Journal of Neurolinguistics 40 (2016) 98–111

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Tracking sentence comprehension: Test-retest reliability in people with aphasia and unimpaired adults

Jennifer E. Mack [a, *], Andrew Zu-Sern Wei [a], Stephanie Gutierrez [a], Cynthia K. Thompson [a, b, c]

[a] Department of Communication Sciences and Disorders, Northwestern University, USA
[b] Cognitive Neurology and Alzheimer's Disease Center, Northwestern University, USA
[c] Department of Neurology, Northwestern University, USA

ARTICLE INFO

Article history:
Received 27 December 2015
Received in revised form 19 April 2016
Accepted 2 June 2016
Available online 24 June 2016

Keywords:
Eyetracking
Aphasia
Sentence comprehension
Test-retest reliability

ABSTRACT

Purpose: Visual-world eyetracking is increasingly used to investigate online language processing in normal and language impaired listeners. Tracking changes in eye movements over time also may be useful for indexing language recovery in those with language impairments. Therefore, it is critical to determine the test-retest reliability of results obtained using this method.

Methods: Unimpaired young adults and people with aphasia took part in two eyetracking sessions spaced about one week apart. In each session, participants completed a sentence-picture matching task in which they listened to active and passive sentences (e.g., *The* [$_{N1+Aux}$ *woman was*] [$_{V}$ *visiting/visited*] [$_{NP/PP2}$ *(by) the man*]) and selected between two pictures with reversed thematic roles. We used intraclass correlations (ICCs) to examine the test-retest reliability of response measures (accuracy, reaction time (RT)) and online eye movements (i.e., the likelihood of fixating the target picture in each region of the sentence) in each participant group.

Results: In the unimpaired adults, accuracy was at ceiling (thus ICCs were not computed), with moderate ICCs for RT (i.e., 0.4–0.58) for passive sentences and low (<0.4) for actives. In individuals with aphasia, test-retest reliability was strong (0.59 < ICC < 0.75) for accuracy and excellent (>0.75) for RT for both sentence types. Similarly, for the unimpaired listeners, reliability of eye movements was moderate for passive sentences (NP/PP2 region) and low in all regions for active sentences. But, for the aphasic participant group, eye movement reliability was excellent for passive sentences (in the first second after sentence end) and strong for active sentences (V and NP/PP2 regions).

Conclusion: Results indicated moderate-to-low reliability for unimpaired listeners; however, reliable eye movement patterns were detected for processes specific to passive sentences (e.g., thematic reanalysis). In contrast, individuals with aphasia exhibited strong and stable performance across sentence types in response measures and online eye movements. These findings indicate that visual-world eyetracking provides a reliable measure of online sentence comprehension in aphasia, and thus may be useful for investigating sentence processing changes over time.

* Corresponding author. 2240 Campus Drive Evanston, IL 60208, USA.
E-mail address: Jennifer-mack-0@northwestern.edu (J.E. Mack).

http://dx.doi.org/10.1016/j.jneuroling.2016.06.001
0911-6044/

1. Introduction

In the last twenty years, visual-world eyetracking has become one of the most fruitful techniques for investigating language comprehension and production. Eyetracking has informed models of normal sentence processing at linguistic levels ranging from phonetic and phonological to lexical, sentence and discourse processing. In addition to the large eyetracking literature on language processing in monolingual adult speakers (see review in Huettig, Rommers, & Meyer, 2011), eyetracking has been used to examine the development of word and sentence comprehension processes in children (e.g. Borovsky, Elman, & Fernald, 2012; Borovsky, Sweeney, Elman, & Fernald, 2014; Fernald, Perfors, & Marchman, 2006; Mani & Huettig, 2012; Marchman & Fernald, 2008; Nation, Marshall, & Altmann, 2003) as well as the mechanisms of word comprehension in bilingual and second-language listeners (Blumenfeld & Marian, 2011; Chambers & Cooke, 2009).

In addition, eyetracking is increasingly used to characterize linguistic impairments in clinical populations. In the aphasia literature, eyetracking has played an important role in characterizing both lexical (Laurinavichyute, Ulicheva, Ivanova, Kuptsova, & Dragoy, 2014; Mirman & Graziano, 2012; Mirman, Yee, Blumstein, & Magnuson, 2011; Yee, Blumstein, & Sedivy, 2008) and sentence processing deficits (Bos, Hanne, Wartenburger, & Bastiaanse, 2014; Cho & Thompson, 2010; Choy & Thompson, 2010; Dickey & Thompson, 2009; Dickey, Choy, & Thompson, 2007; Hanne, Burchert, De Bleser, & Vasishth, 2015; Hanne, Sekerina, Vasishth, Burchert, & De Bleser, 2011; Lee & Thompson, 2011a, 2011b; Mack, Ji, & Thompson, 2013; Meyer, Mack, & Thompson, 2012; Patil, Hanne, Burchert, De Bleser, & Vasishth, 2015; Sheppard, Walenski, Love, & Shapiro, 2015; Thompson & Choy, 2009). For example, in one study (Meyer et al., 2012), we examined eye movements in unimpaired older adults and individuals with aphasia as they listened to active (e.g., *The man was visiting the woman*) and passive sentences (e.g., *The man was visited by the woman*) and selected between two pictures with reversed thematic roles. Unimpaired adults showed evidence of incremental agent-first (A1) processing, initially interpreting (non-case-marked, animate) subjects as agents (cf. Hanne et al., 2015). After presentation of the disambiguating verb morphology (*visit**ing** vs. visit**ed***), they rapidly fixated the target picture in both sentence types; in passive sentences, this likely reflects thematic reanalysis in which the initial A1 interpretation is revised (Hirotani, Makuuchi, Ruschemeyer, & Friederici, 2011; Mack, Meltzer-Asscher, Barbieri, & Thompson, 2013). The aphasic individuals, in contrast, did not show incremental A1 processing, fixating both pictures with equal frequency prior to verb offset. Following presentation of the disambiguating verb morphology, they exhibited delays in fixating the target picture in active sentences and never consistently did so in passive sentences (cf. Hanne et al., 2011). These findings highlight differences between unimpaired adults and aphasic listeners with respect to sentence processing latency and incremental interpretation.

Other populations in which eyetracking has been used to quantify performance patterns in language processing include individuals with apraxia of speech (Lee, Mirman, & Buxbaum, 2014), autism spectrum disorder (Brock, Norbury, Einav, & Nation, 2008; Mirman, Irwin, & Stephen, 2012; Venker, Eernisse, Saffran, & Ellis Weismer, 2013), and children and adolescents with language impairments (Borovsky, Burns, Elman, & Evans, 2013; McMurray, Munson, & Tomblin, 2014; McMurray, Samelson, Lee, & Tomblin, 2010; Nation et al., 2003). Eyetracking is also a potentially useful tool for measuring change over time, reflecting learning and/or language recovery in individuals with impaired language, though little research to date has used eyetracking for this purpose. In one study, Kim and Lemke (2016) used eyetracking-while-reading to test a participant with acquired alexia prior to administration of a text-based reading treatment program. After treatment, using the same task, the participant evinced more normal-like eye movements, reflecting facilitation of a lexical-semantic reading strategy.

Establishing the test-retest reliability of visual-world eyetracking is an important step in evaluating the usefulness of this method as a measure of performance patterns or language change over time. However, relatively little research has addressed this issue. One previous study (Farris-Trimble & McMurray, 2013) investigated test-retest reliability of measures of lexical access derived from visual-world eyetracking. In two test sessions, unimpaired young adult participants heard words (e.g., *horn*) while viewing arrays of four pictures including the target picture, a cohort competitor (e.g., *horse*), a rhyme competitor (e.g., *corn*), and an unrelated competitor (e.g., *box*). Multiple parameters of the time course of fixations to each picture type were modeled for each participant and test session. Correlation analyses comparing the time course parameters across test sessions indicated moderate-to-high reliability, especially for the timing of the rise in fixations to the target picture. In contrast with lexical access, the test-retest reliability of visual-world eyetracking in sentence comprehension tasks has not yet been investigated.

Only one previous study of which we are aware has examined test-retest reliability of sentence comprehension accuracy in unimpaired adults and individuals with aphasia, and no previous studies have examined RT or online eye movements. McNeil and colleagues tested reliability of performance on the Revised Token Test, which examines comprehension accuracy with sentences of varying length and complexity (McNeil et al., 2015). Their unimpaired adults showed modest reliability, likely due to near-ceiling performance on the task. However, the aphasic individuals showed excellent test-retest reliability, suggesting that sentence comprehension ability was reliable across test sessions. However, other studies have revealed a high degree of within-subject variability in performance in aphasia, e.g., across different sentence comprehension tasks (Caplan, Waters, Dede, Michaud, & Reddy, 2007) and – beyond the domain of sentence comprehension – across test sessions in narrative production (Boyle, 2014), confrontation naming (Freed, Marshall, & Chuhlantseff, 1996), and attentional tasks (Villard & Kiran, 2015).

In the present study, we examined the reliability of eye movements as participants performed a sentence-picture matching task (i.e., matching an auditorily-presented sentence to one of two pictures depicting possible interpretations of the sentence, as in Meyer et al., 2012). Measures of response accuracy and latency (RT) in this task have played an important