Brain & Language 169 (2017) 57–63

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Short communication

Longitudinal changes in linguistic complexity among professional football players

Visar Berisha [a,*], Shuai Wang [b], Amy LaCross [b], Julie Liss [b], Pamela Garcia-Filion [c]

[a] Department of Speech and Hearing Science and School of Electrical Computer and Energy Engineering, Arizona State University, United States
[b] Department of Speech and Hearing Science, Arizona State University, United States
[c] University of Arizona, College of Medicine-Phoenix, Dept of Biomedical Informatics, United States

ARTICLE INFO

Article history:
Received 19 July 2016
Revised 12 February 2017
Accepted 25 February 2017

Keywords:
Chronic traumatic encephalopathy
Linguistic complexity
Natural language processing
Type-to-token ratio
Lexical density

ABSTRACT

Reductions in spoken language complexity have been associated with the onset of various neurological disorders. The objective of this study is to analyze whether similar trends are found in professional football players who are at risk for chronic traumatic encephalopathy. We compare changes in linguistic complexity (as indexed by the type-to-token ratio and lexical density) measured from the interview transcripts of players in the National Football League (NFL) to those measured from interview transcripts of coaches and/or front-office NFL executives who have never played professional football. A multilevel mixed model analysis reveals that exposure to the high-impact sport (*vs* no exposure) was associated with an overall decline in language complexity scores over time. This trend persists even after controlling for age as a potential confound. The results set the stage for a prospective study to test the hypothesis that language complexity decline is a harbinger of chronic traumatic encephalopathy.

1. Introduction

Repeated mild traumatic brain injury (mTBI), as experienced by professional athletes in high-impact sports, has been associated with an increased risk of developing chronic traumatic encephalopathy (CTE) Saulle & Greenwald, 2012. CTE is a progressive degeneration of brain tissue characterized by abnormal buildup of tau protein in neurons and glial cells (McKee et al., 2009). A conclusive diagnosis of CTE can be made only through post-mortem evaluation of p-tau deposits in the brain since behavioral manifestations of the disease overlap with other neurodegenerative disorders, such as Alzheimer's disease (Lenihan & Jordan, 2015; McKee et al., 2009).

The clinical presentation of CTE is diffuse and variable, but commonly includes disruptions in mood or behavior, cognitive decline, and/or motor impairments (see Lenihan & Jordan, 2015, for a review). In a review of 51 confirmed cases of CTE in athletes, McKee et al. find that one-third of the individuals were symptomatic when they retired from their sport and one-half showed signs of CTE within 4 years of retirement (McKee et al., 2009). Early and mild manifestations of these symptoms may exist unrecognized for years, masked by compensatory strategies or attributed to other etiologies or general personality traits. Yet early identification of probable CTE is critical for the development and testing of neuroprotective interventions. Fortunately, these subtle deficits can be revealed when pressure is exerted on cognitive resources.

The production of language, in the form of conversation or spontaneous written narrative, is a form of pressure on cognitive resources. It requires identifying words to express an idea, arranging these words in an order allowed by the language, all before initiating the articulatory muscles to even produce the speech, or fine motor control to write. As a result, a number of studies have demonstrated reductions in linguistic complexity of spoken and written discourse in patient populations with cognitive impairment and dementia (Kempler, 1995; Ripich, Vertes, Whitehouse, Fulton, & Ekelman, 1991; Roark, Mitchell, Hosom, Hollingshead, & Kaye, 2011; Snowdon et al., 1996). Our study is an extension of this work to a pre-symptomatic population at risk for CTE: players in the National Football League (NFL). We have compiled an extensive corpus of over 10,000 interviews with 10 NFL players (P), 9 of whom were on active rosters in the NFL as of the start of the 2016 season, and 18 NFL front office executives and coaches (C) who have never played professional football. This provides a rich and unique opportunity to evaluate longitudinal changes in language use for both groups (players and coaches/front-office executives) and to explore whether lexical complexity measures, derived

* Corresponding author at: Department of Speech and Hearing Science and School of Electrical Computer and Energy Engineering, Arizona State University, Tempe, AZ 85284, United States.

E-mail address: visar@asu.edu (V. Berisha).

http://dx.doi.org/10.1016/j.bandl.2017.02.003
0093-934X/

automatically from interview transcripts and tracked longitudinally, exhibit any decline.

There are a number of measures to estimate language complexity in the literature. These have been used to track second-language learning (Lu, 2012), child language development (Richards, 1987), and neurological health (Andreasen & Pfohl, 1976; Berisha, Wang, LaCross, & Liss, 2015; Borovsky, Saygin, Bates, & Dronkers, 2007; Bucks, Singh, Cuerden, & Wilcock, 2000; Fergadiotis, Wright, & West, 2013; Goldstein, Levin, Goldman, Clark, & Altonen, 2001; Kempler, 1995; Ripich et al., 1991; Roark et al., 2011; Snowdon et al., 1996; Tucker & Hanlon, 1998). Because this is an observational study, we focus on two simple, but robust, lexical complexity measures that have been established in other studies as potential markers of cognitive decline (Berisha et al., 2015; Kemper, Thompson, & Marquis, 2001; Le, Lancashire, Hirst, & Jokel, 2011): (1) lexical density (LD) – the ratio of the number of content words (verbs, nouns, adjectives, adverbs) to the total number of words in the text (Lu, 2012); (2) type-to-token ratio (TTR) - a ratio of the number of unique words in a text to the total number of words in the text. Qualitatively, the LD captures the density of ideas in a text and the TTR is a proxy measure for an individual's working vocabulary. Snowdon et al. related idea density in writing with cognitive health later in life (Snowdon et al., 1996). Roark et al. showed that the Content Density (a measure very similar to the LD used here) was a strong predictor of mild cognitive impairment (Roark et al., 2011). Similarly, the TTR has been used extensively by researchers as an analysis measure in a variety of clinical populations, including to assess linguistic differences between people with dementia associated with Alzheimer's and healthy controls (Bucks et al., 2000), to measure lexical diversity in aphasics (Fergadiotis et al., 2013), to assess the extent of lesions on speech production (Borovsky et al., 2007), and as a measure of linguistic diversity in preschool age children with language impairments (Richards, 1987). In our study we analyze longitudinal change in TTR and LD separately and compared between groups (players and coaches/front-office executives) using a multilevel mixed model approach.

Table 1
The table summarizes all players (top) and non-player personnel (bottom) that are included in the present study based on inclusion criterion described in the methods session. For the non-player personnel, we provide their age (as of the date of the first transcript), and the number of years of secondary education (college + postgraduate). For the players, we additionally include the number of years in the NFL (as of the date of the first transcript), their position on the field, and the average number of sacks per game (if Quarterback).

ID	Age	Education	Years in NFL	Position	Sacks/ game
Players (P)					
P1	29.8	4	7.1	Quarterback	1.79
P2	26.1	3	3.7	Nose Tackle	
P3	28.6	4	5.3	Quarterback	1.64
P4	23.7	4	0.27	Quarterback	2.51
P5	24.8	4	1.7	Cornerback	
P6	26.1	3	4.3	Wide Receiver	
P7	23.9	4	0.4	Quarterback	1.95
P8	29.7	4	6.4	Quarterback	1.81
P9	23.8	3	1.3	Quarterback	2.05
P10	23.3	3	1.1	Quarterback	2.74
Avg (St. dev.)	26.0 (2.6)	3.6 (0.5)	3.2 (2.5)		2.1 (0.4)

ID	Age	Education
Coaches + Front-Office Executives (C)		
C1	55.1	4
C2	57.9	4
C3	33.6	4
C4	36.1	4
C5	31.3	3
C6	63.0	4
C7	60.9	4
C8	50.0	5
C9	47.7	6
C10	42.8	6
C11	36.4	4
C12	63.6	2
C13	47.3	4
C14	52.8	4
C15	45.8	4
C16	58.7	5
C17	44.2	2
C18	36.9	4
Avg (St. dev.)	48.0 (10.4)	4.1 (1.0)

2. Results

In Table 1 we provide a list of individuals included in the study and summary statistics: the age (as of the date of the first interview), the number of years of college education, and, for the players, their position on the field, the number of years in the NFL (as of the date of the first interview), and the average number of times sacked per game (for the quarterbacks). An independent-samples *t*-test was conducted to evaluate differences in college education levels and in age between the two groups. There was no significant difference in college education between the two groups; $t(26) = 1.30$, $p = 0.21$. However, the C group mean age was significantly higher than that of the P group; $t(26) = 6.53$, $p < 0.0001$. For the C group, there was a positive correlation for LD with age [$r = 0.104$, $p < 0.001$]. For the P group, there was a negative correlation for TTR with age [$r = -0.425$, $p < 0.001$] and with number of years played [$r = -0.310$, $p < 0.001$].

In Table 2, we show the results of the mixed effects model. The analysis indicate that, compared to the coach/executive group, player status was associated with an overall decline in the TTR and LD scores over time. For TTR, there was a statistically significant decline over time [beta (SE): −0.016 (±0.002) points; $p_{unadj} = 0.012$]. The effect size increased [−0.027 (±0.003) points] after ($p_{adj} < 0.001$) adjusting for a potential confound (age at the time of first transcript). For LD, the decrease [−0.004 (±0.002) points] over time was statistically significant after controlling for age ($p = 0.012$) but not prior to [0.002 (±0.001); $p = 0.166$].

In Fig. 1, we show the slopes of the unadjusted model for the two complexity parameters plotted by group. Note that these measures represent the change in the complexity parameters over time and not their absolute values. A positive value indicates an increase in language complexity over time, whereas a negative value indicates a decrease in complexity over time. It is clear from the plot that there is a distinct difference in the trend exhibited by the two groups. For the P group, 70% of the individuals exhibited either a decline in LD or in TTR, and 40% exhibited a decline in both. For the C group, 44% of the individuals exhibited a decline in either parameter with 28% exhibiting a decline in both. As these figures show, this effect is more pronounced when only considering large negative changes in the language parameters (e.g. slopes < −0.0001).

To visualize data from individual subjects, Fig. 2 shows the TTR and LD plotted over time for two individuals – a coach and a player from the same team during the same period. We also show the line of best fit with 95% confidence bounds for each figure. The line was fit by converting the date to the number of days passed since 1/1/2005. These plots serve as exemplars indicative of the larger trend exhibited by the data. As the figure shows, both measures for the player exhibit a subtle decline over time; for the coach, the same parameters are increasing over time. While there are clear longitudinal trends in the plots, there is also a great deal of local variability. This is to be expected. Extensive work in the