Noisy-Channel Processing in Standard Arabic Relative Clauses

Nicole Dodd (ncdodd@ucdavis.edu)
Department of Linguistics, University of California, Davis, Davis, CA 95616

Fatima Boush (fatimaboush@gmail.com)
Department of Cognitive Sciences, United Arab Emirates University, Al-Ain, UAE

Tommi Leung (leung@uaeu.ac.ae)
Department of Cognitive Sciences, United Arab Emirates University, Al-Ain, UAE

Fernanda Ferreira (ferreira@ucdavis.edu)
Department of Psychology, University of California, Davis, Davis, CA 95616

Emily Morgan (eimorgan@ucdavis.edu)
Department of Linguistics, University of California, Davis, Davis, CA 95616

Abstract

This study investigates sentence processing in Standard Arabic (SA) by examining subject- and object-extracted relative clauses (SRCs and ORCs) through eye tracking. We test memory- and expectation-based theories of processing difficulty, and whether good-enough or noisy-channel processing leads to misinterpretations in ORCs. Our results find increased processing difficulty in ORCs, supporting expectation-based theories; however, this processing difficulty is not localized to the disambiguating region (relative clause verb) as predicted, but rather at the integration of the second noun phrase (relative clause NP). The findings support good-enough/noisy-channel processing theories, suggesting that readers may accept a noisy SRC interpretation of an ORC, and thus bypass integration costs at the RC NP.

Keywords: sentence processing; noisy channel processing; good enough processing; Standard Arabic; eye tracking and reading; working memory; expectations

Introduction

What makes some sentences more difficult to read and comprehend than others? Two main types of theories aim to explain causes of sentence processing difficulty: memory-based theories (e.g., Gibson, 1998; Gibson, 2000) and expectation-based theories (e.g., Hale, 2001; Levy, 2008b). These theories are often tested cross-linguistically using subject- and object-extracted relative clauses. In subject-extracted relative clauses (SRCs), the noun phrase (NP) subject of the matrix clause is also the subject of the RC; in object-extracted relative clauses (ORCs), the NP subject of the matrix clause is the object of the RC (Figure 1).

Memory-based theories predict more processing difficulty when reading structures that utilize more working memory during incremental processing. One example of how this difficulty presents is with structures with long distance dependencies between constituents. Humans have limited computational resources, so readers incur more processing costs the longer they maintain constituents with incomplete

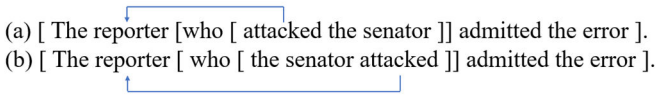

Figure 1: Example (a) SRC and (b) ORC in English. Dependencies between the relative clause verb and matrix clause subject are shown in blue.

dependencies in memory. For example, in English relative clauses, the dependency from the RC verb to the matrix noun is longer for ORCs than for SRCs (Figure 1). This results in more demands on working memory while reading ORCs, and thus more processing difficulty. Further, an additional cost is paid upon integrating the long dependency with the existing structure of the sentence (e.g., integrating "the reporter" upon resolving the dependency at "attacked"). This phenomenon is formalized in the Dependency Locality Theory (Gibson, 2000), which states that the cost of processing and integrating two elements is directly proportional to the length of the dependency between the elements.

Expectation-based theories posit that items that are less expected or lower frequency in context are more difficult to process. In syntactic processing, expectation-based processing difficulty arises when the reader encounters an element that violates their expectations for the upcoming syntactic parse. The reader then pays a processing cost proportional to the difficulty of updating their expectations. In English, SRCs are more common and thus more expected than ORCs. When reading an ORC, readers will incur processing difficulty after reading the relative pronoun "who", where the reader expects to encounter a verb (e.g., "attacked"), signaling an SRC, but instead encounters an NP (e.g., "the senator"), signaling an ORC (Figure 1). Many expectation-based theories operationalize this cost using surprisal theory, calculated as the negative log-probability of a word given previous context (Hale, 2001; Levy, 2008b). Words with a larger surprisal value are more surprising in

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

context and are therefore predicted to be read slower than words with smaller surprisal values.

Violated expectations during reading can both cause increased processing difficulty and result in the misinterpretation of a sentence. This is the case in models of good-enough (Ferreira et al., 2002; Ferreira & Patson, 2007; Ferreira & Lowder, 2016) and noisy-channel processing (Gibson et al., 2013; Keshev & Meltzer-Asscher, 2021; Levy, 2008a; Levy, 2011; Levy et al., 2009, but cf. Cutter et al., 2022). Models of good-enough processing state that readers often construct "good-enough" superficial representations of input during processing. Errors in language comprehension can then occur when readers fail to appropriately access lexical or grammatical constructions. Readers may reanalyze the input when their initial representation is incorrect, but the lingering incorrect interpretation can interfere with arriving at the correct meaning of a sentence. In the case of syntactic processing, encountering a structure that is unexpected in context may cause a reader to reanalyze, yet still accept the interfering "good-enough" interpretation of the input. Noisy-channel processing makes a similar claim, operationalized through statistical reasoning. Language input takes place in noisy circumstances – such as human error and competing environmental conditions – and this noise affects language processing strategies. Noisy-channel processing theories thus suggest that language users weigh the probability of a given sentence structure against the probability of noisy input during sentence processing. In cases where different syntactic structures are possible but one is higher probability than the other, a reader may assume noise in the input and make a number of "edits" to a sentence to arrive at the higher-probability interpretation. Both theories predict that readers experience increased processing difficulty when encountering violated expectations, and may accept the wrong, but more probable, interpretation of the sentence. Since both good-enough and noisy-channel processing theories make similar predictions, we don't attempt to differentiate between them here and consider them jointly.

Early studies that investigated processing differences in SRCs and ORCs cross-linguistically suggested a SRC processing bias, based on evidence from English (e.g., King & Just, 1991), German (Schriefers et al., 1995), Dutch (Mak et al., 2002), Japanese (Ueno & Garnsey, 2008), and Korean (Kwon et al., 2010). Subsequent studies, however, found that this "subject advantage" in relative clause processing was not universal; in Chinese (Hsiao & Gibson, 2003) and Basque (Carreiras et al., 2010), SRCs were harder to process than ORCs. Further, findings from these previous studies do not conclusively support one processing theory over the other. While some studies (e.g., Konieczny & Doring, 2003) have found evidence directly in support of and in contradiction to one theory, others have found evidence that both constraints can contribute to processing costs. For example, Staub (2010) found that both memory-based constraints and expectations contributed to processing difficulties in English SRCs and ORCs. Notably, these difficulties manifested in distinct behaviors: difficulty due to memory constraints presented as longer go-past reading times while difficulty due to violated expectations presented as increased regressive saccades.

The variance among these cross-linguistic findings can often be attributed to differences in typological factors such as word order (e.g., SVO vs. SOV), clause-headedness (head-initial vs. head-final), relative clause position in a sentence (pre-nominal vs. post-nominal), and the use of resumptive pronouns (RPs) (Lau & Tanaka, 2021). Languages with different typological features are not evenly represented in previous research; for example, more research has been done on SVO and SOV languages than VSO languages. Our research takes steps towards diversifying this body of research by investigating Standard Arabic, a morphosyntactically-complex language that is under-represented in psycholinguistic literature.

Standard Arabic

Standard Arabic (SA) is a Semitic language written right-to-left and uses both SVO and VSO word order. SA is mainly used in official governmental or media domains, and native Arabic speakers typically learn SA alongside their regional dialect used for everyday communication. Arabic relative clauses exhibit two linguistic properties that are under-represented in sentence processing literature: VSO word order and grammaticalized resumptive pronoun clitics. A sample SRC and ORC demonstrating these features is shown in Figure 2. Crucially, the only difference between an SRC and ORC with VSO word order in the relative clause is the presence of an object resumptive pronoun clitic on the ORC verb. Many Arabic dialects require the use of resumptive object pronoun clitics; SA strictly requires them only in relative clauses where the matrix subject is indefinite, but they are still strongly preferred in all relative clauses.

One previous study has investigated language processing in Arabic relative clauses. Dodd and Morgan (2022) conducted a self-paced reading task to determine whether memory- or expectation-based theories accounted for processing differences between SRCs and ORCs. Memory-based theories predicted equal processing difficulty for SRCs and ORCs as the matrix subject dependency is resolved at the relative clause verb in both conditions due to the object resumptive pronoun clitic. Expectation-based theories predicted more processing difficulty for ORCs as the less-frequent structure, which was determined through a corpus analysis. They found that ORCs were harder to process than SRCs, supporting expectation-based theories. However, they found that the significant difference in processing times was at the relative clause NP, not the relative clause verb as predicted. They attributed this difference to possible spillover effects in the self-paced reading paradigm. Further, by looking at comprehension question accuracy, they found that some readers accepted SRC interpretations of ORC sentences. Recent research in Hebrew relative clauses, which is typologically similar to Arabic, demonstrated that readers prefer high-frequency but grammatically incorrect interpretations of sentences over their grammatical but infrequent counterparts (Keshev & Meltzer-Asscher, 2021).

Table 1: Predicted behavioral outcomes for veridical processing, misreading, and good-enough/noisy-channel processing.

Theory	Comparison	Predicted behavior at disambiguating region (RC verb)
Veridical processing	Cor ORC vs. Cor SRC	Increased processing difficulty
Misreading	Incor ORC vs. Cor ORC	Decreased processing difficulty due to skipping RP clitic
	Incor ORC vs. Cor SRC	No significant difference; ORC misread as SRC
Good-enough/noisy-channel processing	Incor ORC vs. Cor ORC	No significant difference; both read without skipping RP clitic
	Incor ORC vs. Cor SRC	Increased processing difficulty; incor ORC read similarly to cor ORC

The authors thus hypothesized that a similar good-enough or noisy-channel processing effect could be happening in Arabic.

Current Study

Our study builds upon previous work by explicitly investigating what causes ORC misinterpretations. We do this by conducting an eye tracking experiment, which provides more granular measures of processing behavior than self-paced reading, and by asking comprehension questions directly probing the interpretation of the relative clause.

We first assessed whether ORCs were harder to process than SRCs, in line with Dodd and Morgan (2022) and in support of expectation-based theories, and whether we observe processing differences at the relative clause verb as predicted, or the relative clause NP. We also consider whether processing difficulty from violated expectations manifests as a distinct processing behavior, in line with Staub (2010). We then asked whether readers were initially misreading ORCs as SRCs, or correctly reading ORCs yet accepting noisy SRC interpretations (Keshev & Meltzer-Asscher, 2021). On the one hand, Arabic SRCs and ORCs differ by only a RP clitic, which may be easily missed. On the other hand, SRCs are more frequent than ORCs, so readers may accept a noisy but preferred SRC interpretation.

To answer these questions, we conducted three analyses (Table 1). First, we compared SRCs and ORCs with correct comprehension question answers (Cor ORC vs. Cor SRC) to determine differences during veridical processing. Predictions from the processing theories investigated here all point to the relative clause verb as the primary region of interest. Memory-based theories predict equal processing difficulty between SRCs and ORCs at the RC verb: due to the inclusion of the object RP, the matrix noun can be correctly integrated upon seeing the verb regardless of whether it is the subject or object of the RC. On the other hand, expectation-based theories predict increased processing difficulty at the RC verb as this is where the clause is disambiguated between an SRC or ORC. If Staub's (2010) results on different processing behaviors from memory- versus expectation-based difficulties extend to Arabic, then we would expect any observed expectation-based processing difficulty to manifest in increased regressive saccades.

Our next two analyses investigated misreading versus good-enough/noisy-channel processing by comparing incorrectly understood ORCs to both correct ORCs (Incor ORC vs. Cor ORC) and correct SRCs (Incor ORC vs. Cor SRC). If readers are misreading the ORC verb by skipping the RP, then in those cases the ORC verb would be read as identical to the SRC verb (because the object RP is the only difference between the two). We would then see no significant difference between incorrect ORC trials and correct SRC trials at the RC verb. However, we *would* see a significant difference between incorrect and correct ORC trials, as only correct ORC trials pay the processing cost of reading the resumptive pronoun. On the other hand, if good-enough or noisy-channel processing is occurring, then incorrect ORC trials would be read identical to correct ORC trials at the relative clause verb, with readers paying the cost of reading the resumptive pronoun for all ORCs. Incorrect ORC interpretations would then arise from later good-enough or noisy-channel processing. In this case, we would see no significant difference between incorrect and correct ORC trials at the relative clause verb, but a significant difference between incorrect ORC and correct SRC trials.

Methods

Participants

Forty-seven native Arabic speakers proficient in Standard Arabic (all women; mean age: 19; *SD* = 1.41) were recruited from the United Arab Emirates University (UAEU). Participants were offered both course credit and 40 AED (~15 USD) in cash compensation for their participation. Participants completed a detailed language history questionnaire in which they rated their proficiency in listening, speaking, reading and writing (scale from 1 to 7) in SA and all their other known languages. Participants were considered proficient in SA if they (1) selected Arabic as their native language, and (2) scored their proficiency in each area for SA at 4 or higher. One participant was excluded for selecting English as their native language. Further, we established an a priori criterion to exclude any participant who scored lower than 75% accuracy on comprehension questions on filler items, but all participants performed above this criterion. One final participant was excluded due to a technical error during their experiment.

Materials

We used the 40 items from Dodd and Morgan (2022), then created an additional 45 items following the same design (Figure 2). For each sentence, the matrix clause was SVO and the relative clause was VSO. This meant that readers first read the matrix clause subject, followed by the relative pronoun, and then the RC verb in both conditions. Arabic

	spillover 1	matrix verb	RC NP	RC verb	RC pronoun	matrix NP
(a)	بالخطأ	اعترف	السناتور	هاجم	الذي	الصحفي
	b=il=xaðˤʕ-i	<ʔ>ʕ<ta>rafa	a:=si:na:tu:r	h<a:>ʒam	a:la-ði	a:=sˤaħafi-u
	to=DET=error-ACC	admit<3SG.M.PST>	DET=senator	attack<3SG.M.PST>	who-3SG.M	DET=reporter-NOM

SRC: "The reporter who attacked the senator admitted the error."

(b)	بالخطأ	اعترف	السناتور	هاجمه	الذي	الصحفي
	b=il=xaðˤʕ-i	<ʔ>ʕ<ta>rafa	a:=si:na:tu:r	h<a:>ʒam=ahu	a:la-ði	a:=sˤaħafi-u
	to=DET=error-ACC	admit<3SG.M.PST>	DET=senator	attack<3SG.M.PST>=3SG.M.ACC	who-3SG.M	DET=reporter-NOM

ORC: "The reporter who the senator attacked admitted the error."

Figure 2: Sample (a) SRC and (b) ORC stimuli. Arabic sentences and English glosses are read right to left. The red circles indicate the disambiguating region: the RC verb. The only difference between an SRC and ORC with VSO word order in the RC is the presence of an object RP clitic on the RC verb in the ORC condition.

nouns, verbs, and pronouns are marked for both number and gender, so matrix and RC nouns were matched on number and gender so that the head of the RC would not be disambiguated by number and gender marking on the RC verb. Finally, all stimuli were presented in a non-diacritized format, as is standard for written publications in SA.

We conducted a norming study on our 45 new items to confirm that the subject and object of each RC were equally plausible in both clause conditions (e.g., "the reporter attacked the senator" is as plausible as "the senator attacked the reporter"). Native Arabic speakers ($N = 80$; mean age: 32; $SD = 10.64$) were recruited through Prolific and asked to rate the plausibility of each sentence on a Likert scale (1 = highly implausible, 7 = highly plausible). Plausibility ratings were collected for both the full stimuli sentences (e.g., "The reporter who attacked the senator admitted the error") and the relative clauses as simplified transitive sentences (e.g., "The reporter attacked the senator"). The study also included implausible distractor sentences as an attention check. Five stimuli were excluded after a paired t-test revealed substantial discrepancies between plausibility ratings in the SRC and ORC conditions for those items. The 40 remaining new stimuli combined with the 40 items from Dodd and Morgan (2022) resulted in 80 total stimuli. We also included 80 unrelated filler sentences for a total of 160 sentences. Experimental items were counterbalanced in a Latin square design.

Comprehension questions targeting comprehension of the relative clause appeared after all stimuli ("Which of the following happened?" (a) the reporter attacked the senator, (b) the senator attacked the reporter) and general comprehension questions appeared after all filler items. Dodd and Morgan (2022) had simple "Yes" or "No" options but reported that this framing may have biased the results. We asked a more open-ended question and offered full-sentence options, and randomized whether the correct or incorrect answer was presented first.

Apparatus and Procedure

The experiment took place at the UAEU in the Department of Cognitive Science's eye tracking lab using the Eyelink 1000 Plus eye tracker (SR Research). Right eye gaze movements were recorded via a high-speed 35 mm lens on a desktop mount at a sampling rate of 1000 Hz. Sentences were displayed on a high definition (1920 x 1080 pixels) 24" BENQ ZOWIE XL 2430 monitor at 80 cm viewing distance. Text was presented in 20-point Times New Roman, a proportionally spaced typeface.

Participants were tested individually in a quiet, isolated room. Instructions were given verbally in English and were presented in both English and SA text on the screen. The language of instruction at UAEU is English, so a baseline proficiency in English was assumed. The eye tracker was calibrated and validated using the default 9-point grid calibration. Participants completed two practice trials and were allowed to ask questions before proceeding with the experiment.

The experiment was divided into eight blocks of 20 items each to allow for breaks throughout the experiment. Calibration accuracy was assessed before the start of each block and the eye tracker was re-calibrated as necessary. At the beginning of each trial, a right-aligned asterisk was placed at the onset site of the first letter of the sentence. Once the participant fixated for at least 800 ms, the sentence appeared and replaced the asterisk. Participants pressed the spacebar on the keyboard once they were finished reading the sentence, then used the mouse to select the correct comprehension question answer on the next screen. The experiment took about 60 minutes for each participant.

Analysis and Results

Data were cleaned using SR Research's Data Viewer. Following standard procedures, fixations that were less than 80 ms and within one character of each other were merged, and remaining fixations less than 80 ms or longer than 1,200 ms were excluded. We also excluded trials where significant track loss occurred and fixations where a blink occurred on the target word. This resulted in a total data loss of 5.8%.

After cleaning our data, we investigated the comprehension question accuracy rates for SRCs versus ORCs. We found an overall accuracy rate of 84.4%, with a 93.5% accuracy for SRCs compared to only 75.4% accuracy for ORCs. In terms

Table 2: Model estimates and CrIs for the main predictor variable (Clause Type or Correctness) for all three analyses. Estimates marked with an asterisk are significant.

		Go-past time		Total fixation duration		First pass regression	
Analyses		***RC verb***	***RC NP***	***RC verb***	***RC NP***	***RC verb***	***RC NP***
Cor ORC vs. Cor SRC	β:	5.00	38.66*	33.52	59.97*	0.02	0.11*
	CrI:	[-20.23, 30.30]	[-2.65, 79.61]	[-19.24, 84.15]	[24.67, 94.84]	[-0.13, 0.18]	[0.00, 0.23]
Incor ORC vs. Cor ORC	β:	-2.68	-59.94*	-3.21	-76.91*	-0.13	-0.26*
	CrI:	[-34.09, 28.51]	[-110.93, -8.58]	[-65.70, 61.83]	[-127.94, -26.58]	[-0.43, 0.13]	[-0.46, -0.07]
Incor ORC vs. Cor SRC	β:	17.04	-24.33	44.06	-27.90	0.02	-0.13
	CrI:	[-21.51, 57.79]	[-68.44, 19.02]	[-32.30, 112.93]	[-73.32, 17.96]	[-0.28, 0.28]	[-0.35, 0.06]

of number of trials, we observed 1,747 correct SRC trials, 1,413 correct ORC trials, and 462 incorrect ORC trials. We also calculated accuracy by participant and found substantial variance by individual. ORC scores ranged from 10% to 100%; seven participants (out of 47) scored less than 50%, while only two participants scored 100%. Conversely, the lowest score on SRCs was 70%, with the second lowest score being 75%, and ten participants scored 100%.

For our analysis, interest areas were divided into regions as illustrated in Figure 2. Up to 3 spillover regions (one word each) were analyzed when sentences were long enough. All sentences had at least one spillover region. For each region, we calculated the following eye tracking metrics using the Get Reading Measures package from SR Research: *first fixation duration* (the duration of the first fixation on a region), *first pass duration* (the sum of all first pass fixations before leaving a region for the first time), *go-past time* (the sum of all first pass fixations on a region, including any time spent reading previous material, until progressively leaving the region for the first time), *total fixation duration* (the sum of all fixations on a region), *first pass regression* (a binary measure indicating whether the reader's first pass through a region ended with a regressive saccade to an earlier part of the sentence), and *first pass skip* (a binary measure indicating whether a reader skipped a region on first-pass reading).

Data were analyzed using the *brms* package in R (Bürkner, 2017). Linear and logistic mixed-effects regression models were fit for each eye tracking metric. We fit individual models in each interest area for each analysis (Table 1). The models included sum-coded fixed effects for Clause Type (ORC: 1, SRC: -1 for Cor ORC vs. Cor SRC and Incor ORC vs. Cor SRC models) or Correctness (correct: 1, incorrect: -1 for Incor ORC vs. Cor ORC models). Correctness was a measure of participant-specific accuracy by trial. We also included control predictors of Word Length (continuous) and Trial Index (continuous). We used the maximal random effects structure justified by the design (Barr et al., 2013), resulting in random intercepts for Participant and Item, and random slopes by Clause Type for both Participant and Item. We consider the model estimates as reliable if the credible interval (CrI) does not include 0, or over 95% of the sampled posterior distribution is over or under 0 in the predicted direction. Selected model estimates are reported in Table 2.

Correct ORCs versus Correct SRCs

We found no significant effects for Clause Type at the RC verb for any fixation metric. However, we did find significant effects at the RC NP for go-past time, total fixation duration, and first pass regressions. This indicates that ORCs were associated with significantly longer reading times and higher regression rates.

Incorrect ORCs versus Correct ORCs

We found no significant effects for Correctness at the RC verb for any fixation metric, but again found significant effects at the RC NP for go-past time, total fixation duration, and first pass regressions. This demonstrates that readers spent more time reading and made more regressions away from the RC NP when they correctly interpreted ORCs versus when they incorrectly interpreted them.

Incorrect ORCs versus Correct SRCs

We found no significant effects for Clause Type at the RC verb or the RC NP for any fixation metric.

Post-hoc Analysis: Relative Clause Verb

Despite finding no significant effects on the relative clause verb for any of our analyses, we wanted to determine whether readers paid any sort of additional processing cost for reading the resumptive pronoun clitic on the ORC verb. All of our models included Word Length as a control variable, so it was possible that this extra cost was present but proportional to length. We re-ran all our models in the RC verb region excluding Word Length and found significant effects by Clause Type and Correctness. Correct ORCs had longer first fixation, first pass, go past, and total duration times than correct SRCs, and incorrect ORCs also had longer first pass, go-past and total duration times than correct SRCs. However, there was no significant difference in reading times between incorrect and correct ORCs. These data suggest that readers do register the RP on the ORC verb and read ORC verbs longer than SRC verbs, regardless of whether they correctly interpret the ORC. However, this increased reading time at the ORC verb is proportional to its increased length due to the inclusion of the RP clitic.

Discussion

This study set out to test theories of sentence processing and comprehension in Standard Arabic. Previous research found that ORCs are harder to read than SRCs in Arabic, supporting expectation-based theories. However, questions remained about the frequent misinterpretations of ORCs. We expanded upon this research by investigating the cause of these misinterpretations through an eye tracking study.

We first tested whether SRCs or ORCs were harder to process in Arabic in an effort to corroborate previous findings (Dodd & Morgan, 2022). To answer this question, we analyzed differences during veridical processing, between SRC and ORC trials with correct comprehension question answers. We further tested previous findings from English that these processing difficulties manifest in distinct behaviors (Staub, 2010), with longer go-past times indicating difficulty from memory constraints and increased regressive saccades indicating difficulty from violated expectations.

Our results showed that ORCs were read significantly longer than SRCs overall, in line with previous findings and in support of expectation-based theories. We found that ORCs had longer go-past times, total fixation durations, and increased regressive saccades. According to Staub (2010), this would indicate processing difficulty from both memory limitations and violated expectations. Our findings contradict these predictions – our stimulus design indicates that there should only be expectation-based difficulty, but the observed difficulty manifested in behaviors attributed by Staub to both expectation- and memory-based difficulty.

Despite the relative clause verb being the disambiguating region and the predicted site of processing difficulty, our results showed no significant difference in processing by clause type in this region. Readers do spend more time reading the relative clause verb for ORCs than SRCs, but this is proportional to the added length from the resumptive object pronoun clitic. Rather, we found significant differences in processing at the relative clause NP, in line with previous findings from self-paced reading (Dodd & Morgan, 2022). Taken together, these results suggest that readers pay a processing cost when integrating the relative clause NP in the globally less-expected ORC structure, even though they had previously received the disambiguating information. These results are not predicted by either the memory- or expectation-based theories.

We received a comment from a non-native speaker that readers could initially be interpreting the RC verb with the resumptive object pronoun as a SRC with a RP and null object noun phrase (e.g., "The reporter who attacked him…" where the null object for "him" must be resolved by context). In this case, the reader would maintain uncertainty about the interpretation of the clause until they encountered the relative clause NP, despite seeing an RP clitic. There is some evidence that null object constructions are allowed in SA (Al-Sharafi & Gubaily, 2023), and we conducted a corpus analysis (Maamouri et al., 2010) to see how frequently this structure occurred. We found four examples of this structure occurring, out of 5,725 total SRCs. 1,543 ORCs occurred in the same corpus. If readers are maintaining an SRC interpretation upon reading the RC verb with a RP, they would be relying on a global expectation for SRCs over ORCs, despite the fact that more granular expectations should favor ORCs when the RC verb has a RP. We cannot definitively say whether this is the case or whether readers do probabilistically switch to an ORC interpretation at the RC verb, yet still pay a processing cost at the RC NP. In either case, our findings show that readers are not doing strict incremental processing possible while reading Arabic RCs.

We then asked whether misinterpreted ORCs were caused by readers misreading ORCs as SRCs by skipping the RP clitic, or correctly reading ORCs and instead accepting a noisy but preferred SRC interpretation. To answer this question, we analyzed the differences in processing behaviors in incorrect ORC trials compared to both correct ORC and correct SRC trials. In the case of misreading, we expected to see no significant difference between incorrect ORCs and correct SRCs at the RC verb, but a significant difference between incorrect and correct ORCs. This would indicate that incorrect ORCs were read as SRCs by skipping the RP clitic, and distinctly from incorrect ORCs. In the case of good-enough or noisy-channel processing, we expected to see the opposite: a significant difference between incorrect ORCs and correct SRCs at the RC verb, and no significant difference between incorrect and correct ORCs. This would indicate that readers register the resumptive pronoun clitic on the verb for incorrect ORCs, yet later accept a noisy, incorrect interpretation.

We found no significant differences between incorrect and correct ORC trials at the relative clause verb. This suggests that readers were not misreading the verb when misinterpreting ORCs, as incorrect ORCs behave like correct ORCs. Rather, differences between incorrect and correct ORCs manifested at the relative clause NP, where correct ORCs had longer reading times and higher regression rates. On the other hand, incorrect ORCs and correct SRCs had no significant differences at the relative clause NP, meaning incorrect ORCs behaved similarly to correct SRCs and differently from correct ORCs at the NP. So, incorrect ORCs behave similarly to correct ORCs at the relative clause verb, but similarly to correct SRCs at the relative clause NP.

Overall, our results suggest that the locus of processing difficulty in Arabic ORCs is *not* at the disambiguating region, as predicted by previous theories. Rather, the locus of processing difficulty is at the integration of the second NP. Even though readers already have the disambiguating information to know that this NP must be the subject of the relative clause, integrating this NP into the less frequent structure to arrive at a correct ORC interpretation appears to cause processing difficulty. Accepting a noisy SRC interpretation of an ORC skips the integration cost at the relative clause NP, resulting in comparable processing difficulty between incorrect ORCs and correct SRCs in this region. These results thus support noisy-channel processing over misreading as an explanation for why some ORCs are misinterpreted as SRCs.

Acknowledgements

We would like to acknowledge the valuable native speaker assistance with stimuli design from Manar Al-Shatarat and Shayma Hassouna, and contributions to stimuli design and data processing by our undergraduate research assistants, Mariam Ali and Reem Suleiman. Thank you to Dr. Meera Al-Kaabi for being the initial point of contact for the UAEU research collaboration and assisting with materials revision.

This research was supported by the National Science Foundation Doctoral Dissertation Research Improvement Grant to ND, EM, and FF [grant number 2235106], the UC Davis Society of Hellman Fellows Program to EM, and the UC Davis College of Letters and Science to EM. The Penn Arabic Treebank Part 3 v 3.2 was provided by the Linguistic Data Consortium through a Data Scholarship to ND.

References

Al-Sharafi, Y. M., & Gubaily, M. A. (2023). Investigating the Null Object in Arabic Language. *مجلة أبحاث*, *10*(1).

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4:328*.

Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1).

Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, *115*(1), 79–92.

Cutter, M. G., Filik, R., & Paterson, K. B. (2022). Do readers maintain word-level uncertainty during reading? A pre-registered replication study. *Journal of Memory and Language*, *125*, 104336.

Dodd, N., & Morgan, E. (2022). Expectations and Noisy-Channel Processing of Relative Clauses in Arabic. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1).

Ferreira, F., & Lowder, M. W. (2016). Prediction, Information Structure, and Good-Enough Language Processing. In *Psychology of Learning and Motivation* (Vol. 65, pp. 217–247). Elsevier.

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In *Image, Language, Brain* (pp. 95–126).

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001*, 1–8.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, 1–58.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. *Proceedings of the ICCS/ASCS-2003 Joint International Conference on Cognitive Science*, 1–6.

Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and Linguistic Factors Affecting Subject/Object Asymmetry: An Eye-Tracking Study of Prenominal Relative Clauses in Korean. *Language*, *86*(3), 546–582.

Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa: A Journal of General Linguistics*, *6*(1), 34.

Levy, R. (2008a). A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 234–243.

Levy, R. (2008b). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 1055–1065.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, 106(50).

Maamouri, M., Bies, A., Kulick, S., Krouna, S., Gaddeche, F., & Zaghouani, W. (2010). *Arabic Treebank: Part 3 v 3.2 LDC2010T08*. Philadelphia: Linguistic Data Consortium.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50–68.

Meltzer-Asscher, A. (2021). Resumptive Pronouns in Language Comprehension and Production. *Annual Review of Linguistics*, *7*, 177–194.

Schriefers, H., Friederici, A. D., & Kuhn, K. (1995). The Processing of Locally Ambiguous Relative Clauses in German. *Journal of Memory and Language*, *34*, 499–520.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, *23*(5), 646–688.