Eye-Tracking Measures of Processing Difficulty in Standard Arabic

Nicole Dodd[1], Fatima Boush[2], Tommi Leung[2], Fernanda Ferreira[1], Emily Morgan[1]

[1]University of California, Davis, [2]United Arab Emirates University

What makes some sentences more difficult to process and why?

- **Memory-based theories**: sentences with cognitively-demanding structures (e.g., long distance dependencies) are harder to read [1]
- **Expectation-based theories**: unexpected or infrequent sentence structures are harder to read [2, 3]

Misaligned expectations during reading can also result in **misinterpretations** of a sentence (**good-enough** and **noisy-channel processing**) [4, 5]

Test case: subject- and object-extracted **relative clauses** [6]

(a) [The reporter [who [attacked the senator]] admitted the error].
(b) [The reporter [who [the senator attacked]] admitted the error].

Standard Arabic

Standard Arabic is written right-to-left and is mainly used in official governmental or media domains.

Native Arabic speakers typically learn the Standard dialect alongside their regional dialect for everyday communication.

Key features of Standard Arabic:

- Flexible SVO and VSO word order
- Grammaticalized resumptive object pronoun clitics

spillover 1	matrix verb	RC NP	RC verb	RC pron	matrix NP
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."

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."

Are SRCs or ORCs harder to read in Standard Arabic?

- Memory-based theories: equal processing difficulty for SRCs and ORCs due to similar dependency lengths from disambiguating region (relative clause verb) to matrix subject
- Expectation-based theories: ORCs are harder to process as they are less frequent than SRCs
- Both theories predict the **relative clause verb** as the locus of processing difficulty as it is the probabilistic disambiguating region

(a) الصحفي الذي هاجم السناتور اعترف بالخطأ.
(b) الصحفي الذي هاجمه السيناتور اعترف بالخطأ.

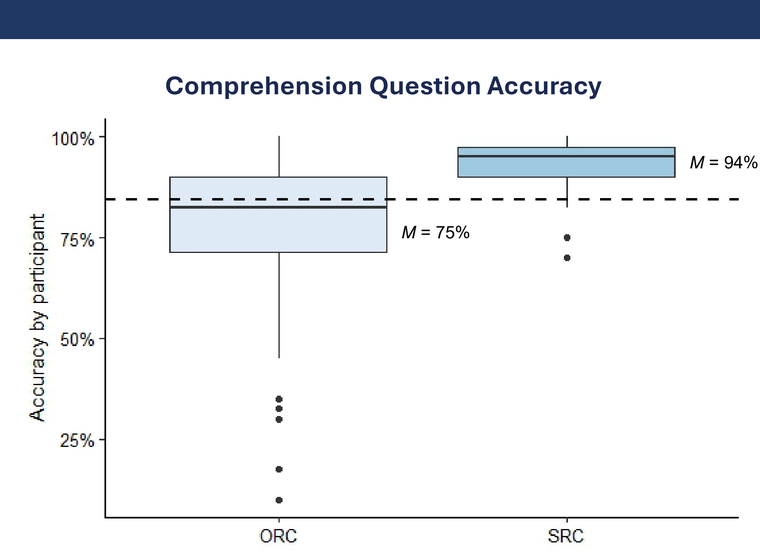
Previous Study: Self-Paced Reading

- ORCs were harder to process than SRCs, supporting **expectation-based theories** [7]
- Locus of processing difficulty was at the **relative clause NP**, not relative clause verb
- **Readers frequently misinterpreted ORCs**: 67% accuracy for ORCs versus 83% accuracy for SRCs

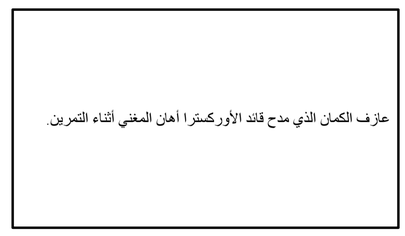
Are people **misreading ORCs as SRCs** by missing the resumptive pronoun clitic, or accurately reading them but **accepting a noisy but preferred SRC interpretation**?

Current Study: Eye Tracking

Theoretical Predictions



- Re-test memory- versus expectation-based processing theories
- Investigate **what causes ORC misinterpretations**: misreading or good-enough/noisy-channel processing



The violinist who flattered the conductor insulted the singer at the rehearsal.

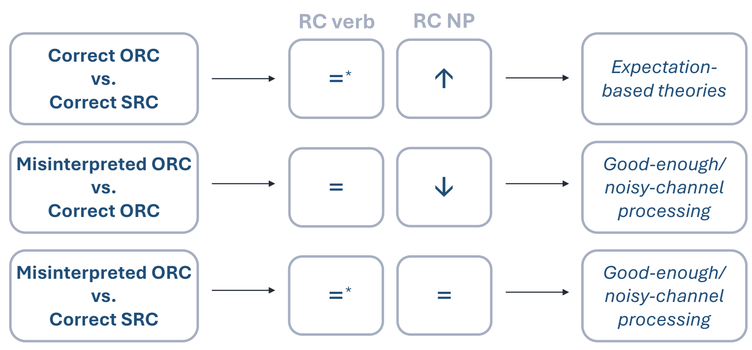
Methods:

- 47 native Arabic speakers from the United Arab Emirates University
- 80 pairs of SRC/ORC sentences with comprehension questions after every item
- Questions targeted interpretation of the relative clause

Results and Discussion

Comprehension Question Accuracy

Results



*no significant difference by clause type, but longer RTs proportional to length of RP clitic

- Longer RTs for correct ORCs compared to correct SRCs
- Misinterpreted and correct ORCs read similarly at the relative clause verb, but misinterpreted ORCs had shorter RTs at the relative clause NP
- Misinterpreted ORCs read longer than correct SRCs at the relative clause verb *proportional to the added length of the resumptive pronoun clitic*, then read similarly at the relative clause NP

Discussion:

- ORCs were read significantly longer than SRCs overall, in line with **expectation-based theories**
- The locus of processing difficulty is the relative clause NP, not the relative clause verb (the probabilistic disambiguating region)
- 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 probabilistic disambiguating information
- **Misinterpretations of ORCs appear to be caused by good-enough/noisy-channel processing:** readers register the resumptive pronoun clitic on the relative clause verb, but skip the integration cost at the relative clause NP to accept a noisy SRC interpretation

Key Takeaways

- ORCs are harder to read than SRCs, but this processing difficulty appears later than expected; readers are not engaging in strict incremental processing with fine-grained statistical expectations
- ORC misinterpretations are due to readers accepting a noisy or "good-enough" SRC interpretation over a dispreferred ORC interpretation

References

[1] Gibson (1998). *Cognition 68(1)*. [2] Hale (2001). *Proceedings of NAACL '01*. [3] Levy (2008). *Cognition 106(3)*. [4] Ferreira et al. (2002). *Curr Dir Psychol Sci 11(1)*. [5] Keshev & Meltzer-Asscher (2021). *Cognitive Psychology 124*. [6] Lau & Tanaka (2021). *Glossa: a journal of general linguistics 6(1)*. [7] Dodd & Morgan (2022). *CogSci '22 44(44)*.

Acknowledgements

- Thank you to our valuable native Arabic speakers Manar Al-Shatarat and Shayma Hassouna, and undergraduate research assistants Mariam Ali and Reem Suleiman for contributions to stimuli design.
- Thank you to Dr. Meera Al-Kaabi for being the initial point of contact for the UAEU research collaboration and assisting with materials.
- This research was supported by the NSF Doctoral Dissertation Research Improvement Grant to ND, EM, and FF, the UCD Society of Hellman Fellows Program to EM, and the UCD College of Letters and Science to EM.

UCDAVIS DEPARTMENT OF LINGUISTICS

Presenter: Nicole Dodd
Email: ncdodd@ucdavis.edu