Expectations and Noisy-Channel Processing of Relative Clauses in Standard Arabic
Nicole Dodd, Emily Morgan
University of California, Davis

An outstanding question in psycholinguistic research is what makes sentences more difficult to process and why. Memory-based constraints theories [1] predict greater processing difficulty when reading cognitively-demanding structures, such as long-distance dependencies. On the other hand, expectations-based theories [2] claim that high-frequency structures are easier to process. One common test case for these theories is the processing of subject- versus object-extracted relative clauses (SRCs and ORCs) cross-linguistically [e.g., 3]. We contribute to this body of research by studying an under-represented language in psycholinguistic research: Standard Arabic.

Expectations-based theories predict that SRCs are easier to process as they are more frequent in Standard Arabic, which we confirmed through a corpus analysis using the Penn Arabic Treebank. However, memory-based theories predict that SRCs and ORCs are equally difficult to process. The relative clause word order (VSO) and use of resumptive object pronoun clitics result in comparable dependencies for both SRCs and ORCs, and thus comparable cognitive load (Fig 1).

We conducted a self-paced reading task. Stimuli included 40 pairs of sentences with an SRC and ORC alternation (Fig 1). 48 native Arabic speakers read each stimulus in one clause condition, plus 80 filler sentences. Readers also answered comprehension questions after all experimental and 20 filler items. Reaction times were analyzed by clause type using a linear mixed-effects model. SRCs had shorter RTs than their ORC counterparts (Fig 2), and were thus easier to process in Arabic, in line with expectations-based theories.

Our post-hoc analysis of comprehension question accuracy demonstrated an interesting trend in relative clause comprehension. We fit a logistic mixed-effects model to the questions that targeted relative clause comprehension, with clause type and correct answer (either "yes" or "no") as fixed effects, including their interaction. The results showed significant effects for both clause type and correct answer condition: participants were more likely to inaccurately respond to ORC comprehension questions, and to falsely respond with "yes" when the correct answer was "no" (Fig 3). The model additionally showed a significant interaction between clause type and correct answer condition, resulting in a subadditive effect. These results lead us to believe that Arabic readers are mistakenly interpreting ORCs as SRCs, especially when a noisy interpretation is suggested by the comprehension question, via a bias for "yes" answers.

As a follow up, we ask whether readers are skipping the resumptive pronoun in ORCs during reading, or reading an ORC yet accepting a noisy SRC interpretation given their expectations [4]. We have taken the first steps towards answering this question with a recall task. Using the same stimuli, we asked participants (n = 80) to read each sentence and then reproduce the sentence word-for-word from memory. Overall, error rates were low (<4%); however, we found that participants both misremembered ORCs as SRCs (71% of errors) and SRCs as ORCs, lending some support to noisy-channel processing as opposed to misreading. These outcomes raise further questions about the strength of semantic, grammatical, and orthographic features in influencing a reader's willingness to accept a noisy interpretation. We plan to conduct an eye-tracking experiment next to ask how these features interact, particularly in the case of reading resumptive pronoun clitics (about which little is currently known).

الصحفي الذي هاجم السناتور اعترف بالخطأ

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

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

الصحفي الذي هاجمه السيناتور اعترف بالخطأ

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

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

Figure 1: Sample stimuli. Arabic sentences are read right to left, and English are read left to right. The red circles indicate the area of interest: the relative clause verb. Arabic word order is such that the only difference between an SRC and an ORC (with matched gender and number for matrix and relative clause nouns) is the inclusion of an object pronoun clitic.
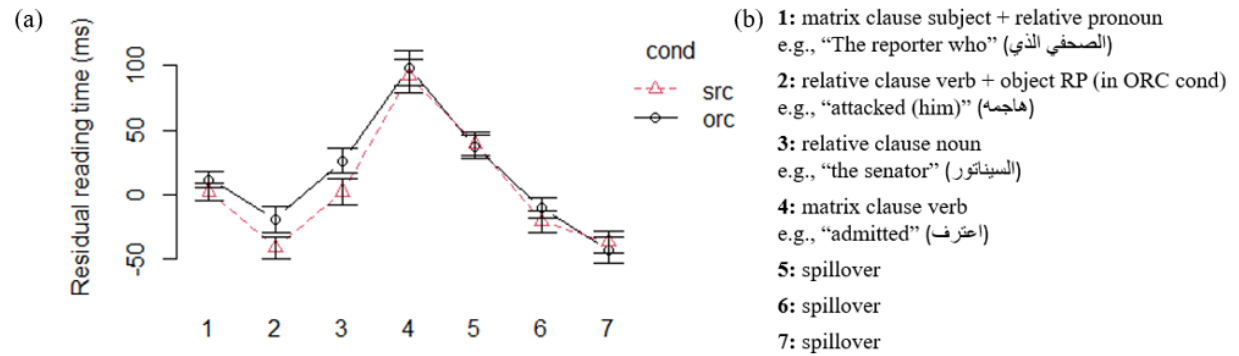

(b)
1: matrix clause subject + relative pronoun
e.g., "The reporter who" (الصحفي الذي)
2: relative clause verb + object RP (in ORC cond)
e.g., "attacked (him)" (هاجمه)
3: relative clause noun
e.g., "the senator" (السيناتور)
4: matrix clause verb
e.g., "admitted" (اعترف)
5: spillover
6: spillover
7: spillover

Figure 2: (a) Average residualized RTs for each region by clause type; (b) Regions of interest with Arabic examples and their English gloss.
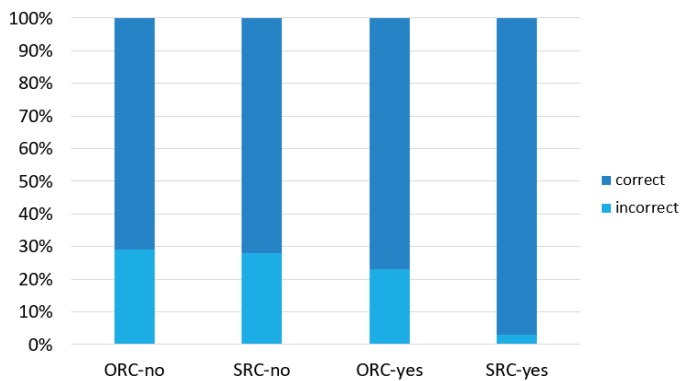

Figure 3: Proportion of correct comprehension question answers by clause type and correct answer condition.

References: [1] Gibson (1998). *Cognition 68(1).* [2] Hale (2001). *Proceedings of NAACL '01.* [3] King & Just (1991). *Journal of Memory and Language, 30(5).* [4] Keshev & Meltzer-Asscher (2021). *Cognitive Psychology, 124.*