Using event-related potentials to examine individual differences in the processing of pronominal reference

Robert Fiorentino[1], Alison Gabriele[1], and Lauren Covey[1]
[1]University of Kansas

Correspondence to: fiorentino@ku.edu

1. INTRODUCTION

Recent studies have shown variability in the processing of pronominal reference in both ambiguous contexts (1a/1b) and in contexts of 'referential failure' in which there is no gender-matching antecedent within the sentence (2a/2b) (Osterhout and Mobley, 1995; Nieuwland and Van Berkum, 2006; Van Berkum and Nieuwland, 2008; Nieuwland, 2014).

These contexts have been shown to give rise to a sustained, frontal negative shift (Nref) or a P600 (or both ERP components) depending on both the accompanying behavioral task and individual differences in working memory (e.g. Nieuwland, 2014). In both contexts, high working memory has been associated with Nref, a component that indexes the inability to assign a unique referent, as opposed to P600 (Nieuwland and Van Berkum, 2006; Nieuwland, 2014).

For 'referential failure,' it has been proposed that the presence of an explicit acceptability judgment task may give rise to P600s, suggesting attempts at co-reference despite the mismatch in gender (Nieuwland, 2014).

We further examine both referential ambiguity and referential failure using an end of sentence recall task, and including a wider range of cognitive measures, to better understand the factors that impact the resolution of pronominal dependencies.

2. STIMULI

Experiment 1: Referential Ambiguity
1a. **Tyler** grabbed **Eric** because *he* was falling down the stairs. (2 referents)
1b. **Janet** grabbed **Eric** because *he* was falling down the stairs. (1 referent)

Experiment 2: Referential Failure
2a. **Nicole** believed **Steven** because *he* was a very genuine person. (1 referent)
2b. **Nicole** believed **Alice** because *he* was a very genuine person. (no referent)

- 40 items per condition, plus 80 filler sentences (all grammatical sentences)
- Selected verbs without strong 1st or 2nd NP bias using ratings from Ferstl et al. (2011)
- Ambiguity manipulation confirmed using online judgment task (n=90 native English speakers) in which participants selected the correct referent for the pronoun or chose "Either"

Condition	Incomplete Sentence (e.g., Tyler grabbed Eric because he__)	Complete Sentence
2 referent (Exp. 1) (ambiguous)	82% "Either" (7% NP1, 11% NP2)	57% "Either" (26% NP1, 17% NP2)
1 referent (Exp. 1)	98% Gender-matching NP (1% other NP, 1% "Either")	94% Gender-matching NP (3% other NP, 3% "Either")
1 referent (Exp. 2)	98% Gender-matching NP (1% other NP, 1% "Either")	96% Gender-matching NP (2.5% other NP, 1.5% "Either")

3. METHOD

Participants
- 33 native English speakers recruited from university population (data collection ongoing)

Procedure
- Sentences presented one word at a time using RSVP (450ms word/300ms pause), with a fill-in-the-blank recall question following 1/3 of the sentences
- Stimulus presentation via Paradigm (Tagliaferri, 2005)
- EEG continuously recorded using Synamps2 amplifier (Compumedics Neuroscan, Inc.) and 32-channel Ag/AgCl electrode cap (Electro-Cap International, Inc.)

Post-EEG Tasks
- Offline judgment task in which participants read the target stimuli and judged which NP is the referent of the pronoun (1st NP, 2nd NP, or either)
- Participants were also tested on the following measures:
- *Reading Span* (Conway et al., 2005): 35.4-98.3% (Mean 63.5, SD 14.9)
- *Counting Span* (Conway et al., 2005): 21-90.4% (Mean 61, SD 15.4)
- *Number Stroop* (Bush & Shin, 2006):
  Accuracy (Incongruent-Congruent): -3.76-10% (Mean 3.3, SD 4)
 RT (Incongruent-Congruent): -34.6-309.9ms (Mean -93.2ms, SD 81.3)

4. EXPERIMENT 1 RESULTS: REFERENTIAL AMBIGUITY

All Participants (n=33)
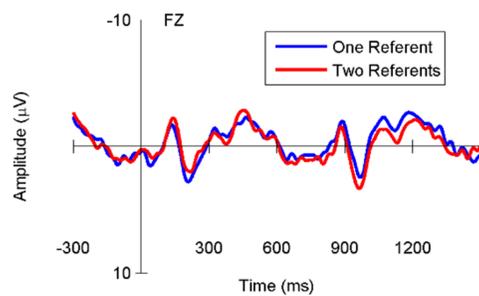
- No significant difference emerged between the two referent (ambiguous) and one referent conditions in the all participants analysis.
- Analyses comparing low and high working memory participants (e.g. Nieuwland and Van Berkum, 2006) did not reveal distinct ERP patterns.
- However, distinct patterns emerged when participants were divided based on whether or not they showed a positivity in the posterior region (500-1400ms) (see Nieuwland and Van Berkum, 2008).
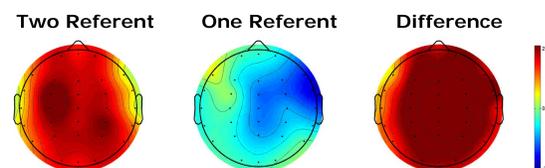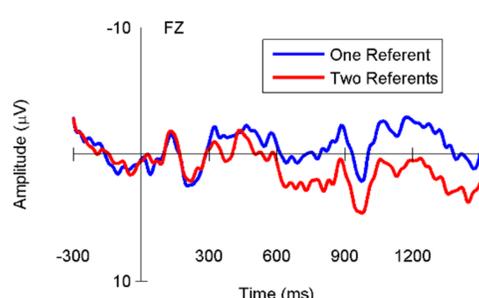
Group 1: Nref (n=15)

Topographic Plots (500-1400ms)

Two Referent / One Referent / Difference
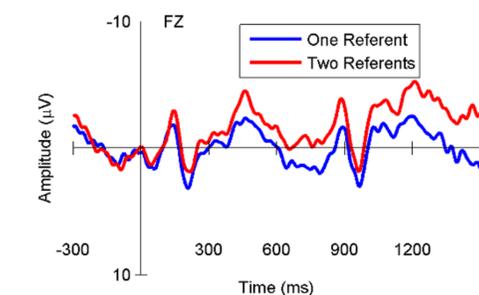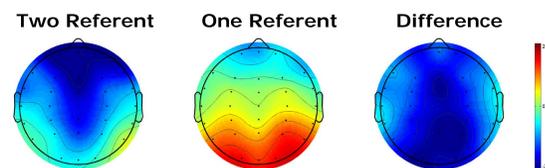
Group 1 (n=15) showed a sustained negativity in both the anterior and posterior regions.

Group 2: Broad Positivity (n=18)

Two Referent / One Referent / Difference
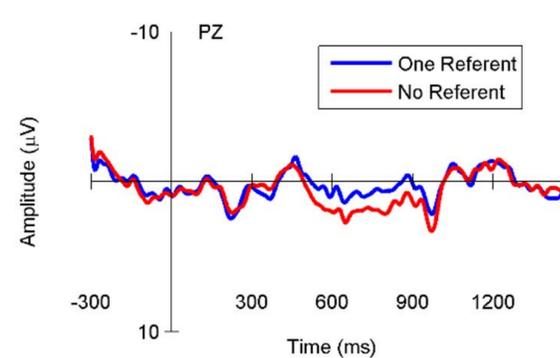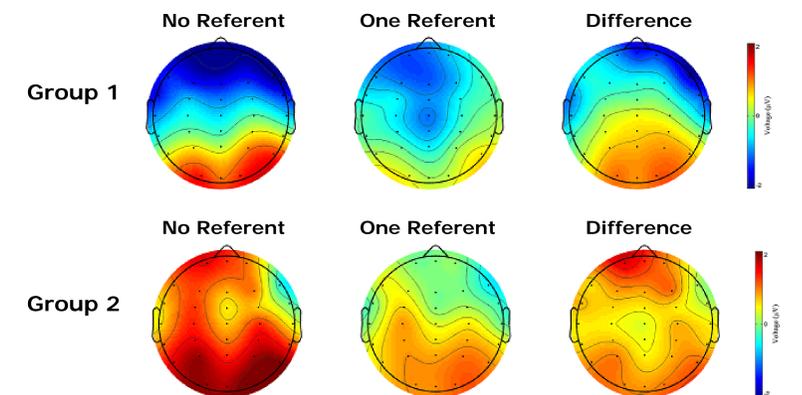
Group 2 (n=18) showed a sustained positivity which was broadly distributed.

5. EXPERIMENT 2 RESULTS: REFERENTIAL FAILURE

All Participants (n=33)

Topographic Plots (500-1400ms)

No Referent / One Referent / Difference
Group 1 / Group 2

6. SUMMARY OF RESULTS

Behavioral Results
End of sentence recall task: 83.8-100% (Mean 94.3, SD 5.3)

ERP Results
2x2 Repeated Measures ANOVA with Referent (one/two, one/none) and Anteriority (anterior, posterior) as within-participants factors

Experiment 1: Ambiguity (500-1400ms)
- All Participants: No significant effects
- Group 1 (n=15): sustained negativity consistent with Nref
- Group 2 (n=18): broad positivity

Experiment 2: Referential Failure
- 500-900ms: Significant positivity restricted to posterior region for all participants. Marginal interaction with Group.
- 500-1400ms: Sustained anterior negativity for Group 1 (not statistically robust)

Individual Differences
- Experiment 1: significant negative correlation between performance on counting span task and effect size in the anterior region (500-1400ms) for Group 1 (n=15).
- Experiment 2: significant positive correlation between performance on counting/reading span tasks and effect size in the posterior region (500-900ms) for all participants.

7. DISCUSSION

Experiment 1: Referential Ambiguity
- Nref was observed only for a subset of participants (Nieuwland and Van Berkum, 2006, 2008): participants who did not show a late positivity in the posterior region showed a sustained negativity consistent with Nref (Nieuwland and Van Berkum, 2008).
- Although working memory did not determine the two distinct patterns that emerged (cf. Nieuwland and Van Berkum, 2006), there was a significant relationship between counting span and Nref effect size (Group 1).

Experiment 2: Referential Failure
- A posterior P600 emerged for all participants despite absence of an explicit judgment task (Nieuwland and Van Berkum, 2006).
- P600 may not necessarily be associated with low working memory (cf. Nieuwland, 2014), as there was a significant positive correlation between working memory and P600 effect size.
- The group differences in Experiment 1 may also impact the processing of referential failure in Experiment 2: an anterior negativity emerges only in Group 1 (500-1400ms), although the effect is not robust.

References
Bush, G. & Shin, L. (2006). *Nature Protocols, 1*(1), 308-313.
Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). *Psychonomic Bulletin and Review, 12*(5), 769-786.
Ferstl, E.C., Garnham, A., & Manouilidou, C. (2011). *Behavior Research Methods, 43*, 124-135.
Nieuwland, M.S. (2014). *Journal of Memory and Language, 76*, 1-28.
Nieuwland, M.S. & Van Berkum, J.J. (2006). *Brain Research, 1118*, 155-167.
Nieuwland, M.S. & Van Berkum, J.J. (2008). *Brain and Language, 106*(2), 119-131.
Osterhout, L. & Mobley, L.A. (1995). *Journal of Memory and Language, 34*, 739-773.
Tagliaferri, B. (2005). Paradigm. Perception Research Systems, Inc, www.perceptionresearchsystems.com.
Van Berkum, J.J., Brown, C.M., & Hagoort, P. (1999). *Journal of Memory and Language, 41*, 147-182.