

## Appendix S1 of “Agreement and Reflexives in Non-Native Sentence Processing”

### Supplementary Analysis of L2 Proficiency

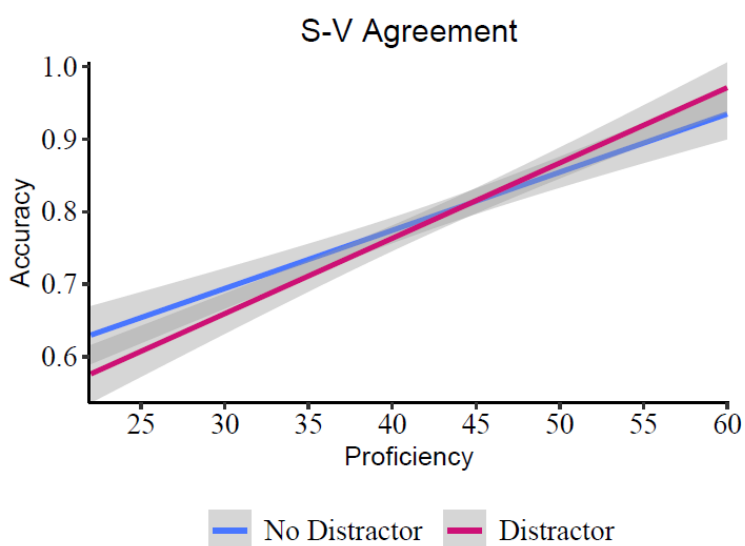
We report here the results of the additional analyses conducted on the L2 group with proficiency as a continuous variable that were conducted for each task in the two studies.

#### Study 1

##### *Grammaticality judgements*

##### *S-V agreement*

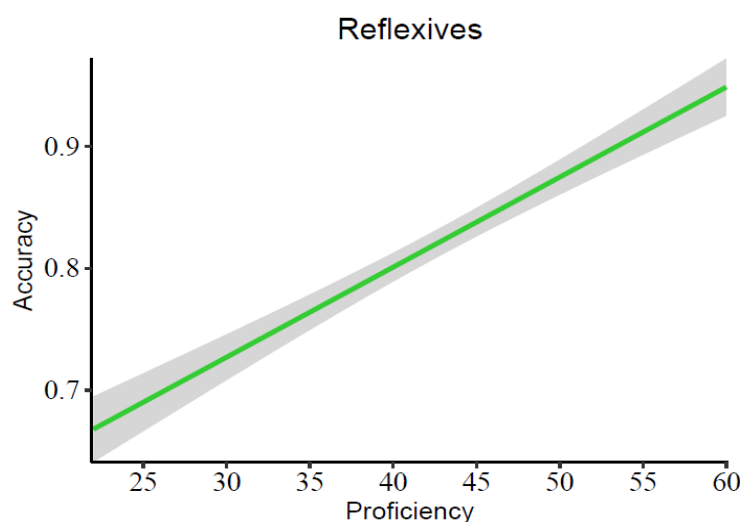
For proficiency effects on L2ers’ judgements, there was a main effect of proficiency (estimate = 0.07 (0.008),  $z = 8.08$ ,  $p < .001$ ) revealing that increasing proficiency induces increasing accuracy rates. A significant interaction between proficiency and distractor was also found (estimate = -0.022 (0.010),  $z = -2.07$ ,  $p = .037$ ). Figure S1 shows that higher-proficiency L2ers are more accurate in distractor than no distractor conditions, while lower-proficiency L2ers show the opposite trend. Note that, as can be seen in Figure S1, this interaction effect is small, especially when compared to the clear main effect of proficiency.



**Figure S1.** Effects of Proficiency and Distractor on L2 Speakers’ Accuracy Rates for S-V Agreement in Study 1.

### *Reflexives*

Proficiency analysis revealed only a significant main effect of proficiency (estimate = 0.061 (0.007),  $z = 8.06$ ,  $p < .001$ ) indicating that accuracy increases as proficiency increases, as shown in Figure S2, with no other significant interactions.

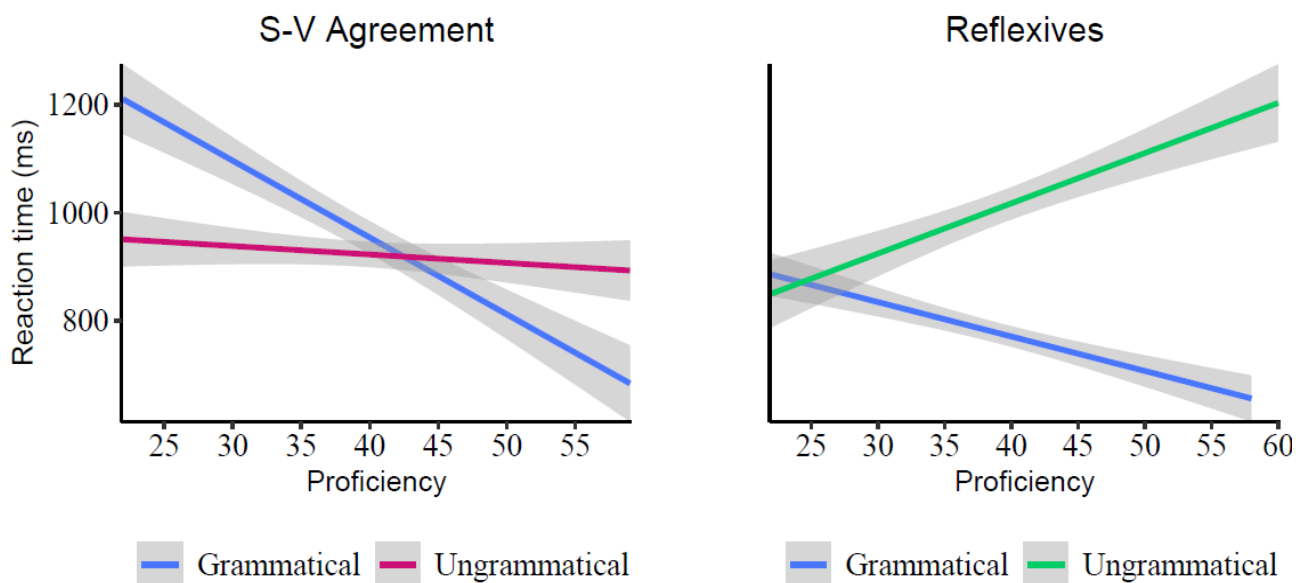


**Figure S2.** Effects of Proficiency on L2 Speakers' Accuracy Rates for Reflexives in Study 1.

### ***Self-paced reading***

#### *S-V agreement*

For proficiency effects on L2ers' reading times, we found a main effect of proficiency (estimate = -0.0006 (0.0002),  $t = -2.11$ ,  $p = .035$ ) showing that reading times were shorter as proficiency increases. There was also a significant interaction between grammaticality and proficiency (estimate < 0.001 (0.0001),  $t = 2.91$ ,  $p = .007$ ). Figure S3 (which appears as Figure 3 in the main paper) shows that higher-proficiency L2ers had longer reading times in ungrammatical conditions compared to grammatical conditions while lower proficient L2ers behaved differently.



**Figure S3.** Effects of Proficiency and Grammaticality on L2 Speakers' Reading Times in Study

1.

#### *Reflexives*

The main effect of proficiency was not significant (estimate < 0.001 (0.002),  $t = 0.127$ ,  $p = .899$ ); however, there was a significant interaction between grammaticality and proficiency (estimate = 0.006 (0.001),  $t = 4.59$ ,  $p < .001$ ). As illustrated in Figure S3, the effect of grammaticality gets larger as proficiency increases.

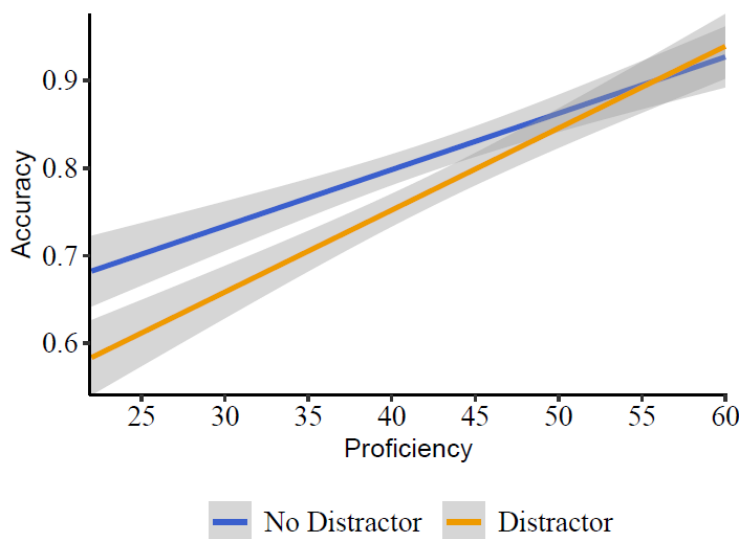
## **Study 2**

### ***Grammaticality judgements***

#### *Reflexives*

L2 proficiency had a significant main effect on L2 speakers' judgements (estimate = 0.065 (0.010),  $z = 6.50$ ,  $p < .001$ ), such that accuracy rates increased as L2 proficiency increases, as illustrated in Figure S4. There was also a significant proficiency by distractor interaction

(estimate = -0.024 (0.011),  $z = -2.07$ ,  $p = .038$ ). As shown in Figure S4, lower proficiency L2ers were more accurate in the no distractor than distractor conditions, while this difference between conditions gets smaller as proficiency increases.

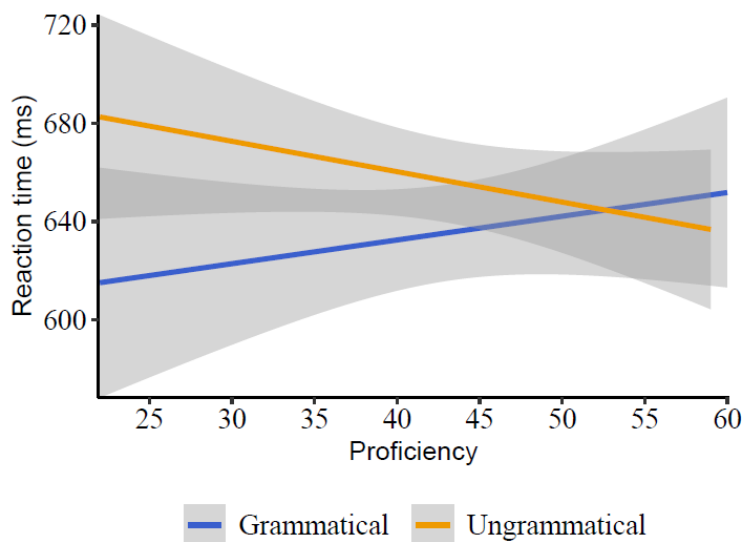


**Figure S4.** Effects of Proficiency and Distractor on L2 Speakers' Accuracy Rates in Study 2.

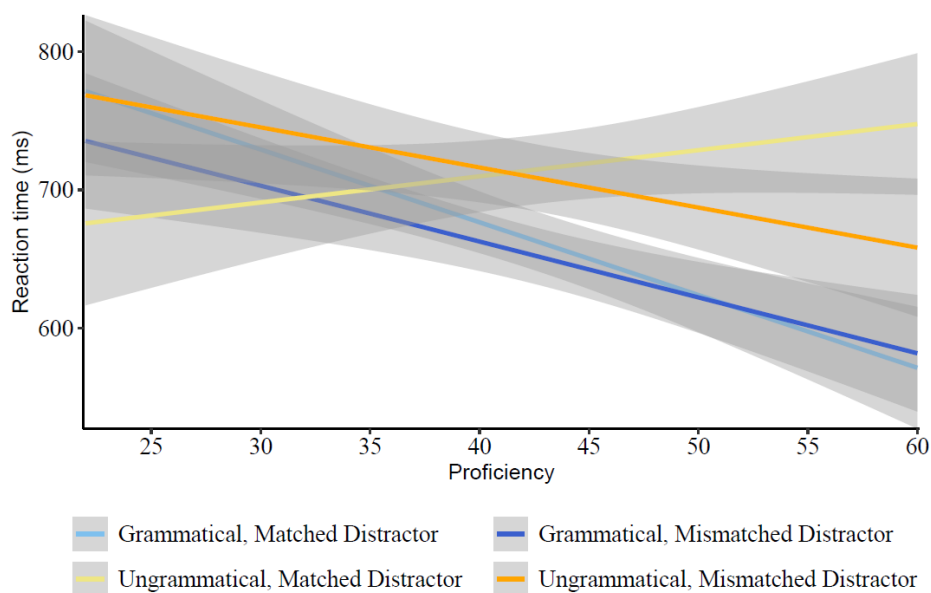
### *Self-paced reading*

#### *Reflexives*

For baseline conditions, the proficiency analysis revealed a significant interaction between grammaticality and proficiency (estimate = 0.004 (0.001),  $t = 3.15$ ,  $p = .002$ ), with lower-proficiency L2ers showing larger grammaticality effects compared to higher-proficiency L2ers as shown in Figure S5.



**Figure S5.** Interaction Effect between Proficiency and Grammaticality on L2 Speakers' Reading Times for Baseline Conditions in Study 2.



**Figure S6.** Interaction Effect Between Proficiency, Grammaticality and Distractor on L2 Speakers' Reading Times for the Main Set of Experimental Items in Study 2.

For the main set of experimental items, analysis of proficiency showed a significant three-way interaction between grammaticality, distractor and proficiency (estimate = -0.004

(0.002),  $t = -2.16$ ,  $p = .041$ ). As illustrated in Figure S6, reading times for grammatical conditions generally decreased as proficiency increased. Participants with lower proficiency also seem to be more influenced by interference from matching distractors, indicative of inhibitory interference, than higher proficiency participants in these conditions. For ungrammatical conditions, lower proficiency learners had shorter reading times when the distractor matched compared to when it mismatched, indicative of facilitatory interference, while higher proficiency learners showed the opposite effect.

## **Discussion**

Based on these results, L2 proficiency appears to have influenced both grammatical judgements and reading times across the two studies. For grammaticality judgements, L2ers generally performed accurately indicating they have the grammatical competence for the linguistic phenomena under investigation. In the main analysis, L2ers were significantly less accurate overall than L1ers in Study 1, but L2ers outperformed L1ers in Study 2. Figure S1 clearly demonstrates that the group effects in the main analysis of Study 1 are influenced by proficiency, with higher proficiency L2ers having higher overall accuracy.

L2ers' reading times also showed their real-time application of syntactic constraints manifested by longer reading times in ungrammatical compared to grammatical conditions. In the main analysis of Study 1 however, we observed less sensitivity to grammaticality effects to S-V agreement for L2ers than L1ers, while for reflexives L2ers exhibited larger grammaticality effects than the L1ers. This relative difference can be explained by L2ers' processing performance across different proficiency levels. As clearly illustrated in Figure S3, for reflexives almost all L2ers with differing proficiency levels showed grammaticality effects in the target direction, with longer reading times in ungrammatical conditions than grammatical

ones. The size of this effect however increased with higher L2 proficiency. However, only higher proficiency L2ers showed a target-like grammaticality sensitivity in agreement. We suggest that these individual differences in L2ers can explain the significant interactions observed between group and grammaticality in the main analysis of Study 1.

Though we did not find significant differences between L1ers and L2ers in Study 2 in the main analysis, our additional analysis here may suggest that grammaticality sensitivity can be modulated by proficiency. Susceptibility to interference from matching distractors also appears to be influenced by proficiency, as lower proficiency L2ers showed a tendency towards facilitatory interference in ungrammatical conditions and inhibitory interference in grammatical conditions, while higher proficiency L2ers did not. This may indicate that lower proficiency L2ers rely more heavily on non-structural cues, such as number, potentially prioritising them over structural cues, leading to increased interference. In contrast, higher proficiency L2ers seem to prioritise structural cues more consistently, like L1ers, leading to greater resilience against interference effects during language processing. However, given we did not observe significant L1/L2 differences in the main analysis, and given that a similar pattern was not found in Study 1, we do not intend to draw strong conclusions here about individual differences in interference effects among L2ers in Study 2.

Finally, we note that in individual differences research, strong conclusions can only be made if the measures used systematically measure individual variation (for discussion, see Hedge et al. (2018). Whether psycholinguistic tasks systematically measure individual variation is currently unclear, though existing research has shown that some existing tasks are not good measures of individual differences (Cunnings & Fujita, 2021; James et al., 2018). To assess this issue in our sample, we calculated the split-half reliability of all effects across our two studies for L2ers. Though it should not be considered a strict cutoff, split-half reliabilities above .7 are typically considered sufficient (Nunnally, 1978).

For grammaticality judgements, split-half reliability for overall accuracy rates were just below .7, ranging from .617 - .697 for the three experiments, suggesting our tasks were reasonable measures in terms of assessing individual differences in overall L2 accuracy on the judgement tasks. Note we also observed interactions between proficiency and distractor in two of our analyses reported above. The distractor effect however had poor split-half reliability (ranging from .021 - .165 across experiments) and so we are cautious in drawing strong conclusions about these interactions.

For self-paced reading, split-half reliabilities for overall reading speed were generally good (ranging from .691 - .866), suggesting our tasks were able to assess individual differences in overall L2 reading speed. Split-half reliabilities for all other experimental effects were low ( $< .274$ ), and as such, while we did observe some L2 proficiency effects, especially in terms of the size of grammaticality effects in Study 1, we are cautious in our conclusions here until further replication. In general, these split-half reliabilities replicate previous findings suggesting that while psycholinguistic tasks provide good measures of individual differences in overall task accuracy and reading speed, they provide poorer measures of individual differences in experimental effects (Cunnings & Fujita, 2021; James et al., 2018).



## **References**

- Cunnings, C. & Fujita, H. (2021). Quantifying individual differences in native and non-native sentence processing. *Applied Psycholinguistics*, 42, 579-599. Retrieved from <https://doi.org/10.1017/S0142716420000648>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50, 116-1186. Retrieved from <https://doi.org/10.3758/s13428-017-0935-1>
- James, A., Fraundorf, S., Lee, E., & Watson, D. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, 102, 155-181. Retrieved from <https://doi.org/10.1016/j.jml.2018.05.006>