Helpful Regardless of Acceptability: Resumption Eases Comprehension of Difficult-to-Process Relative Clauses in the L2 English of L1-Korean and L1-Mandarin Speakers

Fred Zenker and Bonnie D. Schwartz

Language knowledge and language processing interact in complex and sometimes unexpected ways, particularly in a second language (L2). This study investigates the interplay between the representation L2 learners (L2ers) have of the language they are acquiring, on the one hand, and how they process sentences in that language, on the other, by examining a type of relative clause (RC) that is a conspicuous trait of non-native English, namely the resumptive RC, as in (1).

(1) * the man that the detectives arrested him

The fascinating thing about resumptive RCs is that L2ers produce them even when they are disallowed in both the first language (L1) and the target language (TL). Such phenomena, which are traceable to neither L1 transfer nor TL input, allow researchers to gain new insights into properties of the language system and its development in the individual. Drawing on methods from psycholinguistic research on mature native speakers, we test whether resumptive RCs in L2 English constitute licit interlanguage representations and/or a subconscious strategy for managing cognitive load. Participants include L1-English controls and L2ers of English whose L1 is either Korean or Mandarin. Results indicate that despite low acceptability, English RPs can still ease comprehension under processing strain.

1. Filler-gap vs. resumptive relative clauses

In English, RCs are typically of the filler-gap type, with the nominal at the dependency's head being associated with a phonetically empty position (the gap) at the dependency's foot, as in (2a). Resumptive RCs resemble their filler-gap

* Fred Zenker, University of Hawai'i at Mānoa, fzenker@hawaii.edu; Bonnie D. Schwartz, University of Hawai'i at Mānoa, bds@hawaii.edu. Our sincere thanks to: Yu-Tzu Chang, Yunchuan Chen, Ann Tai Choe, In Ji Chun, Kamil Deen, Shin Fukuda, Theres Grüter, Yao Huang, Haerim Hwang, Li 'Julie' Jiang, Hyeyoung Jung, Hyunwoo Kim, Jieun Kim, Jihyun Kim, Kitaek Kim, Eunsun Lee, Wenyi Ling, William O'Grady, Akari Ohba, Amy J. Schafer, Rex A. Sprouse, Jue Wang, Hong Ri Zenker, Jing 'Crystal' Zhong, the Language Analysis and Experimentation Labs, the Language Acquisition Research Group, an NSF Doctoral Dissertation Research Improvement Grant, a Paula Menyuk Travel Award from Boston University, and, of course, all our participants.

 Proceedings of the 47th annual Boston University Conference on Language Development, ed. Paris Gappmayr and Jackson Kellogg, 797-810. Somerville, MA: Cascadilla Press.

counterparts except that a resumptive pronoun (RP) or other nominal occupies the foot position, as in (2b). Such resumptive RCs are regularly produced by L2ers.

(2) a. the only class$_i$ that I'm thinking about dropping __$_i$
b. * the only class$_i$ that I'm thinking about dropping it$_i$

Languages differ in regard to the RC environments in which resumption is permitted. Table 1 shows the distribution of gaps, RPs, and unrelativizable positions in single-clause RCs as reported by Keenan and Comrie (1977) for the languages in our study: English, Korean, and Mandarin. RPs are highly restricted in English and Korean; Mandarin allows them in a variety of environments. This study targets the direct object position, for which English and Korean both disallow RPs but Mandarin optionally allows them. There is ongoing debate about the exact distribution of RPs in Korean (see Han, 2013; Kwon, 2008; Song, 2003) and Mandarin (see Hitz & Francis, 2016; Pan, 2016); we note that our language consultants' judgments align very well with Keenan and Comrie's observations.

Table 1. Distribution of gaps (−), RPs (+), and unrelativizable positions (0) in single-clause RCs (adapted from Keenan & Comrie, 1977, p. 93, Table 2)

Language	SU	DO	IO	OBL	GEN	OCOMP
English	−	−	−	−	−	−
Korean	−	−	−	−	+	0
Mandarin	−	−/+	+	+	+	+

Note. SU = subject, DO = direct object, IO = indirect object, OBL = oblique, GEN = genitive, OCOMP = object of comparison.

2. Previous empirical research on L2 resumption in RCs

Previous empirical research on RCs has found that RPs commonly occur in L2 (oral and written) production data (e.g., Algady, 2013; Gass, 1979; Hyltenstam, 1984; Kim, 2013; Pavesi, 1986). Experimental studies in this line of research have also indicated that RPs are more likely to show up in positions thought to be difficult for relativization (e.g., Algady, 2013; Gass, 1979; Hyltenstam, 1984).

Hyltenstam (1984) was the first to show that L2ers systematically produce resumptive RCs even when they are ungrammatical in both the L1 and the TL. He argued that resumptive RCs are easier for L2ers to process than filler-gap RCs, at least when there is strain on the processor, and that L2 resumption of this ilk thus arises for reasons of processing facilitation. He conjectured that such resumption represents a licit option for relativization in interlanguage grammar. However, these claims have not been thoroughly tested using psycholinguistic methods.

3. Previous experimental research on resumption in L1-English adults' RCs

To probe the sources of L2 English resumption, we follow in the footsteps of the psycholinguistic literature on mature native speakers. Studies in this vein use

offline judgment tasks and online processing tasks in tandem to tease apart representational phenomena from processing phenomena. In such studies on English, acceptability judgment data have shown that RPs receive uniformly low ratings, suggesting they are ungrammatical (e.g., Han et al., 2012; Heestand et al., 2011; Keffala & Goodall, 2011). Elicited production data have shown that RPs are produced at higher rates in environments considered difficult for relativization, indicating that they facilitate RC production (e.g., Ferreira & Swets, 2005; Morgan & Wagers, 2018; Zukowski & Larsen, 2004). While it is widely accepted that RPs facilitate RC production, whether they also ease real-time comprehension is more controversial. Some self-paced reading studies have shown that reading times (RTs) following RPs tend to be faster than those following gaps in (only) difficult-to-process RCs; this has been taken to suggest that RPs facilitate RC comprehension (e.g., Hammerly, 2022; Hofmeister & Norcliffe, 2013; Zenker & Schwartz, 2021). By contrast, Morgan et al. (2020) have recently argued that their self-paced reading data show that RPs hinder RC comprehension. Further RC experimentation is needed to clarify how RPs affect comprehension.

4. The present study

The work on which we report here is part of a larger dissertation project with two sub-studies: one on subject RCs, the other on direct object RCs (ORCs). Each sub-study has four main tasks: an elicited production task examining processing during production; a self-paced reading task (SPRT) examining processing during comprehension; an acceptability judgment task (AJT, with closely-translated versions in English, Korean, and Mandarin) examining offline acceptability of the sentence types in the first two tasks; a C-test (Zenker, in prep.) measuring English proficiency. This report focuses on the SPRT, AJT, and C-test data from the ORC sub-study. Testing L2ers in both their L1 and the TL is important because the distribution of RPs in Korean and Mandarin RCs is still not well understood.

The study addresses three research questions with respect to adult L2ers of English whose L1 is Korean or Mandarin: First, is resumption a licit representational option for ORC formation? Second, does resumption facilitate real-time comprehension of ORCs? And finally, does L2 proficiency influence the acceptability and/or processing of resumptive ORCs vs. filler-gap ORCs?

5. Participants

The tasks were completed by 90 English native speaker controls (ENSs), 69 L1-Korean L2ers of English (KLEs), and 76 L1-Mandarin L2ers of English (MLEs). Participants were tested asynchronously over the internet and compensated with course credit or a monetary award. Instructions were presented in the participant's native language. Table 2 shows that (prior to participant exclusions in either the SPRT or the AJT—see §7.1 and §8.1) the L2 groups were roughly matched in terms of age at testing, C-test proficiency scores, age of onset for acquiring English, and years of residence in English-speaking countries.

Table 2. Participant information (means and ranges, prior to exclusions)

Group	*n*	Age at Testing	C-test Score (max = 50)	Age of Onset	Years of Residence
ENS	90	26.98 (18–71)	42.74 (26–49)	—	—
KLE	69	26.29 (18–41)	29.30 (7–45)	9.35 (8–15)	0.32 (0–7)
MLE	76	28.14 (18–45)	28.86 (7–49)	9.71 (8–14)	0.04 (0–1)

ENSs were recruited on Prolific (www.prolific.co) and via the Linguistics Beyond the Classroom program at the University of Hawai‘i at Mānoa. KLEs and MLEs were recruited on university websites and internet job boards in, respectively, Korea and Taiwan. The experiment was coded using jsPsych (de Leeuw, 2015) and hosted on Cognition (www.cognition.run).

6. Experimental design and data analysis

The SPRT and the AJT had the same 2 × 3 design, crossing dependency type (henceforth, DEPENDENCY: gap vs. RP) and syntactic environment (henceforth, ENVIRONMENT: short-distance vs. long-distance vs. *wh*-island). The assumption behind inclusion of the short-distance, long-distance, and *wh*-island environments (henceforth, the short, long, and island conditions, respectively) is that they are increasingly difficult for relativization (see, e.g., Hawkins, 1999, 2004).

Data analysis and visualization were done in R (R Core Team, 2022), and the *lme4* (Bates et al., 2015) and *ordinal* (Christensen, 2019) packages were used for mixed-effects modeling. When constructing the models, we stayed as close to the maximal random-effects structure as possible while avoiding convergence errors and singular-fit warnings. Categorical predictors received simple coding to allow for ANOVA-style interpretations (reference level: Gap for DEPENDENCY, Short for ENVIRONMENT, and ENS for GROUP). For post-hoc testing, the alpha level (i.e., the significance threshold) was adjusted for multiple comparisons via the multivariate *t* distribution, and pairwise comparisons examined the gap vs. RP contrast for each ENVIRONMENT × GROUP combination. More information about the materials, data, and analyses can be found at the Open Science Framework repository for this study (https://osf.io/dkjb6).

7. Self-paced reading task

The SPRT probed the processing of RCs with gaps vs. RCs with RPs during real-time sentence comprehension. There were 30 critical items (6 conditions × 5 tokens, Latin-squared) and 42 fillers. The critical conditions are shown in Table 3. Fillers were designed to be comparable to critical items in both length and complexity. There were six filler types (7 tokens each): four types designed to be grammatical and two types designed to be ungrammatical. The presentation order for critical and filler trials was pseudorandomized so that no two critical items appeared consecutively.

Table 3. Self-paced reading task conditions (regions of interest underlined)

Environment	Example Stimuli in the Gap and RP Conditions
Short	I think Mary knows the man that these detectives arrested {__/*him} at the beginning of the week.
Long	Mary knows the man that I think these detectives arrested {__/*him} at the beginning of the week.
Island	Mary knows the man that I wonder which detectives arrested {*__/*him} at the beginning of the week.

The stimulus sentences were presented in word-by-word fashion on a computer screen in moving-window format. Each stimulus was displayed as a single line of text, and participants pressed the space bar to advance through the sentence. If the reading speed for a sentence was less than, on average, 200 ms per word, a message would appear asking the participant to read more carefully.

A comprehension question in multiple-choice format followed each stimulus. The purpose of these questions was twofold: (a) to ensure that participants were reading the sentences for meaning and (b) to test whether they correctly associated the gap or RP with the ORC's head. For critical trials, correctly answering the question relied on accurate dependency resolution. Comprehension questions for fillers targeted information in different parts of the sentence to discourage participants from developing strategies for answering (i.e., strategies that do not rely on a real understanding of the sentence). Participants were notified when their responses were incorrect. After reading the instructions, they needed to complete three practice items without errors before proceeding to the main phase of the task.

7.1. Data processing and analysis

Data from one KLE and four MLEs were lost due to experimental error. A further two KLEs and five MLEs were removed for having less than 50% accuracy on comprehension questions across all critical and filler trials. These exclusions left 90 ENSs, 66 KLEs, and 67 MLEs for the SPRT analyses. Following standard data-cleaning procedures for SPRTs (see Keating & Jegerski, 2015), RTs slower than 3,000 ms or faster than 200 ms were excluded, leading to the removal of 6% of the critical-trial data (ENS: 7%; KLE: 6%; MLE: 4%). Residual reading times (RRTs) were then calculated via a linear regression model with fixed effects for WORD LENGTH and WORD POSITION and with a random intercept for PARTICIPANT. RRTs over two standard deviations above the mean for each ENVIRONMENT × GROUP × REGION combination were replaced with that cutoff value, affecting 4% of critical-trial data (ENS: 3%; KLE: 5%; MLE: 5%).

RRTs at the two regions of interest (see §7.2) were analyzed using separate linear mixed-effects models. We predicted that when there is processing strain, RTs would be faster after RPs than after gaps at one or both regions of interest. Such a result would suggest that RPs facilitate ORC comprehension. The binary comprehension-question accuracy scores were analyzed using a mixed-effects

logistic regression model. Here, we expected accuracy to be higher with RPs than with gaps, at least in environments that put strain on the processor; this would provide further evidence that RPs facilitate ORC comprehension. Finally, to test for L2 proficiency effects, we repeated the RRT and comprehension question analyses for the L2 data, adding scaled English proficiency scores (henceforth, PROFICIENCY) as a continuous fixed effect. We reasoned that if processing becomes more efficient as proficiency increases, there should be a negative relationship between PROFICIENCY and magnitude of the RP facilitation effect.

7.2. Results: Reading time data

Figure 1 shows the mean RRTs in each condition plotted out for the three participant groups. Region numbers were coded such that Region 0 was where the gap or RP occurred. Following Hofmeister and Norcliffe (2013), we chose the two words following the gap or RP (i.e., Regions 1 and 2) as the regions of interest.

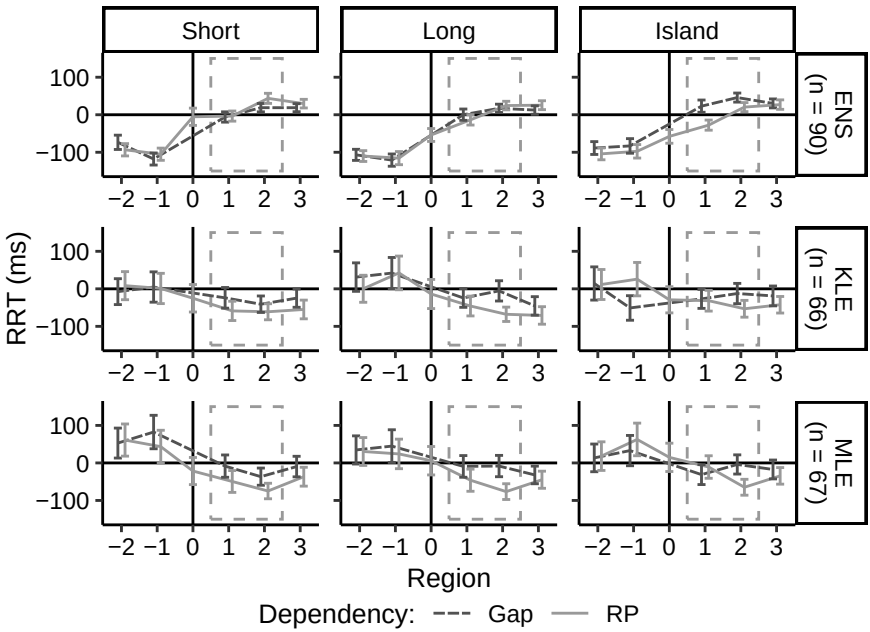

Figure 1. Mean residual reading times in the self-paced reading task; Region 0 is the gap/RP region; error bars are 95% confidence intervals; dashed rectangle indicates regions of interest

For ENSs, the relationship between RRTs for gap trials vs. RP trials depended on environment. In the short environment, where processing is expected to be relatively easy, there was a slight numerical advantage for gaps, with gap trials tending to be read somewhat faster than RP trials. By contrast, in the island environment, where processing should be most difficult, the pattern reversed, and

a substantial RP advantage in RRTs can be seen. For KLEs and MLEs, there was a numerical advantage for RP trials in one or both regions of interest across all environments, and the difference was largest in the long and island environments.

Our linear mixed-effects model for RRTs at Region 1 (model formula: RRT ~ DEPENDENCY * ENVIRONMENT * GROUP + (1 + DEPENDENCY | PARTICIPANT) + (1 + DEPENDENCY | ITEM)) found effects for DEPENDENCY (RP − Gap: $\beta = -20.15$, $SE = 6.68$, $p = .004$) and GROUP (KLE − ENS: $\beta = -30.69$, $SE = 12.73$, $p = .017$; MLE − ENS: $\beta = -22.65$, $SE = 12.66$, $p = .075$) as well as two significant three-way interactions (DEPENDENCY[RP − Gap] × ENVIRONMENT[Island − Short] × GROUP[KLE − ENS]: $\beta = 84.32$, $SE = 30.41$, $p = .006$; DEPENDENCY[RP − Gap] × ENVIRONMENT[Island − Short] × GROUP[MLE − ENS]: $\beta = 112.15$, $SE = 30.16$, $p < .001$). Pairwise comparisons revealed that the RP advantage in RRTs at Region 1 was significant only for ENSs and only in the island environment ($\beta = -51.04$, $SE = 15.22$, $p = .007$). ENSs thus treated RPs differently in the island condition vs. the other two conditions; they read RP trials faster in that environment, suggesting that when there is strain on the processor, the RP helps. The gap vs. RP contrast did not reach statistical significance for any of the other ENVIRONMENT × GROUP combinations in the Region 1 analysis.

At Region 2, our analysis (model formula: RRT ~ DEPENDENCY * ENVIRONMENT * GROUP + (1 + DEPENDENCY | PARTICIPANT) + (1 | ITEM)) found significant effects for DEPENDENCY (RP − Gap: $\beta = -30.97$, $SE = 4.75$, $p < .001$), ENVIRONMENT (Island − Short: $\beta = 13.87$, $SE = 4.82$, $p = .004$), and GROUP (KLE − ENS: $\beta = -68.56$, $SE = 16.02$, $p < .001$; MLE − ENS: $\beta = -74.50$, $SE = 15.94$, $p < .001$). All but one of the two-way interactions also reached at least marginal significance ($p < .10$; for details, see https://osf.io/6zafq). Pairwise comparisons showed that the gap vs. RP contrast was not significant for ENSs in any environment (all ps > .05). By contrast, the RP advantage here was significant for KLEs and MLEs: for KLEs in the long environment ($\beta = -65.71$, $SE = 13.39$, $p < .001$) and the island environment ($\beta = -40.41$, $SE = 13.40$, $p < .023$), for MLEs in all three environments (Short: $\beta = -36.43$, $SE = 13.04$, $p = .046$; Long: $\beta = -64.23$, $SE = 13.06$, $p < .001$; Island: $\beta = -58.88$, $SE = 13.06$, $p < .001$).

To review: Each group had shorter RRTs with RPs than with gaps in at least the island condition (ENS, KLE, MLE) if not also the long condition (KLE, MLE), the two syntactic contexts where processing should be harder. This pattern of results, when paired with complementary comprehension-question accuracy data, would constitute clear-cut evidence that RPs ease real-time ORC comprehension.

7.3. Results: L2 proficiency effects for the reading time data

The analysis of L2 proficiency effects in the RRT data at Region 2 (model formula: RRT ~ DEPENDENCY * ENVIRONMENT * GROUP * PROFICIENCY + (1 + DEPENDENCY | PARTICIPANT) + (1 + GROUP | ITEM)) found a significant positive relationship between L2 proficiency scores and RRTs ($\beta = 22.62$, $SE = 10.14$, $p = .028$). That is to say, as L2 proficiency increased, RRTs tended to get slower. Post-hoc testing revealed that this pattern was driven primarily by the KLE data in RP trials. However, we found no significant DEPENDENCY × PROFICIENCY

interaction (p = .295), signaling that there was no evidence of a meaningful relationship between proficiency score and strength of the RP advantage in RRTs.

7.4. Results: Comprehension-question accuracy data

Figure 2 shows the mean rates of comprehension question accuracy for each condition plotted out for the three participant groups. Strikingly, the mean accuracy rate is consistently higher for RP trials than for gap trials, although the magnitude of the difference varies across groups and environments.

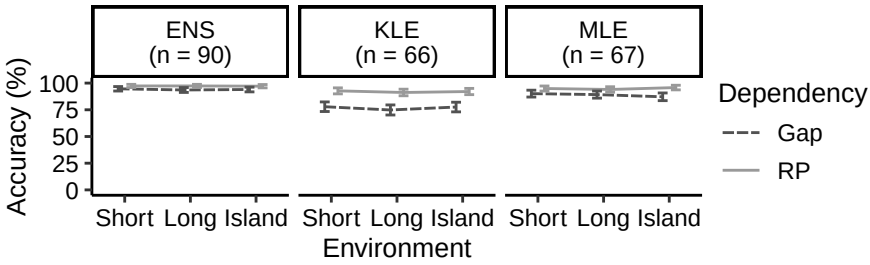

Figure 2. Mean accuracy on comprehension questions in the self-paced reading task; error bars are 95% confidence intervals

Our logistic mixed-effects regression analysis (model formula: ACCURACY ~ DEPENDENCY * ENVIRONMENT * GROUP + (1 | PARTICIPANT) + (1 + DEPENDENCY | ITEM)) showed that overall accuracy was significantly higher for RPs than for gaps (β = 1.13, SE = 0.12, p < .001). Accuracy was also significantly higher for ENSs than for KLEs (β = −1.59, SE = 0.24, p < .001) and MLEs (β = −0.68, SE = 0.25, p = .006). Also significant was the two-way DEPENDENCY[RP − Gap] × GROUP[KLE − ENS] interaction (β = 0.56, SE = 0.26, p = .030). Pairwise comparisons revealed that for ENSs, the gap vs. RP contrast reached marginal significance only in the long environment (β = 1.02, SE = 0.37, p = .053). This may have been due to a ceiling effect, as the mean in each condition was above 90%. For KLEs, the RP advantage was significant across all three environments (Short: β = 1.53, SE = 0.27, p < .001; Long: β = 1.44, SE = 0.25, p < .001; Island: β = 1.45, SE = 0.27, p < .001). Finally, for MLEs, the RP advantage was found to be significant only in the island environment (β = 1.41, SE = 0.34, p < .001).

To summarize: Mean accuracy on comprehension questions was numerically higher with RPs than with gaps for all groups and in all environments, and our analyses revealed the difference to be statistically significant for both L2 groups in at least the island condition, where processing was expected to be most difficult.

7.5. Results: L2 proficiency effects for comprehension question data

Our analysis of L2 proficiency effects in the comprehension question data (model formula: ACCURACY ~ DEPENDENCY * ENVIRONMENT * GROUP * PROFICIENCY + (1 | PARTICIPANT) + (1 + GROUP | ITEM)) found a significant positive

relationship between ACCURACY and PROFICIENCY ($\beta = 0.55$, $SE = 0.11$, $p < .001$): As L2 proficiency increased, accuracy increased as well. However, in regard to this comprehension question accuracy, no statistically meaningful relationship was detected between proficiency and strength of the RP advantage ($p = .562$).

8. Acceptability judgment task

The AJT tested the offline acceptability of the sentence types from the SPRT. As Table 4 illustrates, the AJT had the same design and stimuli as the SPRT had, but with a shortened temporal phrase at the end of the sentence (cf. Table 3).

Table 4. Conditions in the English acceptability judgment task

Environment	Example Stimuli in the Gap and RP Conditions
Short	I think Mary knows the man that these detectives arrested {__/*him} last week.
Long	Mary knows the man that I think these detectives arrested {__/*him} last week.
Island	Mary knows the man that I wonder which detectives arrested {*__/*him} last week.

There were English, Korean, and Mandarin versions of the task. The stimuli in the Korean and Mandarin AJTs were close translations of those in the English AJT. Importantly, we did not assume that the "island condition" in the Korean and Mandarin AJTs constituted genuine syntactic islands. To minimize effects of L1 knowledge on L2 performance, we had L2ers first complete the English AJT and then the version in their L1. Participants received different running lists for the SPRT, the English AJT, and (for L2ers) the Korean or Mandarin AJT such that they never saw any one item more than once in the same condition.

Participants were asked to rate the sentences on a 1−6 scale with an additional off-scale I-don't-know option, in the event they were unable to rate the sentence for whatever reason (e.g., unfamiliar words). As in the SPRT, a three-trial practice phase preceded the main phase of the experiment (in all three versions).

8.1. Data processing and analysis

Experimental error led to the loss of one KLE in the English AJT, one KLE in the Korean AJT, and one MLE in the Mandarin AJT. Data removal comprised two ENSs with a *yes* bias and 29 I-don't-know responses. The remaining raw 1−6 acceptability ratings were analyzed using a cumulative link mixed model. For English RPs, we expected mean ratings from ENSs to be low in all conditions and those from L2ers to be perhaps higher where processing is harder—i.e., the island condition, if not also the long condition—especially for MLEs, whose L1 has licit ORC resumption. For English gaps, we predicted mean ratings would be high for all groups in at least the short condition but low (at least for ENSs) in the island condition (where processing is most difficult and gaps are ungrammatical).

To test for L2 proficiency effects in the English AJT data, we repeated the mixed-effects model analysis for the L2 data, adding (English) PROFICIENCY as a continuous fixed effect. Here, we predicted that there would be a positive relationship between (scaled) proficiency score and strength of the preference for gaps over RPs, at least in environments where gaps are grammatical in English.

8.2. Results: Rating data

Figure 3 displays the mean ratings in the six critical conditions for each GROUP × LANGUAGE pairing (henceforth, PAIRING).

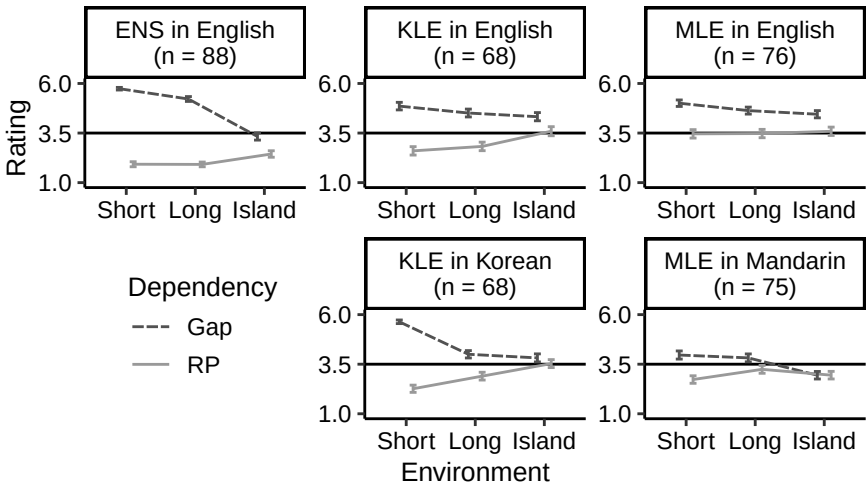

Figure 3. Mean ratings in the acceptability judgment task; horizontal line marks midpoint of rating scale; error bars are 95% confidence intervals

In the English AJT, the mean gap-trial rating for each group was on the top third of the rating scale in the short and long environments, where gaps are grammatical in English. The gap-trial rating in each case was at least somewhat lower in the island environment, where gaps are disallowed in English. By contrast, the mean ratings for RP trials were generally quite low, being either on the lower half of the scale or at the midpoint of the scale. Notably, there were no cases where the 95% confidence interval for RP-trial ratings was entirely on the top half of the rating scale, showing that no group consistently accepted RPs.[1]

[1] Some individuals did consistently accept RPs in English: A minority of participants in each group gave "acceptable" ratings (i.e., 4, 5, or 6) on at least four out of five RP trials in the short (ENS: 3%; KLE: 24%; MLE: 42%), long (ENS: 3%; KLE: 29%; MLE: 42%), and island (ENS: 14%; KLE: 43%; MLE: 38%) conditions. L2ers who were consistent RP acceptors were found to have significantly lower L2 proficiency scores than other L2ers.

The cumulative link mixed model on the rating data (model formula: RATING ~ DEPENDENCY * ENVIRONMENT * PAIRING + (1 + DEPENDENCY + ENVIRONMENT | PARTICIPANT) + (1 + DEPENDENCY + ENVIRONMENT + PAIRING | ITEM)) found significant effects for DEPENDENCY (RP − Gap: $\beta = -2.28$, $SE = 0.15$, $p < .001$), ENVIRONMENT (Long − Short: $\beta = -0.33$, $SE = 0.07$, $p < .001$; Island − Short: $\beta = -0.53$, $SE = 0.10$, $p < .001$), and, in English, PAIRING (KLE − ENS: $\beta = 0.40$, $SE = 0.20$, $p = .048$; MLE − ENS: $\beta = 0.73$, $SE = 0.20$, $p < .001$). Most of the two-way and three-way interactions were also significant (for details, see https://osf.io/6zafq). Crucially, pairwise comparisons revealed that in English, ratings were significantly higher for gaps than for RPs in all ENVIRONMENT × PAIRING combinations (all ps < .001). Taken together with the generally low mean ratings for RPs, these results suggest that no group (as a group) regarded resumption as an acceptable means of English ORC formation in any environment tested. In the Korean and Mandarin AJTs, the group results were similar to those in English in that the gap-over-RP advantage was significant in the short and long environments (all ps < .05); however, unlike in English, for neither KLEs nor MLEs was the contrast significant in their L1 "island condition" (both ps > .05).

8.3. Results: L2 proficiency effects for rating data

Our analysis of L2 English proficiency effects in the English AJT data (model formula: RATING ~ DEPENDENCY * ENVIRONMENT * PAIRING * PROFICIENCY + (1 + DEPENDENCY * ENVIRONMENT | PARTICIPANT) + (0 + DEPENDENCY * ENVIRONMENT | ITEM)) found a significant DEPENDENCY × PROFICIENCY interaction ($\beta = -1.63$, $SE = 0.41$, $p < .001$); as L2 proficiency rose, then, magnitude of the preference for gaps over RPs also rose. Post-hoc tests revealed this proficiency effect to be strongest for KLEs in the short and long environments (Short: $\beta = -3.46$, $SE = 0.74$, $p < .001$; Long: $\beta = -2.76$, $SE = 0.65$, $p < .001$), but the same numerical trend obtained in all ENVIRONMENT × GROUP combinations. These results indicate that as gains are made in L2 English proficiency, L2ers become more sensitive to the ungrammaticality of RPs in English ORCs.

9. Discussion and conclusion

For the SPRT data, one of our research questions asked whether resumption facilitates real-time ORC comprehension. The results offer compelling evidence from both reading time data and comprehension-question accuracy data that RPs ease ORC comprehension when the parser is under processing strain, and this was true for all participant groups. We also asked whether L2 proficiency influences the processing of gaps vs. RPs. Here, no significant relationship between English proficiency score and strength of the RP processing advantage emerged in either reading times or comprehension-question accuracy scores. Overall, the SPRT findings contribute to a growing body of evidence indicating that resumption can facilitate real-time comprehension of RCs (see Hammerly, 2022; Hofmeister & Norcliffe, 2013; Zenker & Schwartz, 2021). The fact that the processing effects

peaked later for KLEs and MLEs than for ENSs (i.e., at Region 2 rather than at Region 1) corroborates other research (e.g., Kilborn, 1992) which has found that processing tends to be slower and more effortful in an L2 than in an L1.

For the AJT data, one of our research questions asked whether pronominal resumption is a licit option for ORC formation. Our data analyses show that no group in this study consistently accepted RPs in any version of the AJT. We also found that all groups preferred gaps over RPs in English. These findings indicate that no group regarded resumption as an acceptable means of forming English ORCs.[2] A second question for the AJT data asked whether L2 proficiency affects the acceptability of gaps vs. RPs. Our L2 data analysis revealed that the size of the gap preference in the English AJT increased as English L2 proficiency rose. Taken together, these findings support the claim that RPs are ungrammatical in English (see Han et al., 2012; Heestand et al., 2011; Keffala & Goodall, 2011).

The present SPRT findings stand in opposition to Morgan et al.'s (2020) claim that RPs hinder comprehension. One possible reason that Morgan et al.'s data patterned how they did is that their SPRT stimuli were designed in such a way that the RPs could not assist with dependency resolution; this is demonstrated in (3), where both *Miss Piggy* (the dependency's head) and *Miss Cat* (another NP in the sentence) are legitimate potential antecedents for *her* (the RP). Any possible processing facilitation conferred by the RP would have been entirely undermined by this ambiguity (indeed, this ambiguity could in fact have instead been a source of processing difficulty).

(3) It was Miss Piggy that Miss Cat understood why Mr. Dog poked {__/her} with a pencil.

(Morgan et al., 2020, p. 6, Table 3)

This issue with Morgan et al.'s (2020) stimuli serves as a reminder as to why such RPs are thought to facilitate processing in the first place: RPs ease production/parsing by overtly establishing/maintaining coreference in difficult-to-process RCs (e.g., Asudeh, 2004; Hawkins, 1999; Hyltenstam, 1984; Keenan & Comrie, 1977). The RP in (3) cannot operate as an aid to processing precisely because it has more than one possible antecedent in the sentence. In our stimuli, by contrast, the RP always uniquely identified the head of the RC.[3]

---

[2] This holds of all groups (as groups). For L2ers, this minimally signifies that with rising English proficiency, they become sensitive to the ungrammaticality of English resumptive RCs. Still, as mentioned in fn. 1, a nontrivial number of our L2 participants did accept RPs in the two more difficult-to-process environments. Determining whether these results support Hyltenstam's (1984) claim that L2ers' (initially) create resumptive RCs as a licit means of RC formation—when doing so is ungrammatical in their L1—is still underway.

[3] The fact that the RP advantage we observed varied across environments (for L2ers as well as ENSs) should go some way towards allaying Morgan et al.'s (2020) concern that in other studies of this kind (examining native speakers), there may have been pragmatic cues in the stimuli that were primarily responsible for the processing facilitation that was sometimes found. See Morgan et al. for discussion of this nontrivial methodological point.

Directions for future research include testing whether RPs facilitate RC processing in L1/L2 Korean and L1/L2 Mandarin (see also fn. 2). In particular, we plan to investigate why the MLEs in our study did not consistently accept RPs in the Mandarin AJT (cf. Table 1). One possibility is that the positioning of the RP may have been too close to the head noun in the sentences we used, as in (4).

(4) 黃 麗 認識 那 個 我 好奇 哪 些 偵探
Huang Li renshi na ge wo haoqi na xie zhentan
Huang Li know DEM2 CL 1S wonder which CL detective
上 個 星期 逮捕 {__/他} 的 男人。
shang ge xingqi daibu {__/ta} de nanren
last CL week arrest {__/3MS} REL man
'Huang Li knows the man that I wonder which detectives arrested *(*him) last week.'

In conclusion, our data analyses strongly indicate that pronominal resumption is not a licit relativization option for any group in this study; however, RPs can still ease real-time comprehension of ORCs when there is strain on the processor.

References

Algady, Dola. (2013). *The acquisition of relative clauses: How do second language learners of Arabic do it?* [Doctoral dissertation]. University of Wisconsin-Milwaukee.

Asudeh, Ash. (2004). *Resumption as resource management* [Doctoral dissertation]. Stanford University.

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Christensen, Rune H. B. (2019). *ordinal—Regression models for ordinal data* [R package version 2019.12-10]. https://CRAN.R-project.org/package=ordinal

de Leeuw, Joshua R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavioral Research Methods*, *47*, 1–12.

Ferreira, Fernanda, & Swets, Benjamin. (2005). The production and comprehension of resumptive pronouns in relative clause "island" contexts. In Anne Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Routledge.

Gass, Susan M. (1979). Language transfer and universal grammatical relations. *Language Learning*, *29*, 327–344.

Hammerly, Christopher. (2022). The pronoun which comprehenders who process it in islands derive a benefit. *Linguistic Inquiry*, *53*, 823–835.

Han, Chung-hye. (2013). On the syntax of relative clauses in Korean. *Canadian Journal of Linguistics*, *58*, 319–347.

Han, Chung-hye, Elouazizi, Noureddine, Galeano, Christina, Görgülü, Emrah, Hedberg, Nancy, Hinnell, Jennifer, Jeffrey, Meghan, Kim, Kyeong-min, & Kirby, Susannah. (2012). Processing strategies and resumptive pronouns in English. In Nathan Arnett & Ryan Bennett (Eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics* (pp. 153–161). Cascadilla.

Hawkins, John A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, *75*, 244–285.

Hawkins, John, A. (2004). *Efficiency and complexity in grammars*. OUP.

Heestand, Dustin, Xiang, Ming, & Polinsky, Maria. (2011). Resumption still does not rescue islands. *Linguistic Inquiry*, *42*, 138–152.

Hitz, John, & Francis, Elaine J. (2016). On the usefulness of formal judgment tasks in syntax and in second-language research: The case of resumptive pronouns in English, Turkish, and Mandarin Chinese. *Linguistics*, *54*, 1241–1280.

Hofmeister, Philip, & Norcliffe, Elisabeth. (2013). Does resumption facilitate sentence comprehension? In Philip Hofmeister & Elisabeth Norcliffe (Eds.), *The core and the periphery* (pp. 225–246). CSLLI Publications.

Hyltenstam, Kenneth. (1984). The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In Roger W. Andersen (Ed.), *Second languages: A cross-linguistic perspective* (pp. 39–58). Newbury House.

Keating, Gregory D., & Jegerski, Jill. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*, 1–32.

Keenan, Edward L., & Comrie, Bernard. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*, 63–99.

Keffala, Bethany, & Goodall, Grant. (2011). *Do resumptive pronouns ever rescue illicit gaps in English?* [Poster]. CUNY 2011 Conference on Human Sentence Processing, Stanford University, Stanford, CA.

Kilborn, Kerry. (1992). On-line integration of grammatical information in a second language. In Richard J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 337−350). North-Holland.

Kim, Chae-Eun. (2013). *Asymmetries in the production of relative clauses: First and second language acquisition* [Doctoral dissertation]. University of Hawai'i at Mānoa.

Kwon, Nayoung. (2008). *Processing of syntactic and anaphoric gap-filler dependencies in Korean: Evidence from self-paced reading time, ERP, and eye-tracking experiments* [Doctoral dissertation]. University of California, San Diego.

Morgan, Adam M., von der Malsburg, Titus, Ferreira, Victor S., & Wittenberg, Eva. (2020). Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive pronouns hinder comprehension. *Cognition*, *205*, Article 104417.

Morgan, Adam M., & Wagers, Matthew W. (2018). English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry*, *49*, 861–876.

Pan, Victor Junnan. (2016). *Resumptivity in Mandarin Chinese: A minimalist account*. de Gruyter.

Pavesi, Maria. (1986). Markedness, discoursal modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition*, *8*, 38–55.

R Core Team. (2022). *R: A language and environment for statistical computing*. https://www.R-project.org

Song, Jae Jung. (2003). Resumptive genitive pronouns in Korean relative clauses: Distribution and explanation. *SKY Journal of Linguistics*, *16*, 139–160.

Zenker, Fred. (in prep.). *Development and validation of a C-test for the assessment of general English ability*.

Zenker, Fred, & Schwartz, Bonnie D. (2021). Resumptive pronouns facilitate processing of long-distance relative clause dependencies in second language English. *Proceedings of the Linguistic Society of America*, *6*, 325–339.

Zukowski, Andrea, & Larsen, Jenny. (2004, March). *The production of sentences that we fill their gaps* [Poster]. 17th CUNY Conference on Human Sentence Processing, University of Maryland, College Park, MD.

Proceedings of the 47th annual Boston University Conference on Language Development

edited by Paris Gappmayr
and Jackson Kellogg

Cascadilla Press Somerville, MA 2023

ISSN 1080-692X
ISBN 978-1-57473-087-6 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com