracted relative clauses (SRCs and ORCs) with native Thai and English speakers using a self-paced reading paradigm. Although both Thai and English exhibit post-nominal RCs, Thai uses neither determiners nor inflectional morphemes, which warrants further investigation. The processing advantage of the SRC over the ORC was predicted and confirmed, while the primary location of difficulty was predicted by the Locality-based Integration or Surprisal/Expectation models. For English, the results indicated that the primary source of difficulty was at the ORC embedded verbs, consistent with predictions from the Locality-based Integration account. Although a similar pattern

	<p>of findings was obtained for Thai, with significant differences between ORC and SRC at the embedded verbs, there was a tendency for the primary difficulty to occur at the ORC embedded nouns, which aligns more with the Surprisal/Expectation account. Thus, the absence of determiners (e.g., articles) in Thai appears to account for differences in the primary location of difficulty in the processing of Thai and English relative clauses. The English results are consistent with prior studies using phrase-by-phrase segmentation. The results with Thai, particularly from the by-item analysis, indicate a tendency toward an effect of word category frequency. We recommend that future studies investigate post-nominal RCs in other languages to shed light on language-specific properties (such as bare NPs in Thai) and to ascertain the role of word category frequency in initial sentence processing.</p> <p><b>Keywords:</b> Thai relative clauses, English relative clauses, self-paced reading, primary difficulty, Surprisal/Expectation model, Locality-based Integration model</p>
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## Introduction

In online sentence processing, participants tend to process texts immediately. As soon as the parser encounters each word in a sentence it integrates different sources of knowledge to construct an interpretation in a moment-by-moment fashion (Pickering & van Gompel, 2006). At the outset the parser may encounter some obstacles such as structural complexity, inducing a certain amount of delay. Relative clauses are one type of potential obstacle because they are quite complex in structure. In (1), on a standard assumption in generative grammar, there exists an empty category ( $e_i$ ) or a gap, which is coindexed with the head noun *the actor*, via the operator-variable binding fashion (Poole, 2002). This operation enables the identification of the gap to be restricted by the clause introduced by the complementizer *that* (or the filler). Sentences (1a) and (1b) are termed subject-extracted relative clauses (SRC) and object-extracted relative clauses (ORC), due to the raising of the operator from the embedded subject and object position, to bind the head noun, respectively.

- (1)     a. The actor<sub>i</sub> that  $e_i$  hated the writer complained about the film script.  
           b. The actor<sub>i</sub> that the writer hated  $e_i$  complained about the film script.

Semantically, in (1a) the head noun *the actor* bears the Experiencer role in both the matrix and embedded clauses, and *the writer* bears the Theme role. The sequence of Experiencer preceding Theme follows a common thematic hierarchy (e.g., the Uniformity of Thematic Hierarchy; Baker, 1988). In (1b), *the actor* bears the Agent role in the matrix clause and becomes the Theme in the embedded clause, while *the writer* is the Experiencer. That the head noun changes its semantic roles instigated by the embedded subject and verb in the relative clause in (1b) presumably imposes a challenge to the parser, unlike the head noun in (1a), whose semantic role is identical with the one in the gap. The semantic role change, an effect of the following relative clause, could potentially incur a disadvantage in the processing of the ORC such as (1b), over that of the SRC such as (1a). Structural complexity that is associated with object relative clauses such as (1b) may provide insight into the broader issue of what guides initial processing, an aim this study attempts to achieve.

In early experimental studies, subject and object relative clauses were employed to evaluate sentence parsing methods. Holmes and O'Regan (1981) conducted eye-tracking experiments with French speakers to assess whether a canonical sentence pattern would account for the asymmetry between the two structures. Ford (1983) examined whether the Continuous Lexical Decision technique, in which participants judge whether each letter string within a sentence is a word or non-word, was sensitive to structural complexity. The disadvantage of the ORC was suggested by more comprehension errors in Holmes and O'Regan's (1981) study and longer response times at the word level in Ford's (1983) study.

Subsequent research directly investigated whether SRC or ORC sentences induce longer reading times, consistently finding that ORCs were read more slowly than SRCs in studies with English as the language of investigation. In self-paced reading (SPR) experiments, ORC verbs and SRC nouns (i.e., object complements of verbs in SRCs) occupy the same critical region—the region for *hated* and *the writer* in the earlier sentence (1). Exploiting this property, King and Just (1991) found that the ORC verb and the matrix verb involved longer reading times than the SRC noun and the matrix verb. Gordon et al. (2001), using the same critical words, similarly found slower reading times for ORC verbs compared with SRC nouns. They further extended the analysis beyond individual words to entire relative clauses (e.g., *the barber praised* vs *praised the barber*), demonstrating that ORCs continued to be processed more slowly than SRCs. Traxler et al. (2002) used eye-tracking to examine processing across larger critical regions, including all words in the relative clauses and matrix verbs, except the relative pronoun. They found that participants spent more time on ORCs and ORC matrix verbs than on SRCs and SRC matrix verbs, with animacy controlled across clause types.

However, studies with speakers of Chinese (Hsiao & Gibson, 2003; Xu et al., 2019), Japanese (Miyamoto & Nakamura, 2003), and Korean (Kwon et al., 2013), using self-paced reading paradigms or event-related potentials (ERPs), have yielded inconclusive evidence on whether SRCs are less complex than ORCs. These languages differ in their word order patterns (i.e., SVO for Chinese, SOV for Japanese and Korean), and in relative clause structures, including pre-nominal RCs, case marking, and the presence or absence of relative markers, yet the syntactic distance between filler and gap is consistently longer in SRCs than in ORCs. The Locality-based Integration model predicts greater difficulty for SRCs than ORCs, due to the longer distance required to integrate resources (details provided in the discussion of sentence (3)). In Miyamoto and Nakamura (2003) and Kwon et al. (2013), ORCs were processed more slowly than SRCs. By contrast, Hsiao and Gibson (2003) and Xu et al. (2019) reported faster processing for ORCs than for SRCs. These mixed results suggest that the Locality-based Integration model better accounts for the Chinese data than for Japanese or Korean. Xu et al. (2019) attribute the SRC or ORC advantage to language-specific differences and stress that these topological differences should be further investigated in sentence processing.

While data from pre-nominal RCs remain unclear regarding whether SRCs are less complex than ORCs, data from post-nominal RCs suggest a more consistent asymmetry. In line with the ORC–SRC complexity contrast, additional investigations have probed whether ORCs become less costly when modulated by working memory (WM) capacity (King & Just, 1991) and by the animacy of noun phrases (Gennari & MacDonald, 2008; Traxler et al., 2002). King and Just (1991) found that readers with higher WM capacity processed both the embedded verb and the matrix verb in ORC sentences faster than those with low WM capacity, suggesting reduced ORC processing difficulty among readers with greater working memory resources. Traxler et al. (2002) found a decrease in difficulty when an inanimate NP occupied the sentential subject, as did Gennari and MacDonald (2008) when an inanimate NP occupied the relative clause subject. A decrease in difficulty in the ORC when the sentential or the embedded subject is inanimate, in turn, suggests that it is important to use animate sentential and embedded subjects to ensure sufficient processing demands arise for the ORC, a point taken into account in our study.

In addition to investigating working memory capacity and animacy as factors, several previous studies have identified the ORC verb as the primary locus of processing difficulty (Holmes & O'Regan, 1981; Ford, 1983; Gordon et al., 2001; Grodner & Gibson, 2005). Other studies have pointed to the ORC subject, or both the ORC subject and verb, as loci of processing difficulty (Forster et al., 2009; Gennari & MacDonald, 2008; Staub, 2010).

Notably, English embedded subjects in ORC trials of studies conducted with English speakers (e.g., Forster et al. 2009; Gennari & MacDonald, 2008; Gordon et al., 2001; Grodner & Gibson, 2005) include determiners (i.e., the definite article *the*). This raises the question of where within these trials processing costs arise, as the data may inform how the remaining strings are affected. Several studies (Gennari & MacDonald, 2008; Gordon et al., 2001; Grodner & Gibson, 2005) have reported reading times for *the* in combination with nouns, while Forster et al. (2009) report reading times for *the* per se.

Forster et al. (2009) employed a maze task, a variant version of the self-paced reading task, in which participants choose one of two words as a possible continuation of the preceding word. For example, in Forster et al. (2009), an ORC trial: *The .../banker amazes/ that truck/ the across/ lawyer inside/ irritated because/ played none/ tennis loosens/ every yes/ Saturday. caught./* was presented (aiming at the target *The banker that the lawyer irritated played tennis every Saturday*). After the pair *that truck* was presented and *that* was selected, participants would then choose *the* as the next possible continuation. Forster et al. (2009) found a longer reading time at the determiner *the* in the ORC than in the SRC counterpart, while the noun and verb showed relatively similar reading times across ORCs and SRCs. They therefore argue that ORC difficulty may arise at the very onset of the embedded clause, namely at the determiner. This finding is particularly relevant for the present study, as it highlights a potential source of processing cost that is absent in languages without determiners, such as Thai. The question is whether the presence of Thai NPs without determiners (i.e., bare nouns) will also involve cost.

In terms of theoretical models that address the location of difficulty in relative clauses, approaches are numerous. The Active Filler Strategy (Frazier, 1987), Perspective Shifting (MacWhinney, 2005), and Surprisal/Expectation approaches (Hale, 2001; Levy, 2008) all predict the source of difficulty at the ORC subject. The Locality-based Integration (Gibson, 1998; Grodner & Gibson, 2005) and the Adaptive Control of Thought–Rational (Lewis & Vasishth, 2005) predict difficulty at the OR verb, and at both the ORC subject and verb, respectively. In the present study, we restrict our predictions to the Surprisal/Expectation and Locality-based Integration approaches, as they predict clearly different locations of difficulty. In terms of initial processing, the Surprisal/Expectation and Locality-based Integration approaches may represent the role of input frequency (or conformity to incoming input) and the integration of resources in linear order, respectively.

With respect to the Surprisal/Expectation approach (Hale, 2001; Levy, 2008), the parser's pace depends on the frequency of encountering the structure in question. Levy (2008) developed a formula, based on Hale's

(2001) Earley Parser, which takes into account the probability of the immediately preceding input, a negative log probability, and the frequencies of given structures from a corpus (e.g., the Penn Treebank project). This formula assigns magnitudes, corresponding to degrees of difficulty, to words in linear order. Levy (2008) identifies difficulty of a word (e.g.,  $w_i$ ) to *surprisal*, which he defines quantitatively as the negative log-probability of  $w_i$  in its sentential context (i.e., the immediately preceding input sequence  $w_1 \dots w_{i-1}$ ) and extra-sentential context. As Levy (2008) remarks, *surprisal* approaches 0 when a word is obligatory—when the probability of the immediately preceding input along with the extra sentential context is 1— and gets close to infinity as a word becomes highly unlikely.

Magnitudes by Hale's (2001) Earley Parser standards are displayed in (2a) and (2b) below. Note that the magnitude values of the individual words appear in parentheses. As shown in (2b), the ORC embedded subject, with a value of 4.53, has the greatest degree of difficulty/surprisal among all the words, and is predicted to be more difficult to process than those with values of 1.6 or below (e.g., *the* and *me*).

- (2) a. The (1.60) man (0) who (1) saw (0.20) you (1.60) saw (1) me (1)  
 b. The (1.60) man (0) who (1) you (4.53) saw (0) saw (1) me (1.60)

In addition to the above formula, another crucial element is the relative pronoun, which creates a clausal domain. According to Levy (2008), when the parser reaches the relative pronoun, it expects a verb. However, exactly what kind of verb is expected and whether a subject will precede it remains uncertain. This uncertainty causes the parser to process the embedded verb in the SRC more slowly than its counterpart in the ORC, because the SRC verb follows immediately after the relative pronoun, whereas the ORC verb is preceded by the subject NP. In addition, the parser does not expect object relative clauses. When it encounters the embedded subject, the parser slows considerably upon encountering an unexpected element after maintaining the incorrect expectation. Broadly speaking, in the Surprisal/Expectation model, the frequency with which an upcoming input appears in the context of the prior material determines the degree of processing complexity. The greatest difficulty arises when the input deviates from the expected preceding structure or context.

While the Surprisal/Expectation approach identifies the ORC subject as the locus of greater difficulty, the Locality-based Integration account (Gibson, 1998; Grodner & Gibson, 2005) predicts the ORC verb to be the costliest region. According to this account, the parser faces perceptual limitations in processing lexical resources, and linear order distance

determines the cost of integrating one resource with another for comprehension. When new sentences such as (3a) and (3b), reproduced from (1a) and (1b), are presented, the lexical resources (measured in units) are treated as new to the discourse events depicted in the sentences. Note that the Locality-based Integration approach is generally referred to as a resource-based model because it relies on the integration of lexical resources for comprehension.

- (3) a. The (0) actor (1) that (0) hated (1) the (0) writer (1) complained  
(3) about (1) the (0) film script (1)  
b. The (0) writer (1) that (0) the (0) actor (1) hated (1+2) complained  
(3) about (1) the (0) film script (1)

According to the Locality-based Integration approach, with respect to the head and the RC part of (3a), three new discourse referents are introduced: *the actor*, *hated*, *the writer*. Each element incurs one unit; *the* does not carry a unit independently. From left to right, when the parser reaches the matrix verb *complained*, it associates *complained* with *the actor*. To enable this linking, the parser integrates three units: a hating state, a writer, and a complaining event. In the remaining PP, *about* and *film script* incur two units, as they are also new discourse referents. In (3b), while the integration costs of the matrix clause subject and the PP are the same as in (3a), the difference between the two sentence types involves the embedded verb *hated*. In (3b), one unit occurs when *hated* crosses its subject *the actor* (a new discourse referent) to enable a local link. Subsequently, the object of *hated*, is linked to the complementizer *that*, which is coindexed with *the writer*. This integration crosses the embedded verb and the embedded subject, amounting to two units.

Given the above integration costs, the Locality-based Integration approach predicts two sources of difficulty: the ORC verb and the matrix verbs of both the ORC and SRC. In essence, if integration costs arise from limited computational resources, the greatest difficulty should occur when distant dependents are associated for comprehension, such as the linking of the object of *hated* with *that*. As Grodner and Gibson (2005) remark, this type of association typically occurs at a point where the parser can easily predict its syntactic category, such as the ORC verb, given the prior context.

The above approaches make differing predictions regarding the ORC, prompting Staub (2010) to question where exactly the difficulty lies; that is, the RC verb, the region posited by resource-based models (e.g., the Locality-based Integration), the RC subject, the region posited by Surprisal/Expectation model, or both. He investigated native English speakers' processing of SRCs and ORCs using the eye-tracking paradigm and

found that mean go-past times for the determiner and noun regions in ORCs were much longer than their SRC counterparts. In the relative clause verb region, the three means (i.e., first fixation, gaze, and go-past) for ORCs were reliably longer than the corresponding SRC measures. All the means of the matrix verb regions revealed non-significant differences.

Staub's (2010) findings, therefore, support the alternative prediction, that both the RC subject and verb equally induce difficulty. Notably, there were substantial regressive saccades from the determiner and NP subject, and an inflated first-pass reading time for the relative verb. Staub (2010) remarks that such differences are crucial because tasks such as SPR alone could not reveal them, and they are presumably associated with different theoretical approaches (e.g., the Surprisal/Expectation and Locality-based Integration models). We also examined the primary location of difficulty in English SRCs and ORCs and their Thai counterparts using the SPR paradigm.

### Present Study

The data from post-nominal RCs (Ford, 1983; Gordon et al., 2001; Holmes & O'Regan, 1981; King & Just, 1991; Traxler et al., 2002) provide supporting evidence that the SRC has an advantage over the ORC in reading experiments. Determining the primary location of difficulty remains unsettled, which motivates further investigation in a language such as Thai, given its grammatical differences from English. SRCs and ORCs in Thai and English are actually structurally similar, as they are both post-nominals. However, as seen in (1), determiners are not expressed in Thai; thus, bare nouns (e.g., *náksadæɛŋ* and *nákkhǎan*) can appear independently, as in (4).

#### (4) Thai SRC and ORC

a. *náksadæɛŋ*<sub>i</sub> [RC *thǐi* *e<sub>i</sub>* *kliat nákkhǎan*]  
actor COMP hate writer

*tamni rúan bòtlakhǎon*  
blame matter script.play

“The actor that hated the writer complained about the film script.”

b. *náksadæɛŋ*<sub>i</sub> [RC *thǐi* *nákkhǎan kliat e<sub>i</sub>*]  
actor COMP writer hate

*tamni rúan bòtlakhǎon*  
blame matter script.play

“The actor that the writer hated complained about the film script.”

The head noun *náksadēy* is modified by the following RC in both (4a) and (4b); the operator-variable binding mechanism enables restriction in the identification of the head noun. Regarding definiteness, when a bare noun such as *náksadēy* appears without an RC, it receives a definite singular reading by default (Singhapreecha & Sybesma, 2015). For example, in a context where a book and a magazine are lying on the table, a speaker may say “*nǎṅsūuu pen khǒṅ chǎn*” meaning “the book is mine,” in which the bare noun *nǎṅsūuu*, corresponding to “the book”, has a definite singular reading.

In the context of (4) where an RC is available, its definite property is mainly derived from the relative clause, rather than its default property, while past time reference can be implied in the absence of a time adverb. The presence and absence of determiners in English and Thai may affect reading times and the point at which difficulty emerges, especially in the ORC in both languages.

As noted earlier, both Japanese and Chinese have pre-nominal RCs. Reading studies with speakers of these languages (e.g., Hsiao & Gibson, 2003; Miyamoto & Nakamura, 2003) therefore inform the question of whether SRCs are less complex than ORCs. Evidence from Thai addresses another key issue: the primary locus of processing difficulty. As discussed earlier, Forster et al. (2009) found that the definite article incurred longer reading times in ORCs than in SRCs, while the following nouns showed similar reading times. Compared with English NPs with definite articles, Thai definite NPs without determiners are expected to require substantial processing time. In linear order, the absence of a determiner in Thai allows the NP subject in the ORC to follow the complementizer immediately, unlike in the English ORC, where the definite article intervenes between the complementizer and the noun. Given this adjacency between the complementizer and bare NPs, the Locality-based Integration and Surprisal/Expectation models can be evaluated straightforwardly using Thai. Consequently, the findings will have implications for the explanatory power of these theoretical models and provide evidence about the factors guiding initial sentence processing. We take these cross-linguistic differences into account in our predictions below.

Firstly, following the potential asymmetry between subject and object relative clauses and previous findings with post-nominal RCs (Gordon et al., 2001; King & Just, 1991; Traxler et al., 2002), we predict ORCs will be processed more slowly than SRCs in both Thai and English. Secondly, regarding the primary location of difficulty, two alternative loci are considered. Based on the Locality-based model (Gibson, 1998; Grodner & Gibson, 2005), difficulty is predicted at the embedded verb in the ORC and should be more pronounced in English than in Thai due to greater verb

morphological complexity. In terms of NPs, the Locality-based model does not assign independent units to determiners. As Thai definiteness can be inferred from bare NPs, and as NPs are assigned the same weight (1 unit), embedded NPs are not expected to incur additional difficulty in either language. Alternatively, according to the Surprisal/Expectation model (Hale, 2001; Levy, 2008), the primary difficulty is predicted to involve the embedded subject in the ORC, which is unexpected following the complementizer.

We aimed first to determine whether the ORC is processed more slowly than the SRC across Thai and English using the self-paced reading paradigm. Regions three (the RC verb or noun), four (the RC noun or verb), and five (the matrix verb) were treated as critical regions, following prior studies (Staub, 2010; Traxler et al., 2001). Secondly, we addressed the unresolved issue of the primary locus of difficulty between the two clause types. Our goal was to acquire evidence that identifies this locus in Thai and compare it to English. Conducting such a cross-linguistic comparison within a single study is valuable, as it enables finer-grained analyses and contributes critical evidence to current theoretical debates.

## Method

### Norming

Firstly, we wanted to ensure that the animate NPs were equally plausible to serve as Agent/Experiencer and Patient/Theme. Secondly, we wanted to confirm that subject relatives occurred more frequently than object relatives in both English and Thai.

We initially constructed eleven English experimental sentence pairs containing SRCs and ORCs, based on Traxler et al.'s (2002) Experiment 1 (e.g., *the nurse that liked the doctor wore a floral skirt* and *the nurse that the doctor liked wore a floral skirt*). To check for plausibility, we extracted the sentential subject and the relative clause verb and object to form a new sentence (e.g., *the nurse liked the doctor*) and created an alternation in which the subject became the object and vice versa (i.e., *the doctor liked the nurse*). We refer to these as sentence conditions 1 and 2, respectively. We chose the top 8 pairs (8 ORCs and 8 SRCs), which were ranked in descending p values by the Wilcoxon signed-rank test. These eight pairs were also translated into Thai. Twenty-four native English speakers, and twenty-nine native Thai speakers rated these items for plausibility on a 1–5 Likert scale (1 = *not at all plausible*, 5 = *extremely plausible*). Data were analyzed using repeated measures ANOVAs, with language group (English vs. Thai) as the between-subjects variable, and the plausibility ratings of sentence conditions 1 and 2 as the dependent variable. The interaction between sentence condition and language group was not significant in either

the by-subject or by-item analyses ( $F_1(1, 51) = .177, p = .733$ ;  $F_2(1, 14) = .029, p = .866$ ), which indicates that neither of the sentence conditions differed in terms of plausibility across both English and Thai. Thus, we deemed the NPs in these sentences as appropriate target stimuli for the experimental trials.

An additional twenty-five English speakers and twenty-six native Thai speakers were asked to produce complete sentences from sentence fragments. NPs with the complementizer *that* or *thîi* acted as prompts for completion. The results indicated that SRC and ORC sentences accounted for 89% and 11% in English, and 81% and 10% in Thai. The ratios of SRC to ORC (8/9 to 1) were similar for both English and Thai speakers, as expected.

## Participants

Thirty-six native Thai-speaking participants (26 females) ranging in age between 18-21 (aged 19 on average) participated. The native Thai-speaking (NTS) participants were first and second-year undergraduate students at a public university in Bangkok. An additional native English-speaking group consisting of 21 persons (2 females) participated. Fourteen resided in Thailand and seven were visitors to Japan. The native English-speaking (NES) participants in Thailand were invited by the first author, and those in Japan by the fourth author in person or by email. The NES participants in Thailand ranged in age between 20-60 (aged 46 on average). Those in Japan ranged in age between 19-31 (aged 24 on average). All participants in the study had normal or corrected-to-normal vision. They gave informed consent and received a fee for taking part.

## Design and Materials

The experiment employed a 2 x 2 x 3 mixed factorial design, with clause type (SRC vs. ORC) and critical region (3 vs. 4 vs. 5) as the within-subjects factors, and language group as the between-subjects factor (NTS vs. NES).

The plausibility study yielded normed English and Thai NPs and verbs from the eight trials. This enabled us to select the corresponding target sentences for use in the SPR experiment. To cancel out effects from presenting only one version of the sentential subject of a pair such as (1a) and (1b), we created an additional set of eight pairs in English (and translated them into Thai) to allow for the NPs and the verb in the RCs in the earlier set to serve as the sentential subjects as well, with different matrix predicates (e.g., *the writer/ that/ hated/ the actor/ had/ a car accident/ last week* and *the writer/ that/ the actor/ hated/ had/ a car accident/ last week.*) This yielded a total of sixteen target sentence pairs (16 SRC & 16 ORC sentences). Since English and Thai

differ in the presence and absence of determiners, phrase-by-phrase segmentation was employed for consistency in alignment between the language segments/regions. As appropriate, each author ensured that there was meaning correspondence and the segmentations were done in parallel fashion across English and Thai.

Sentences (5) and (6) illustrate English and Thai SRC and ORC sentence pairs with region numbers appearing in subscript. In Thai orthography, words appear adjacent to one another. For naturalness, additional (region seven) tokens were created for selected Thai (items 1, 10, 11, 12, 14, 15, and 16) and English (items 1, 11, 12, 14, and 16) sentence pairs; these were not included in the analysis (see Appendices A and B).

For all stimuli, critical regions for analysis were regions three, four, and five, corresponding to the embedded verb/subject, object/verb, and matrix verb (see Traxler et al., 2002). The mean word, syllable, and character lengths of the target stimuli, excluding region seven, are reported in Table 1.

#### (5) English and Thai SRCs and comprehension questions

- a. The writer<sub>1</sub> that<sub>2</sub> hated<sub>3</sub> the actor<sub>4</sub> had<sub>5</sub>  
a car accident<sub>6</sub> last week<sub>7</sub>
- b. นักเขียน<sub>1</sub> ที่<sub>2</sub>เกลียด<sub>3</sub> นักแสดง<sub>4</sub> ได้รับ<sub>5</sub>  
อุบัติเหตุ<sub>6</sub> อาทิตย์ที่แล้ว<sub>7</sub>  
*nákkhǎan<sub>1</sub> thǐi<sub>2</sub> kliat<sub>3</sub> náksadeɛŋ<sub>4</sub> dâjráp<sub>5</sub>*  
writer COMP hate actor get  
*ʔùbàttihèt<sub>6</sub> ʔathít.thǐiléew<sub>7</sub>*  
accident week.past
- c. Did the writer have a car accident? นักเขียนได้รับอุบัติเหตุไหม

#### (6) English and Thai ORCs and comprehension questions

- a. The writer<sub>1</sub> that<sub>2</sub> the actor<sub>3</sub> hated<sub>4</sub> had<sub>5</sub>  
a car accident<sub>6</sub> last week<sub>7</sub>
- b. นักเขียน<sub>1</sub> ที่<sub>2</sub>นักแสดง<sub>3</sub>เกลียด<sub>4</sub> ได้รับ<sub>5</sub>  
อุบัติเหตุ<sub>6</sub> อาทิตย์ที่แล้ว<sub>7</sub>  
*nákkhǎan<sub>1</sub> thǐi<sub>2</sub> náksadeɛŋ<sub>3</sub> kliat<sub>4</sub> dâjráp<sub>5</sub>*  
writer COMP actor hate get  
*ʔùbàttihèt<sub>6</sub> ʔathít.thǐiléew<sub>7</sub>*  
accident week.past
- c. Did the actor have a car accident? นักแสดงได้รับอุบัติเหตุไหม

**Table 1**

*Mean Number of Words and Word Length per Region in Syllables and Characters for English and Thai ORC Stimuli*

	REG1	REG2	REG 3	REG 4	REG 5	REG 6	Total
<i>Words</i>							
English	2.06	1	2.06	1	1	2.5	9.62
Thai	1.13	1	1.13	1.13	1.06	2.13	7.58
<i>Syllables</i>							
English	3.19	1	3.19	1.63	1.44	3.81	14.26
Thai	2.81	1	2.81	1.75	1.50	3.81	13.68
<i>Characters</i>							
English	10.63	4	10.63	5.63	5.69	12.31	48.88
Thai	7.75	3	7.75	5.63	4.94	11.81	40.88

Note—English NPs were significantly longer in character length than the corresponding Thai NPs for Regions 1 and 3 ( $F(1, 31) = 10.737, p < .05$ ; Regions 4-6 did not differ significantly in character length between English and Thai (all  $F$ s  $< 1$ ; all  $p$ s  $> .05$ ). The number of characters per region in the SRC trials matched those in the ORC trials; the only difference was a switch between regions three and four.

The number of characters in Thai included the consonants, vowels, and tone mark symbols; those in English included spaces. As Thai lacks determiners and uses a non-segmental orthography, mean word and character lengths were relatively shorter for Thai compared to their English counterparts. For the matrix verb region, the Thai stimuli included three bimorphemic serial verbs: *dāj.ráp* (get.receive), *paj.jām* (go.visit) and *paj.phóp* (go.meet) meaning ‘receive’, ‘visit’ and ‘meet’; the remaining verbs were monomorphemic. In English, monomorphemic verbs inflected for past tense were employed. The combination of serial verbs in Thai and past tense morphemes in English likely resulted in the average number of characters being equivalent across languages for this region.

## ***Procedure***

The 16 pairs of target stimuli (32 items in total) were allocated to two lists (List 1, List 2) using a Latin-Square design format. Each list consisted of 8 SRC and 8 ORC stimuli. The participants were presented with both lists (e.g., 1a and 2b in Appendices A and B), and list order was counterbalanced. We also included 48 filler sentences. The total set of 64 stimuli (16 targets, 48 fillers) was presented randomly using *Linger* on a laptop computer, in a non-cumulative moving window fashion. Four practice sentences were included prior to the test phase, and participants were able to take a short break after completing approximately 50 trials.

Two laptop computers were set up in an empty classroom or a study room, and two participants performed the SPR task simultaneously. Participants were instructed to read sentences one segment at a time by the experimenter or research assistant. Participants were also informed that they should complete the task quickly but accurately. After reading the sentence, they answered a comprehension question, as shown in (4c) and (5c), with the *F* key indicating a Yes response and the *J* key indicating a No response. These comprehension questions ensured that participants paid attention to the meaning of each sentence. When participants answered correctly, they received positive feedback, and when they answered incorrectly, they were informed that the answer was incorrect.

## **Results**

For the NTS group, data from the trials with incorrect answers and those exceeding 3,000 ms were removed, which accounted for 7.09% of the total data. We considered 3,000 ms an extreme value that might affect data analysis, observing behavioral science standards (e.g., Miller's (2024) median outlier latency). Overall, 92.91% of the NTS data remained for statistical analyses.

With respect to the NES group, the data from incorrect answers and those exceeding 4,000 ms were excluded, accounting for 7.64% of the total data. The 4,000 ms absolute outlier value was determined, following the same standards. Overall, 92.36% of the NES data, comparable to NTS group, remained for statistical analyses.

Mean accuracy rates for the NTS group were 95% and 93% for the SRC and ORC conditions, respectively, and for the NES group, 96% and 90%. Reading time data were computed only for trials with correct responses to the comprehension questions. Reading times were averaged across six

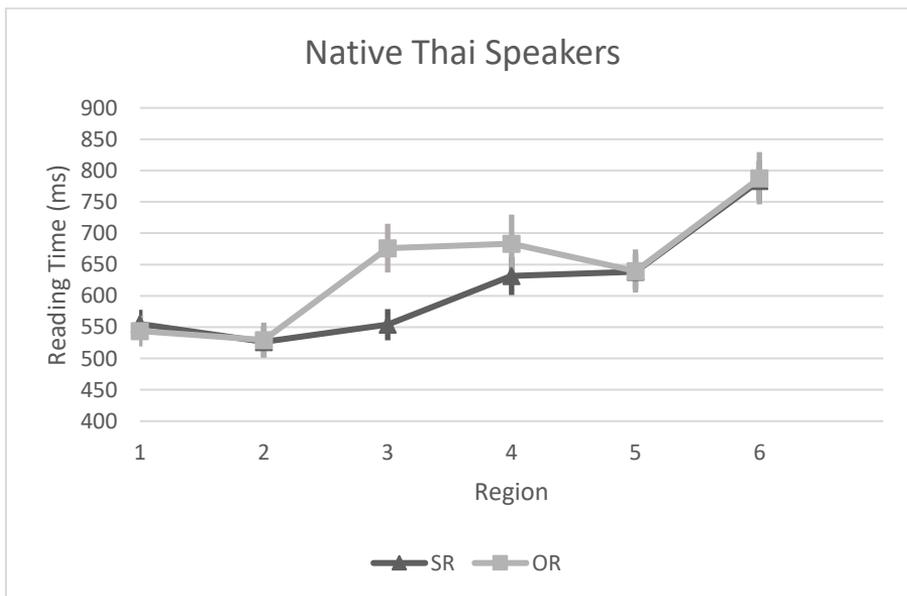
regions ( $M_s = 657$  ms &  $756$  ms;  $SD_s = 464$  ms &  $576$  ms for the NTS & NES groups, respectively).

Figures 1 and 2 illustrate the mean reading times (by subject) for the individual regions of the NTS and NES groups. For the NTS group, there was remarkable slowing between the complementizer and the embedded subject in the ORC. Speed then remained quite stable till the clausal boundary (the embedded verb). In the SRC, there was steady slowing between the complementizer and the embedded verb; the slow trend continued to rise moderately when reaching the clausal boundary (the embedded object). After this point (the matrix verbs), the two structures were processed consistently quickly but parsing speed slowed again at region six for both the ORC and SRC, presumably due to phrase lengths and in preparation for the upcoming comprehension questions.

For the NES group (see Figure 2), there was moderate slowing in the ORC between the complementizer and the embedded subject, but processing speed increased markedly when reaching the embedded verb. In the SRC, slowing was stable between the complementizer and the embedded verb and the slow trend continued to rise moderately to the embedded object. Like in the Thai data, the matrix verbs were processed consistently more quickly across the ORC and SRC, with a noticeable increase in reading time at region six.

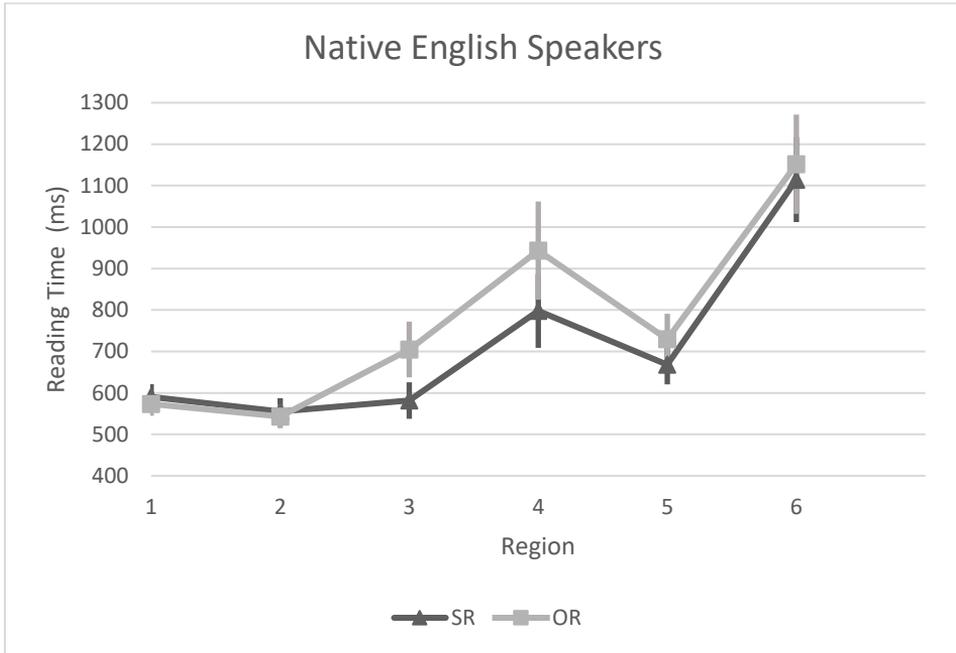
**Figure 1**

*Native Thai Speakers' Reading Times for SRCs and ORCs in the By-Subject Analysis*



**Figure 2**

*Native English Speakers' Reading Times for SRCs and ORCs in the By-Subject Analysis*



We initially checked if there were interactions among the combined regions three and four and the individual region five, clause type, and language group. Region three alone and region four alone had different lexical items. When they were combined (summed), the combinations had the same lexical items, but in reverse order (see Table 2). Since the combined regions three and four contained the same lexical items, they were analyzed together (cf. Gordon et al., 2001). In addition, region five was included because spill-over effects might occur at this position. In the upcoming statistical reports, the F-values of the by-subject and by-item analyses are represented by  $F_1$  and  $F_2$ , respectively.

**Table 2**

*Mean Differences between the Combined Regions 3 and 4 vs. Region 5 (Standard errors are in parentheses.)*

Analysis/Region	By subject	By item
Region 3&4	1445.363 <small>(56.259)</small>	1393.027 <small>(69.536)</small>
Region 5	685.151 <small>(21.541)</small>	668.976 <small>(26.928)</small>
Mean difference	760.232* <small>(45.774)</small>	724.051* <small>(51.217)</small>

Note. \*  $p < .05$

There were significant main effects for regions ( $F_1(1, 55) = 199.851, p < .001; F_2(1, 30) = 275.841, p < .001$ ), clause type ( $F_1(1, 55) = 21.869, p < .001; F_2(1, 30) = 9.002, p < .05$ ), and interactions between clause type and region ( $F_1(1, 55) = 16.606, p < .001; F_2(1, 30) = 9.678, p < .05$ ). The interaction effect between region and language group was significant in the by-item analysis ( $F_2(1, 30) = 6.588, p < .05$ ), but just approached significance in the by-subject analysis ( $F_1(1, 55) = 3.138, p = .082$ ). However, there were no interactions between clause type and language group ( $F_1(1, 55) = 2.075, p = .155; F_2(1, 30) = 1.334, p = .257$ ), nor were there significant three-way interactions between region, clause type, and language group ( $F_1(1, 55) = .135, p = .715; F_2(1, 30) = .317, p = .578$ ). These results suggest that the patterns of processing for the combined Regions 3 and 4 and the individual Region 5 across SRCs and ORCs were the same across the two languages. The main effects for regions suggest that the combined regions involved significantly longer reading times than Region 5 (see Figures 1 and 2). The difference between the combined and the matrix verb regions was clear, and consequently, there were no spill-over effects from the RCs to the matrix verbs.

To examine clause types within each language, repeated measures ANOVAs were conducted on the summed reading times for regions three and four, comparing ORC and SRC conditions. For the NTS group, ORCs were read more slowly than SRCs by both subject and item ( $F_1(1, 35) = 12.909, p < .05; F_2(1, 15) = 17.737, p < .05$ ), and the same pattern was observed for the NES group ( $F_1(1, 20) = 10.342, p < .05; F_2(1, 15) = 4.767, p < .05$ ). Thus, both Thai and English data confirm the prediction that ORCs are processed more slowly than SRCs.

To determine the primary locus of difficulty, a three-way ANOVA with word category (nouns vs. verbs), clause type (SRC vs. ORC), and

language group (NTS vs. NES) was performed. This analysis examined whether processing was affected by different word categories across regions (i.e., Region 3 was occupied by a noun in the ORC but a verb in the SRC), following the approach of Forster et al. (2009) and Staub (2010). The analysis revealed significant three-way interactions ( $F_1(1, 55) = 12.601, p < .05$ ;  $F_2(1, 30) = 12.468, p < .05$ ), indicating that the effect of clause type on reading times varied as a function of both word category and language group. There were also main effects of clause type ( $F_1(1, 55) = 24.292, p < .001$ ;  $F_2(1, 30) = 10.483, p < .05$ ) and interactions between word category and clause type ( $F_1(1, 55) = 26.800, p < .001$ ;  $F_2(1, 30) = 23.902, p < .001$ ). No other main effects or interactions involving word category or language were observed (all  $F_s < 2.1$ ; all  $p_s > .05$ ). These results indicate that NTS and NES participants differed in how they processed nouns and verbs across ORC and SRC conditions, prompting further analyses to pinpoint the primary locus of difficulty for each group.

Comparisons were made between the nouns in the ORC and SRC and for the verbs and the matrix verbs, by means of ANOVAs with two within-subjects factors (i.e., Region 3 of the ORC and Region 4 of the SRC). Within the NTS group, the reading times of verbs in the ORC were significantly longer than those in the SRC ( $F_1(1, 35) = 12.599, p < .05$ ;  $F_2(1, 15) = 19.199, p < .05$ ). However, the reading times of nouns in the ORC were not significantly longer than those of the SRC in the by-subject analysis ( $F_1(1, 35) = 2.796, p = .103$ ), but there was a marginally significant difference in the by-item analysis ( $F_2(1, 15) = 3.881, p = .074$ ). There were no significant differences at the matrix verb position (both  $F_s < 1$ ;  $p_s > .05$ ).

With respect to the NES group, the reading times of the verbs in the ORC were significantly longer than those in the SRC ( $F_1(1, 20) = 17.790, p < .001$ ;  $F_2(1, 15) = 14.140, p < .05$ ). However, the reading times of the nouns were not significantly different between the ORC and the SRC ( $F_1(1, 20) = 2.907, p = .104$ ;  $F_2(1, 15) = 1.480, p = .243$ ), nor were there any significant differences in reading times at the matrix verb position ( $F_1(1, 20) = 2.032; p = .169$ ;  $F_2(1, 15) = 2.714, p = .120$ ).

The verbs in the ORC involved longer reading times than their counterparts in the SRC in both groups. The nouns in the ORC and SRC were processed similarly in the NES group; however, the nouns in the ORC tended to incur longer reading times than their SRC counterparts in the NTS group, as seen in the by-item analysis. For the NES group, this result is compatible with the Locality-based Integration model; that is, the primary difficulty occurred at the verb in the ORC. With respect to the NTS group, there was a tendency for difficulty to appear at the noun initially and continue to the verb in the ORC, a trend more in line with the Surprisal/Expectation model.

## Discussion

We investigated the processing of Thai and English SRCs and ORCs using a self-paced reading paradigm. Our findings revealed that ORCs were processed more slowly in both Thai and English, as expected. The absence of any significant three-way interaction between the combined regions three and four, clause type, and language group, alongside the significant difference observed between the SRC and ORC conditions, suggests that the NTS group read ORCs more slowly than SRCs, with the same pattern obtained for the NES group. These results are consistent with previous findings (e.g., Ford, 1983; Gordon et al., 2001; Holmes & O'Regan, 1981; King & Just, 1991; Traxler et al., 2002).

A further aim was to determine the primary location of difficulty in the ORC, a consequence of which would illuminate what guides initial processing. Our finding that the embedded verb in the English ORC involved difficulty is similar to a number of previous studies (e.g., Ford, 1983; Gordon et al., 2001; Grodner & Gibson, 2005; Holmes & O'Regan, 1981). This result is also consistent with the Locality-based Integration model, but is inconsistent with recent findings, in which the primary difficulty involved the ORC subject (Forster et al., 2009; Gennari & MacDonald, 2008; Staub, 2010). Aside from Staub's eye-tracking experiments, it is likely that when a determiner is presented in isolation in a word-by-word presentation (e.g., Gennari & MacDonald, 2008), slowing emerges as it normally occurs with a noun, but when a determiner-noun pair is presented together, as in our procedure, the more naturalistic presentation might cause the parser to withhold the processing of the RC until the subsequent item, as the Locality-based Integration model predicts.

The Thai NPs were processed more slowly than the English NPs. The fact that Thai and English differ with respect to the absence and presence of determiners provides insight into how the parser manages bare NPs versus NPs with determiners. In the absence of determiners, the parser probably encounters an NP which, despite its default definite property, awaits further identification in the following constituent (i.e., exactly what the head noun refers to). This semantically unclear condition, as discussed with sentence (4), could delay the processing of the bare NPs in Thai, in addition to the cost associated with the unexpected NP predicted by the Surprisal/Expectation model. In English, the determiner presenter together with a noun may reduce the semantic ambiguity of the embedded subject and enable the parser to maintain identification until the end of the relative clause. While the statistical results for Thai indicate difficulty at the verb in the ORC, it is plausible that the difficulty started earlier (i.e., at the noun), as suggested by the by-item

findings. Overall, the Surprisal/Expectation model (Hale, 2001; Levy, 2008) appears more felicitous for the Thai data than the Locality-based Integration model (Gibson, 1998; Grodner & Gibson, 2005). If our assertion that the primary difficulty involves the noun for Thai ORCs is tenable, these results are consistent with findings in previous research (Forster et al., 2009; Gennari & MacDonald, 2008; Staub, 2010).

Viewed from the issue of initial sentence processing, it is plausible that the frequency of word category plays a role in facilitating or delaying processing, particularly for Thai, whereas resource-based accounts are able to explain the facilitation in processing for English. When the parser encounters a Thai NP or a determiner alone in the ORC (Gennari & MacDonald, 2008), a difficulty arises, resulting in a slowing of reading speed, possibly due to the lower frequency of the bare NP and determiner. The parser's plausible slow reading of nominal subjects without determiners in ORCs also suggests that language-specific properties should be taken into account in initial processing, in line with Xu et al.'s (2019) call for further investigations into language-specific mechanisms in sentence processing.

When the parser encounters an English NP in the ORC, which includes a determiner and a noun (the present study; Gordon et al., 2001; Grodner & Gibson 2005), the parser likely holds off processing until it reaches the upcoming verb, where slow reading speed or difficulty emerges, in line with resource-based accounts such as the Locality-based Integration model. Given that the current study and several previous studies (see Gordon et al., 2001; Grodner & Gibson, 2005) commonly report results from phrase-by-phrase segmentation, it is noteworthy that supporting evidence for the Locality-based Integration model for English may be limited to the methodology of phrasal presentation.

Another important note relates to Staub's (2010) findings, in which both the ORC subject and verb involved processing difficulty with observable behavioral differences. As discussed earlier, Staub (2010) suggests that these differences are associated with different theoretical models. Our results on Thai participants' processing of SRCs and ORCs indicate a plausible uniformity toward Surprisal/Expectation model, similar to those reported for English (e.g., Forster et al., 2009; Staub, 2010), despite the use of a self-paced reading paradigm in Thai.

Regarding the matrix verbs, Holmes and O'Regan (1981) found that regressive eye-movements occurred less often at the clause boundary than at the preceding phrase. In Ford's (1983) word-by-word experiments, significantly longer reading times persisted at the matrix verb and the determiner. In Grodner and Gibson's (2005) self-paced reading Experiment 3, the matrix verb was processed faster than the embedded verb, but there were no differences between the matrix verbs across SRCs and ORCs. Forster

et al. (2009) and Staub (2010) similarly found that the matrix verbs were processed equally quickly. With the exception of Ford's study, the current study as well as other previous studies (Forster et al., 2009; Grodner & Gibson, 2005; Holmes & O'Regan, 1981; Staub, 2010), suggest that the parser establishes an association between the matrix verb and the sentential subject relatively quickly.

### Limitations of the Study

At this institution, there were a small number of NES, most of whom taught English to Thai undergraduate students. Consequently, the average NES participant in Thailand was considerably older than the NTS counterpart. We do not consider these limitations as seriously affecting the interpretation of the results. Our NES results, which suggest difficulty on the ORC verbs, are in line with prior investigations (e.g., Gordon et al., 2001; Grodner & Gibson, 2005). Therefore, the NES participants were deemed eligible for inclusion.

In terms of stimuli, an NP with the adjective *young* (i.e., *the young woman*), occupied segment 3 of (11b), in Appendix A. This adjective was inserted to distinguish between *the young woman* and *the woman* in (14b). Since the average reading times of these tokens—505 ms in (11b) and 574 ms in (14b)—were relatively similar, the results were unlikely to be affected by the presence of *young*. There were four Thai trials—(4a), (4b), (14a) and (14b) in Appendix B, in which the embedded object and subject NPs contained common nouns, classifiers, and demonstratives. Common nouns, unlike career-related nouns in the other stimuli, typically convey generic readings (i.e., *phûuchay* “men” or *phûuyǐn* “women”). If presented alone (albeit briefly) in our phrase-by-phrase SPR experiments, the resulting reading would sound unnatural, as they would normally be associated with generic readings. In addition, the verbs *visited* and *admired* denote specific events; the alternative in Thai to avoid unnaturalness and remain consistent with the specific events denoted by the verbs was to add a demonstrative and the proper classifier to these common nouns. Therefore, we inserted *khon*, a proper classifier, and *nán*, a distal demonstrative, meaning ‘that man and that woman’, respectively.

### Conclusion

We investigated the processing of subject and object-extracted relative clauses (SRC and ORC) in native Thai and native English speakers, testing two main predictions. The first predicted an SRC advantage over the ORC, and the second concerned the primary locus of difficulty. The first prediction was confirmed. Regarding the primary location of difficulty, reading times for

the embedded verb in the ORC were significantly longer than for its SRC counterpart in both languages. Comparisons of the reading times for the nouns were non-significant in English, whereas Thai showed a marginally significant difference by item. The English data support the Locality-based Integration model, with primary difficulty at the embedded verb, whereas Thai data are better accounted for by the Surprisal/Expectation model, which predicts the embedded noun as the primary locus of difficulty. This cross-linguistic difference may reflect the absence of determiners in Thai. For matrix verbs, reading times were shorter than preceding regions and were similar across ORC and SRC conditions and groups. Regarding initial sentence processing, both word-category (noun or verb) frequency and resource-based factors appear influential. After the complementizer, the less frequently occurring noun and the more frequently occurring verb may account for RC processing in both Thai (the current study) and English (Forster et al., 2009; Staub, 2010), whereas resource-based models are evident mainly in phrase-by-phrase segmentation in English. Further research on embedded nouns and verbs in post-nominal RCs is recommended to clarify language-specific processing mechanisms and the role of word-category frequency.

### **Ethical Approval**

This research was approved by the Human Subject Research Ethics Committee (in Social Sciences) of Thammasat University (application number 045/2562) in Thailand and that of University of Tokyo (application number 629) in Japan.

### **Acknowledgements**

We thank Likkasit Dansakul for his kind assistance with setting up *Linger* and coding trials, Naruemon Khunwong for supplying filler sentences, and the late Petcharat (Aom) Thanasuptawee for programming the English and Thai stimuli on *Linger*. Aom's contribution to this study will remain in our minds. We are thankful to the two anonymous reviewers for their helpful and constructive comments. The first author, particularly, is grateful to her late father, Police Lieutenant General Samrit Singhapreecha, for enabling her to eventually complete this article. This project was partially supported by Japan Society for the Promotion of Science [Kakenhi Grant in Aid No. 20H01254], to which we are grateful.

### About the Authors

**Pornsiri Singhapreecha:** Professor of Linguistics and part-time faculty at the Language Institute of Thammasat University. Her areas of research are syntactic theory, second language acquisition (within the Generative model), and L1 and L2 processing.

**Benjamin Clarke:** Assistant Professor in the Department of Linguistics, Faculty of Liberal Arts, Thammasat University. He conducts psycholinguistic research on the cognitive mechanisms underlying comprehension, production, and acquisition in both L1 and L2 speakers.

**Naparat Meechanyakul:** Assistant Professor of Linguistics at the Language Institute of Thammasat University. Her areas of research are theoretical and experimental linguistics, syntactic theory and its interfaces, and second language acquisition.

**Yuki Hirose:** Professor in the Graduate School of Arts and Sciences at the University of Tokyo, Japan. Her research focuses on first and second language processing and acquisition, with recent work focusing on the role of prosody in predictive processing.

### References

- Baker, M. (1988). *Incorporation. A theory of grammatical function changing*. University of Chicago Press.
- Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(2), 203-218.
- Forster, I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41, 163–171.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5(4), 519-559.
- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161-187.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E., & Wu, H. H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, 28(1-2), 125-155.

- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1411-1423.
- Grodner, D. J., & Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29, 261-291.
- Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. *Proceedings of NAACL*, 2, 159-166.
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3-27.
- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 20, 417-430.
- King, J., & Just, M.A. (1991). Individual differences in syntactic parsing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.
- Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: evidence from ERP data. *Language*, 89(3), 537-585.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375-419.
- MacWhinney, B. (2005). The emergence of grammar from perspective. In D. Pecher & R. A. Swaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 198-223). Cambridge University Press.
- Miller, J. (2024). Estimating the proportions and latencies of reaction time outliers: a pooling method and case study of lexical decision tasks. *Behavioral Research Methods*, 56(7), 7280-7306. doi:10.3758/s13428-024-02419-y.
- Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding and M. Tsujimura (Eds.), *WCCFL 22 Proceedings* (pp. 342-355). Cascadia Press.
- Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 455-503). Academic Press.
- Poole, G. (2002). *Syntactic theory*. Palgrave.

- Singhapreecha, P., & Sybesma, R. (2015). An ellipsis analysis of split nominals in Thai. *Cahiers de Linguistique Asie Orientale*, 44(1), 57-88.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71-86.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.
- Xu, K., Duann, J., Hung, D.L., & Wu, D.H. (2019). Preference for object relative clauses in Chinese sentence comprehension: evidence from online self-paced reading time. *Frontiers in Psychology*, 10(2210). doi: 10.3389/fpsyg.2019.02210.

## Appendix A

### English Stimuli in Segments

- 1a. The writer/ that/ hated/ the actor/ had/ a car accident/ last week.
- 1b. The writer/ that/ the actor/ hated/ had/ a car accident/ last week.
- 2a. The actor/ that/ hated/ the writer/ complained/ about the film script.
- 2b. The actor/ that/ the writer/ hated/ complained/ about the film script.
- 3a. The man/ that/ visited/ the director/ married/ a famous actress.
- 3b. The man/ that/ the director/ visited/ married/ a famous actress.
- 4a. The director/ that/ visited/ the man/ made/ great films.
- 4b. The director/ that/ the man/ visited/ made/ great films.
- 5a. The nurse/ that/ liked/ the doctor/ wore/ a floral skirt.
- 5b. The nurse/ that/ the doctor/ liked/ wore/ a floral skirt.
- 6a. The doctor/ that/ liked/ the nurse/ moved/ to another hospital.
- 6b. The doctor/ that/ the nurse/ liked/ moved/ to another hospital.
- 7a. The salesman/ that/ met/ the lawyer/ carried/ a big bag.
- 7b. The salesman/ that/ the lawyer/ met/ carried/ a big bag.
- 8a. The lawyer/ that/ met/ the salesman/ specialized/ in taxes.
- 8b. The lawyer/ that/ the salesman/ met/ specialized/ in taxes.
- 9a. The policeman/ that/ hurt/ the robber/ won/ the lotto.
- 9b. The policeman/ that/ the robber/ hurt/ won/ the lotto.
- 10a. The robber/ that/ hurt/ the policeman/ fled/ to the mountains.
- 10b. The robber/ that/ the policeman/ hurt/ fled/ to the mountains.
- 11a. The musician/ that/ rescued/ the young woman/ gave/ an interview/ in public.
- 11b. The musician/ that/ the young woman/ rescued/ gave/ an interview/ in public.
- 12a. The young woman/ that/ rescued/ the musician/ posted/ the story/ on Facebook.

12b. The young woman/ that/ the musician/ rescued/ posted/ the story/ on Facebook.

13a. The woman/ that/ admired/ the painter/ worked/ for the UN.

13b. The woman/ that/ the painter/ admired/ worked/ for the UN.

14a. The painter/ that/ admired/ the woman/ had/ a new exhibition/ in town.

14b. The painter/ that/ the woman/ admired/ had/ a new exhibition/ in town.

15a. The driver/ that/ killed/ the manager/ possessed/ cocaine.

15b. The driver/ that/ the manager/ killed/ possessed/ cocaine.

16a. The manager/ that/ killed/ the driver/ owned/ a condo/ in Pattaya.

16b. The manager/ that/ the driver/ killed/ owned/ a condo/ in Pattaya.

## Appendix B

### Thai Stimuli in Segments

1a. นักเขียน/ที่/เกลียด/นักแสดง/ได้รับ/อุบัติเหตุ/อาทิตยที่แล้ว

1b. นักเขียน/ที่/นักแสดง/เกลียด/ได้รับ/อุบัติเหตุ/อาทิตยที่แล้ว

2a. นักแสดง/ที่/เกลียด/นักเขียน/ตำหนิ/เรื่องบทธละคร

2b. นักแสดง/ที่/นักเขียน/เกลียด/ตำหนิ/เรื่องบทธละคร

3a. ผู้ชายคนนั้น/ที่/ไปเยี่ยม/ผู้กำกับหนัง/แต่งงาน/กับนักแสดงหญิงชื่อดัง

3b. ผู้ชายคนนั้น/ที่/ผู้กำกับหนัง/ไปเยี่ยม/แต่งงาน/กับนักแสดงหญิงชื่อดัง

4a. ผู้กำกับหนัง/ที่/ไปเยี่ยม/ผู้ชายคนนั้น/สร้าง/หนังดังหลายเรื่อง

4b. ผู้กำกับหนัง/ที่/ผู้ชายคนนั้น/ไปเยี่ยม/สร้าง/หนังดังหลายเรื่อง

5a. พยาบาล/ที่/ชอบ/หมอ/ใส่/กระโปรงลายดอก

5b. พยาบาล/ที่/หมอ/ชอบ/ใส่/กระโปรงลายดอก

6a. หมอ/ที่/ชอบ/พยาบาล/ย้าย/ไปโรงพยาบาลอื่น

6b. หมอ/ที่/ พยาบาล/ชอบ/ย้าย/ไปโรงพยาบาลอื่น

7a. พนักงานขาย/ที่/ไปพบ/ทนายความ/หิว/กระเป๋ไปใหญ่

7b. พนักงานขาย/ที่/ทนายความ/ไปพบ/หิว/กระเป๋ไปใหญ่

8a. ทนายความ/ที่/ไปพบ/พนักงานขาย/เชี่ยวชาญ/เรื่องภาษีอากร

8b. ทนายความ/ที่/พนักงานขาย/ไปพบ/เชี่ยวชาญ/เรื่องภาษีอากร

9a. ตำรวจ/ที่/ทำร้าย/โจร/ถูก/ลือตเตอร์

9b. ตำรวจ/ที่/โจร/ทำร้าย/ถูก/ลือตเตอร์

10a. โจร/ที่/ทำร้าย/ตำรวจ/หนี/เข้าไปในป่า

10b. โจร/ที่/ตำรวจ/ทำร้าย/หนี/เข้าไปในป่า

11a. นักดนตรี/ที่/ช่วยชีวิต/หญิงสาว/ให้/สัมภาษณ์/ทางทีวี

11b. นักดนตรี/ที่/หญิงสาว/ช่วยชีวิต/ให้/สัมภาษณ์/ทางทีวี

- 12a. หญิงสาวที่/ช่วยชีวิต/นักดนตรี/โพสต์/เรื่องราว/ในเฟซบุค
- 12b. หญิงสาวที่/นักดนตรี/ช่วยชีวิต/โพสต์/เรื่องราว/ในเฟซบุค
- 13a. ผู้หญิงคนนั้นที่/ชื่นชม/จิตรกร/ทำงาน/ให้สหประชาชาติ
- 13b. ผู้หญิงคนนั้นที่/จิตรกร/ชื่นชม/ทำงาน/ให้สหประชาชาติ
- 14a. จิตรกรที่/ชื่นชม/ผู้หญิงคนนั้น/แสดง/ภาพเขียนใหม่/ในงานนิทรรศการ
- 14b. จิตรกรที่/ผู้หญิงคนนั้น/ชื่นชม/แสดง/ภาพเขียนใหม่/ในงานนิทรรศการ
- 15a. คนขับรถที่/ฆ่า/ผู้จัดการ/มี/โคเคน/ในครอบครอง
- 15b. คนขับรถที่/ ผู้จัดการ/ ฆ่า/ มี/โคเคน/ในครอบครอง
- 16a. ผู้จัดการที่/ฆ่า/คนขับรถ/เป็นเจ้าของ/คอนโด/ที่พัทยา
- 16b. ผู้จัดการที่/คนขับรถ/ฆ่า/เป็นเจ้าของ/คอนโด/ที่พัทยา