Open Library of Humanities

Part of the Ubiquity Partner Network

Digital Studies / Le champ numérique

Method

How to Cite: Sapach, Sonja. 2020. "Tagging My Tears and Fears: Text-Mining the Autoethnography." *Digital Studies/Le champ numérique* 10(1): 4, pp. 1–28. DOI: https://doi.org/10.16995/dscn.328

Published: 31 August 2020

Digital Studies /
Le champ numérique

Sapach, Sonja 2020. "Tagging My Tears and Fears: Text-Mining the Autoethnography." *Digital Studies/ Le champ numérique* 10(1): 4, pp. 1–19. DOI: https://doi.org/10.16995/dscn.328

METHOD

Tagging My Tears and Fears: Text-Mining the Autoethnography

Sonja Sapach
University of Alberta, CA
sapach@ualberta.ca

This article, which was presented for the first time at CSDH/SCHN 2018, outlines a mixed methodology that combines digital text-mining techniques, like XML tagging and topic modelling, with autoethnography. In this article, I describe my intimate and rocky relationship with my data, referred to as my "life-as-text." As someone who has Complex Post Traumatic Stress Disorder, I have an extremely difficult time putting many of the events of my life into situational and linear context. The process of formally tagging my data requires me to overcome psychological blocks that have been created as defense mechanisms; thus, allowing me to honestly read and create associations in my dissociated "life-as-text." Conducting a macroanalysis of data that was gathered through both journaling, and the recording and transcription of "let's play" videos, has given me a new lens through which to view myself and my experiences. I see myself in my data. I see myself in ways that I cannot otherwise put into words. Text-mining my deepest memories and emotions provides me with an odd sense of stability in a seemingly chaotic assortment of words. In this article, I explain my autoethnographic journey and the powerful impact that digital text-mining has had on my relation to myself and my research.

Keywords: Autoethnography; Topic Modelling; XML; Mixed Methodologies; Text-Mining; Alienation; C-PTSD

Cet article, qui a été premièrement présenté à CSDH/SCHN 2018, décrit une méthodologie mixte qui marie des techniques numériques de la fouille de texte, tels que le marquage XML et la modélisation thématique, et l'auto-ethnographie. Dans cet article, je décris ma relation intime et difficile avec mes données, ce que je nomme ma « vie en tant que texte ». En tant qu'une personne ayant le Syndrome de stress post-traumatique complexe, j'ai d'énormes difficultés à contextualiser d'une façon situationnelle et linéaire beaucoup des évènements de ma vie. Le processus de marquage formel de mes données m'exige de surmonter des blocages psychologiques qui ont été créés comme mécanismes de défense, me permettant de lire honnêtement et de développer des associations dans ma « vie en tant que texte » dissociée. Effectuer une macroanalyse de données, qui ont été récoltées

à travers la tenue d'un journal, ainsi qu'à travers l'enregistrement et la transcription de vidéos « jouons » ("*let's play*"), m'a offert une nouvelle perspective par laquelle je me considère et par laquelle je considère mes expériences. Je me vois dans mes données. Je me vois d'une façon dont je ne trouve pas les mots pour le décrire. La fouille de textes de mes souvenirs et de mes émotions les plus profondes me fournit une sensation étrange de stabilité au sein d'une gamme de mots qui est chaotique en apparence. Dans cet article, j'explique mon parcours auto-ethnographique et l'impact important qu'a eu la fouille de textes sur ma relation avec moi-même et avec ma recherche.

Mots-clés: Autoethnographie; Modélisation de sujets; XML; Méthodologies mixtes; Exploration de texte; Aliénation; SSPT-complexe

My tears and fears: Trauma, C-PTSD, and alienation

Before I begin tackling the digital methods of XML tagging, text-mining, and topic modelling, I need to start with an introduction to the life that will be examined in this paper, mine. I have experienced a great deal of trauma in my life! Immediately, this statement is wrought with difficulties. How does one describe the experience of trauma? How is trauma defined? How is it quantified? Trauma can be physical, as occurs through an injury. I have personally experienced seven concussions, two emergency surgeries, and countless cuts and bruises. Trauma can be social, as described by Jeffrey Alexander (2012) who developed a sociological model of trauma that emphasizes its collective nature and cultural transmission. Trauma can be a discourse that explores issues around collective memory, bearing witness, politics and power, and shattered narratives, as explored by Dori Laub and Andreas Hamburger (2017).

Having formally researched trauma since 2014, I continue to struggle to define trauma in a way that can be neatly summarized in an introductory paragraph, such as this. In reality, such a feat would be impossible. Trauma can be physical, emotional, sexual, social, political, cultural, shared or individual. It can appear in one or all of these forms depending on who is telling the story. Whomever I choose to cite, I acknowledge that I am doing an injustice to the field of trauma studies as a whole.

I have therefore chosen to conduct an autoethnographic project that focuses on my own subjective experiences of trauma. My research explores my "life-as-text", also referred to as "life narrative" by Smith and Watson (2010), in order to develop grounded theories connecting a relational theory of alienation, developed by Rahel Jaeggi (2014) and discussed below, to my personal struggles with complex post-traumatic stress disorder (C-PTSD). My research is primarily exploratory though it is guided by an informal hypothesis that techniques for the mediation and resolution of alienation may be found in the ways that I have mediated and partially resolved my own C-PTSD.

In order to ethically protect the people in my life, I am choosing to refrain from specific descriptions of traumatic incidents in this paper. Instead, I will draw your attention to the *CDC – Kaiser Permanent Adverse Childhood Experiences (ACE) Study* conducted between 1995 and 1997 by Felitti, Vincent, Robert Anda, Dale Nordenberg, David Williamson, Alison Spitz, Valerie Edwards, Mary Koss, and James Marks. The study involved conducting a survey regarding childhood abuse and neglect, the results of which were then connected to health, risk behaviour, and disease in adulthood (Felitti et al. 1998). Important for our purposes is the continued research that has been conducted by the Centers for Disease Control and Prevention, as well as academic researchers (Waite 2019), utilizing adapted versions of the original ACE questionnaire. These studies reinforce the validity of the ACE questions in measuring the impact of adverse childhood experiences.

In a 2010 report, for example, Bynum et al. (2010) administered an 11question version of the ACE questionnaire to 26, 299 adults across five states. The 11 questions covered incidences of verbal, physical, and sexual abuse as well as the relationship status of parents, substance abuse by parents, exposure to mentally ill household members, and witnessing domestic violence, all before the respondent turned 18 years old. The study found that 40.6% of respondents received an ACE score of 0, 22.4% a score of 1, 13.1% a score of 2, 8.8% a score of 3, 6.5% a score of 4, and 8.7% a score of 5 or more. The higher the ACE score, the more exposure the individual had to a diverse range of adverse childhood experiences, which correlates to higher

incidences of health, risk behaviour, and disease in adulthood. After taking a self-report, 10 question version of the ACE questionnaire which can be found on the news site ACESTooHigh (Stevens n.d.), my ACE score is 8. When considering the severity of my own score, please note that my primary reason for mentioning it here is to concretize the intensity and diversity of trauma that I faced growing up. It is the duration and intensity of this trauma that led to my diagnosis of C-PTSD.

As the name suggests, C-PTSD, has left me suffering with long-term symptoms of Post Traumatic Stress Disorder (PTSD), such as flashbacks, nightmares, a heightened startle reflex, and memory difficulties related to the traumatic events. The addition of the word "complex" describes the often difficult to articulate core psychological changes that occurred as a result of long-term exposure to trauma. The following description of people who suffer from C-PTSD is the one that hits closest to home for me:

> Severe and long-standing trauma introduces a profound destabilization in the day-to-day existence of many victims…They have symptoms that alter their perceptions of their environment, disrupt their cognitive functioning, *and interfere with a sense of continuity in their lives* [emphasis added]….They experience self-hate and self loathing and feel little kinship with other human beings. They long for a sense of human connection but are profoundly alone, regarding other people with great mistrust and suspicion. They want to feel understood but cannot even begin to find the words to communicate with others about their most formative experiences (Chu 2011, 18).

Destabilization is the most difficult thing to deal with in my life. I have spent my entire life attempting to find a stable footing, a foundation upon which I could build a "normal" life. I have lived, and continue to live, with constant feelings of unpredictability, chaos, and lack of control. Even now, as a PhD candidate at the age of 39, I continue to search for stability. Ultimately, I understand that I may never find the foundation that I am lacking, and I am learning to embrace my destabilization as a part of who I am. I have also learned to identify and accept my fairly consistent

sense of dissociation which can be best described as a feeling of being on the outside looking in.

While the destabilization leaves me feeling out of control, dissociation leaves me feeling distant, separated, and alienated. James A. Chu (2011) describes pathological dissociation as a discontinuous sense of identity resulting from disconnected state changes. Everyone experiences a wide range of state changes daily as they transition between various social roles (that of worker, parent, friend, etc.). While people usually experience these state changes as fluid, unconscious transitions between various parts of a stable identity, dissociation causes me to experience these state changes as a disruptive and uncomfortable process requiring a lot of conscious effort. I am left feeling alienated, somehow separated from myself as I constantly work to adapt to my social and physical environments.

Due to the fact that I often describe my own symptoms as alienating, I am exploring the connection between my own experiences with trauma and C-PTSD, and social experiences of trauma and alienation. While there are a multitude of definitions and theories about alienation throughout the literature, I am focused on the definition provided by Jaeggi (2014) who, after careful consideration of the history of alienation theory, concludes that alienation can ultimately be defined as a "relation of relationlessness" (893). Jaeggi clarifies this definition as follows:

> If, as I have claimed, alienation is a relation of relationlessness rather than a mere absence of a relation, then giving an account of this relation will be especially complicated: as clear as it is, on the one hand, that we are somehow antecedently connected to ourselves, it is, for precisely this reason, just as unclear how it is possible for this connection to break down....The thesis, then, is that a relation to oneself is a relation that can be disturbed in various ways. And in this relation, too, we find...we are alienated from something that is simultaneously our own and alien, we are involved in relations in which we alienate ourselves, we are in a certain sense at once perpetrator and victim (Jaeggi 2014, 1352).

Jaeggi's conception of alienation is that it is about deficient relations, as opposed to non-existent ones. Alienation from the self, from others, and from the environment, is not a lack of a relation to these things, but an explicit relation that is characterized by an inability to relate. The resolution of alienation, for Jaeggi, involves appropriation: regaining and reworking the relations between the self and the self, the self and the social world, and the self and the material world. To me, C-PTSD also represents a relation of relationlessness, to my inner child, to my memories, and to my daily experiences, through dissociation and destabilization.

Autoethnography

Before moving into the discussion of tagging and text-mining my above outlined tears and fears, as promised in the title of this paper, I want to briefly outline why I felt compelled to add digital text analysis methods to a methodology that is, at its core, subjective and personal. My own experiences with trauma, which have resulted in consistent feelings of destabilization and dissociation, have severely disturbed the relations that I have with myself in many ways. Conducting my autoethnographic research has, as a result, been simultaneously highly therapeutic, in that it has allowed me to take control over the fear of my memories, and extremely difficult, in that I often struggle to solidify specific memories into articulable narratives. As a research methodology, autoethnography, perhaps more than any other, requires an explicit recognition of the deep subjectivity involved (Ellis, Carolyn. 2004). Audrey Giles and DJ Williams (2007) note that the method is a blending of fact and fiction that draws from multiple layers of consciousness and is always partial and situated. Norman Denzin (2014) discusses the multiple layers of narratives that must be acknowledged in autoethnography:

> Stories within stories told to groups remind us that every life story is a multiplicity of stories that could be told. There is no single life story or self-autobiography that grasps or covers all that a life is for a person. There are only multiple stories that can be told. Each storyteller can only tell the stories his or her biography allows to be told (Denzin 2014, 55).

Attempting to reconstruct and tell my own biography through the lens of C-PTSD introduces an additional layer to the already layered stories of my life. Dissociation disrupts my memory of events, people, places, and temporalities. The goal of my research is to draw a connection between my C-PTSD and the relation of relationlessness that is alienation, but it is the inherently alienating nature of my trauma that stands in my way. When I recall an event and attempt to write it down, I consistently have to fight against my immediate emotional reactions of fear and shame, as well as my memory of events as clouded by dissociation. I have multiple stories to tell, but have a difficult time situating myself within them.

Fortunately, research has been done, and theories developed, that led me to understand that my trauma may not be "standing in my way" at all. Ron Eyerman (2013) argues that a close reading of the trauma behind social and political theory can reveal a hidden world of meaning through an acknowledgment of "sudden breaks, lapses and absences" (50). This has led me to realize that in order to honestly and accurately develop social theory through my own subjective lens, I need to acknowledge my own breaks, lapses, and absences which have become a significant part of my "life-as-text" due to dissociation and destabilization. Geoffrey H. Hartman (1995) also seeks unpredictability and ruptures in literature as noted in the following particularly poetic and relevant quote:

> But I share with trauma studies a concern for the absences of intermittences in speech (or of conscious knowledge in speech); for the obliquity or residual muteness of "flowers of speech" and other euphemic modes; for the uncanny role of accidents; for the "ghosting" of the subject; for the connection of voice and identity; for interpretation as a feast not a fast; and for literature as a testimonial act that transmits knowledge in a form that is not scientific and does not coincide with either a totally realistic (as if that was possible) or analytic form of representation (Hartman 1995, 552).

Trauma creates ruptures in lived experience, in meaning development, and in identity formation. My biography contains more intermittences, accidents, breaks, lapses,

absences, and self-ghosting than it does coherent narrative. Acknowledging this led me to realize that I would be doing an injustice to my autoethnographic research if I did not somehow incorporate the parts of my biography that seemed inaccessible to me. This is where the XML tagging, text mining, and topic modelling came into play.

In order to find that which seemed inaccessible in my biography, I needed to come up with a way to both expose, and then analyze, that which remained lost in my dissociation and destabilization. I came up with a two step process. Step two involved the macroanalysis techniques just mentioned. As Matthew Jockers (2013) states: "[t]he macroanalytical approach reveals details about texts that are, practically speaking, unavailable to close readers of the texts" (Jockers 2013, 26). I understood that due to my C-PTSD and traumatic relationship with my life narrative, I would likely dissociate from acknowledging certain themes and patterns during qualitative coding (microanalysis). I wanted to see if text mining and topic modelling would reveal themes and patterns, breaks, lapses, and absences, about my biography that I would consciously or unconsciously avoid when approaching my "life-as-text". Before I could conduct an analysis, I needed something to analyze. Step one therefore involved expanding my data gathering beyond the limits of writing directly from memory. I will explain step one in more detail below before describing the results of step two.

Step one: Constructing my life narrative/life-as-text

This is a good place to note that I completed an application for ethics approval from the University of Alberta research ethics office and was granted approval on June 5th, 2017. Although it may seem harmful to refer to myself as a participant or object of study, I felt compelled to write out an ethics application as a form of contract between myself as a "researcher" and myself as a "participant" in order to reinforce the fact that I took all of the necessary steps to ensure my psychological safety. These steps include(d): continuing to attend therapy, continuing to take my medication, and agreeing to remain in control of what I disclosed. The last step was the most important as I had/have to maintain a balance between what is useful to include/analyze, and what is doing more harm than good. Additionally, I am highly

aware of the need to protect the people in my life who may be harmed by what I publish: the ethics contract covered this as well.

As I began to gather data, I employed three different techniques in order to allow myself to piece together my life narrative. As Sidonie Smith and Julia Watson (2010) note: Both memoir and autobiography are encompassed in the term *life writing*. We understand *life narrative*, by contrast, as a general term for acts of self-presentation of all kinds and in diverse media that take the producer's life as their subject, whether written, performative, visual, filmic, or digital (Smith and Watson 2010, 4). First, I wrote digital journal entries (private blogs) that reflected on: life events that occurred during my research, for example returning to my home town for my grandmother's funeral; nightmares and night terrors that I had; analyses of activities I engaged in (video game playing was a large part of it as you will see below) and academic literature I read; and random childhood memories. The blogs resulted in an informally organized narrative that allowed me to examine the development of my thoughts and theories as the project evolved. I did not write the entries at predictable intervals, but instead wrote whenever inspiration or need struck. I wrote a total of 35 official journal entries which amounted to a total of 38,214 words.

The second technique that I used was intended as more of a "traditional" ethnographic method. I carried a notebook with me, basically everywhere that I went, which allowed me to take field notes. I was able to record observations of events, thoughts, symbols, places, and social interactions as they occurred. I actively noted important details such as time, date, location, and my emotional state of mind in an attempt to discover my own breaks, lapses, and absences in daily life. While the journal writing allowed me to interpret and elaborate on various thoughts and experiences, the field notes provided me with context during episodes of dissociation or ongoing traumatic triggers. One common symptom of C-PTSD is memory problems due to dissociation. The field notes reminded me of where I was, who I was around, and what I was doing in order to compliment the blogs, and to guide my autoethnographic writing.

Context is highly important for an autoethnographic project as noted by Giles and Williams (2007) who argue that authors must "provide sufficient social context, move forward and backward in time, and inward and outward regarding the deeply personal and the social" (191). I therefore used my field notes to clarify and add content to my journal entries. For example, my grandmother died during my data gathering and I attended her funeral, which also involved visiting my father's grave for the first time since his passing in 2004. I was flooded with intense and overwhelming emotions and experiences from my childhood. My family left me alone as I stood in the rain, crying, and talking to my father's headstone. I talked to him about many things, including how much I missed playing games with him. Later that afternoon, I took some brief field notes about gaming with my father and how much of an impact the activity had on my life. Those notes were then incorporated into a set of journal entries that centered around how playing video games with my father, uncles, and cousins, made me "feel human" and "safe" when I was younger. The notes provided me with access to deeply personal emotions and reactions that might not have made it into my "life-as-text" otherwise. I chose not to include the field notes in an analyzable corpus of text as all important contextual and textual information was reproduced in my blogs. Instead, I have used the notes during the qualitative analysis of my blogs, and the overall analysis of my data, to aid with the narrative flow of my autoethnography as a whole.

Finally, and perhaps most importantly, I utilized web cameras and game-play recording software to create a series of videos documenting my reactions during video game play. I refer to this method of data gathering as the "let's play" which is a term that describes a player recording their own reactions and commentary while playing a video game (Sapach 2018). Let's plays are extremely popular online and have wide audiences on both YouTube and Twitch. I have personally invested hundreds of hours into watching let's plays by some popular YouTube stars like Jacksepticeye (Wikipedia 2018) and Markiplier (Wikipedia 2018). I am therefore very familiar with the technique, descriptions of hardware setups, and ability for the process to capture unique, unexpected, and often entertaining reactions to video game play. Unlike the

traditional let's play, I chose not to publish my videos online due to the sensitive and deeply personal nature of much of the content.

At this point, I should explain why I only recorded myself playing video games. My earliest childhood memory involves playing a video game. The vast majority of my positive social, and individual, experiences during the first 16 or so years of my life involved playing video games. My autoethnographic project is part of my larger dissertation that also argues that video games, and video game culture, have unique characteristics that helped me to mediate and partially resolve my C-PTSD. I feel comfortable playing video games. I feel more open to experiencing emotion when playing video games. I also find that my memory is highly connected to visual and spatial cues. I chose to record myself playing video games from throughout my childhood with the hope that I would be able to trigger memories through not only the visual and spatial feedback from the game, but also through hearing the music and sound effects, and physically holding game controllers and performing actions as I did during my childhood. Interestingly, there is research that argues for the benefits of video game play in relation to PTSD, specifically around the game "Tetris." (Holmes et al. 2009; McGonigal 2015) I will not get into it here, but I highly recommend it for anyone interested in how games can help prevent harmful flashbacks.

Returning to Jaeggi (2014), the resolution of alienation comes through a process of successful appropriation. In order to develop/repair relations of relationlessness with the self, appropriation needs to be an active and involving process of integration and transformation of meaning and experiences. Recording myself playing video games allowed me to record the active process of exposing myself to triggering memories while utilizing the benefits of the games for both memory work, and appropriation work. My goal was to observe myself during moments of trauma recollection and trauma mediation through game play. This self-observation through a medium different from written text allowed for a more detailed creation of my life narrative.

In total, I recorded 39 videos that are included in my dataset. I actually recorded many more videos; however, they either focused entirely on my hardware setup and acted as test videos, or I chose to exclude them due to their content as per my

ethics agreement with myself. I only ended up excluding one video due to traumatic content, but I did make note of the fact that I was silent and nearly catatonic for over 75% of the video. I later blogged about the experience in order to preserve the trigger that led to the extreme dissociation. There are a total of 24 hours and 34 minutes of video that I eventually transcribed manually, resulting in a 100,038-word corpus of analyzable video data.

There is a speculative question that lies at the heart of this project; namely, how can collective play and participation in video game culture help us cope with, understand, and in some ways overcome, trauma and alienation? When gathering data, I actively kept this question in the back of my mind. A large part of my dataset, therefore, is comprised of descriptive personal narratives exploring how video games and video game culture impacted my experiences with trauma and coping with my C-PTSD. I intentionally triggered memories and emotional reactions related to game play experiences, game culture related symbols, and social interactions around video game culture. An unexpected result of recording the let's plays was the amount of current, real life trauma that I described. Many of the videos blended childhood memories with things I was experiencing and feeling as the project progressed. This has resulted in an extremely personal corpus of data that not only explores my childhood trauma, but also provides important insights into how video games, and my academic study of them, currently works to mediate my C-PTSD symptoms.

Step two: Data analysis techniques

One aspect of my data analysis involved a qualitative open and axial coding of all blogs, field notes, and let's play video transcriptions. I conducted this coding before running any digital text analysis in order to determine what themes and patterns I could derive from the data. As an exploratory study based in grounded theory, this coding was highly personal and relatively informal. A detailed analysis of my findings is not appropriate for this paper; however, I will say that the process was an emotional and often triggering one. For example, each time I read the name of a particular person who caused me a lot of harm, I noticed myself skimming through the section as quickly as possible. My reaction to reading my own recollection of my trauma was often dissociation in the form of avoidance. This difficulty, which I did

mediate through controlling the amount of time I would allow myself to read each session (no more than 30 minutes) reinforced my anticipated need for an alternative method of analysis.

I should note here that my video transcriptions are somewhat unorthodox. I chose to manually transcribe each video as opposed to using a speech-to-text program as I wanted to do a detailed content analysis. Having said that, I also decided against following formal multimedia/video content analysis techniques, such as those found in Ohm (2016) and Krippendorff, (2019). My reason for this choice primarily involves the highly intimate nature of the process. As an exploratory study, I did not want to overwhelm the process with formalized annotation techniques. I ended up watching each video three times. The first time, I took brief notes about the general content and themes and my overall demeanor. The second time I meticulously typed out each word I spoke, including words like "um" which ended up being my most frequently used word. During this second viewing, I also made notes related to mumbling and stuttering by using three asterisks followed by the description followed by three more asterisks; for example: "um um I am talking ***mumble*** I th think ***stutter***." I wanted to have a clear way of making note of my breaks and verbal intermittences while being able to quickly identify them during qualitative coding. The third viewing involved me taking note of emotional reactions, long pauses in speech, and obvious signs of dissociation. For the pauses, I made side notes about how long I sat in silence for. For the emotions and dissociation, I coded them as I perceived them which is another reason that I did not follow any formal guidelines. My C-PTSD alters how I express emotions in various situations. A single tear can represent anger while a smile can represent shame. As an intimately subjective process, I felt that using my own transcription techniques would provide me with the best information to accurately transcribe my life narrative into my "life-as-text" for digital analysis.

XML tagging

As this is a paper about an exploratory project aimed at developing grounded theory, it is important to address the things that both worked, and did not work, in allowing me to find the patterns and themes that qualitative coding missed. XML tagging is one of the things that did not work for me…as it was intended to work. Admittedly

a novice in this area, my original plan did not include XML tagging as an option. During a discussion with one of my dissertation supervisors, we began contemplating the best ways to organize my life narrative into a cohesive, easy to navigate, format. He suggested that I incorporate structured, hierarchical metadata into my corpus through the use of XML. Having worked with HTML previously, I did not anticipate a problem with the learning process.

I installed *Oxygen XML Editor 19.1*, and began researching the best way to tag my transcripts and blogs. I read through the book *XML Pocket Reference* 3rd ed (St Laurent and Fitzgerald 2005) and discovered that *RELAX NG* seemed to be the appropriate schema to utilize due to its simplicity and flexibility. I began by following a simple hierarchy, including the use of a <head> element which contained the date of the entry as well as the time of recording and video game information for video transcriptions (see **Figure 1**). I then grouped paragraphs thematically. For video recordings, I used the category "setup" when I was talking about recording issues,

```
<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="schema.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<transcription game="the adventure of link">

    <head>

        <date when="2017-02-15">
            <year>2017</year>
            <month>02</month>
            <day>15</day>
        </date>

        <time>
            <hour>12</hour>
            <minute>04</minute>
        </time>

        <biblgame>
            <game name="the adventure of link">Zelda 2: The Adventure of Link</game>
            <developer>Nintendo Co. Ltd.</developer>
            <publisher>Nintendo Co. Ltd?</publisher>
            <releasedate>1987-01-14</releasedate>
            <platform>NES</platform>
        </biblgame>

    </head>

    <body>

    <p category="setup">
        Ok so I have commentary on, it seems to pick up the same in game sound as before. This is <game name="the ad
    </p>

```

Figure 1: Example of XML tagging for a video game recording transcript using a basic RELAX NG schema with personalized tags.

"description" when I was talking about the video game being played, "memory" when I was talking about memories, "methodology" when I was breaking the fourth wall and talking about the process of recording as a research method, and "interpretation" when I brought up academic theory or insights about my experience replaying the game. Figure one displays the general hierarchy and the header tag.

As I worked through each transcript and blog, I began adding tags to people, pets, games, and places. I also began tagging actions like sighing, reactions like laughing, pauses, stutters, and emotions. I attempted to categorize each emotion as either positive or negative in anticipation of conducting a sentiment analysis. I began to realize, the more I drifted away from a formal schema and into categorizing each word as I interpreted it, that I was second guessing my tags. Many of my "emotion" tags, for example, <emotion type="negative">dissociation</emotion>, were actually descriptions of C-PTSD symptoms. My tagging began to emulate my qualitative open coding in that I found myself searching for themes and patterns where I should have been focused on providing simple metadata.

After spending countless hours tagging, and re-tagging documents, I made the decision to stop after only completing work on 29 out of 73 files. I was getting lost in attempting to categorize my life narrative. The process itself was very insightful as it allowed me to explicitly acknowledge my breaks, lapses, and absences through the tagging of pauses and stutters. It allowed me to visualize the connections between my pauses, stutters, games, people, and emotional (symptomatic) states. XML tagging brought me face-to-face with things that I might otherwise ignore or shut-out due to my C-PTSD. Ultimately, the process became more harmful than beneficial in that I was "forcing" myself to attempt to categorize my symptoms as emotional responses. In my lived experience, the primary emotion I can genuinely identify is fear. I do feel things like love, happiness, sadness, and anger, but they are more often than not accompanied by some sense of fear, or shame. My trauma created this confusion of emotions, and I have learned to live with and navigate them. When conducting XML tagging, I was asking myself to pick an emotion, and concretize it as either positive or negative. I eventually acknowledged that I was incapable of doing it, so I stopped.

Before moving on, I would be doing a disservice to this article if I did not mention the following research. My focus on attempting to categorize my emotions using XML as a precursor to potential sentiment analysis stemmed from work that has been done on utilizing text-analysis and sentiment analysis to identify PTSD related information for the purposes of identifying potential PTSD sufferers. Vadim Kagan, Edward Rossini, and Demetrios Sapounas (2013) gathered PTSD data from two distinct corpra: "(1) clinical documentation describing all aspects of the psycho-physiological disorder and (2) forum discussions authored by individuals suffering from the disorder or their friends and family." (Kagan, Rossini, and Sapounas 2013, 24) They then had experts score the documents providing a human annotation training set on which the software was trained, after which they determined:

> The results of processing validation data sets (distinct from the training set mentioned above) through the software demonstrated the high accuracy of the automated analysis results are closer to those of the psychologists. The agreement between the automated analysis results and those of human experts ranged between high 70 and high 80% – a very good result for any text analytics study, and an excellent one for a project investigating something as complicated as psychological signals (Kagan, Rossini, and Sapounas 2013, 81).

They then moved to user testing with a small set of users and further demonstrated that text analysis of structured questionnaires through automated algorithms could provide a fairly accurate indicator of the presence of PTSD.

I was inspired by the research and had hoped to find indicators of C-PTSD through an analysis of my own "life-as-text." Not in a formally scientific way as the above study did: my project is not a formal scientific study. What I had hoped was that there would be patterns to my emotions that might provide more insight into my own symptoms. As I noted however, as I began confounding my symptoms with emotions, I realized that the "complex" part of C-PTSD is thus named for a reason.

I still find the above study fascinating and will pursue extension of the results to C-PTSD in the future.

Python/NLTK/R

I originally began my descriptive analysis using the *Python* NLTK library but I gradually transitioned over to *R* largely due to the fact that I was required to learn and work with *R* for an unrelated project. I will first describe the interesting issue around the use of stopwords I encountered while analyzing word frequencies using *Python* followed by a discussion about dictionary/sentiment analysis conducted using *R.*

My initial descriptive analysis using *Python* provided one significant question that gave me a lot to think about. How do I deal with the issue of stopwords in my "life-as-text?" This question stemmed from the above discussed XML tagging and qualitative coding. I realized that as an exploratory autoethnography, my goal was to understand my corpus in relation to how I experienced my life, my trauma, and my alienation. With the understanding that standard practice is to exclude stopwords from natural language processing (Bird, Klein, and Loper 2009), I cleaned my corpus accordingly. Upon running my word frequency analysis, I realized that the word "I" was excluded as it is one of the stopwords in the NLTK library. As I was analyzing a first-person life narrative, I was dissatisfied with the exclusion of such an important word, so I reran the analysis on a version of the corpus with the majority of the stopwords reincorporated. I made the choice to exclude the words "the" and "and" as I was not particularly interested in exploring their frequency and had found no significant patterns of their usage in my qualitative analysis. Again, as an exploratory study, I was fundamentally playing with my data to see what would happen. **Figure 2** shows the frequency distribution of the reintroduced stopwords only as I was interested in understanding how often "I", and other singular first-person pronouns, actually appeared.

The fact that the word "I" appears 7,129 out of 128,252 times is not overly surprising. And yet, for an autoethnography, I actually expected the word "I" to occur more than 5% of the time, as I not only spoke in the first person during my videos but also wrote in the first person in my journal entries. In order to compare my results to another autoethnography out of curiosity, I used *Voyant Tools* (voyant-tools.

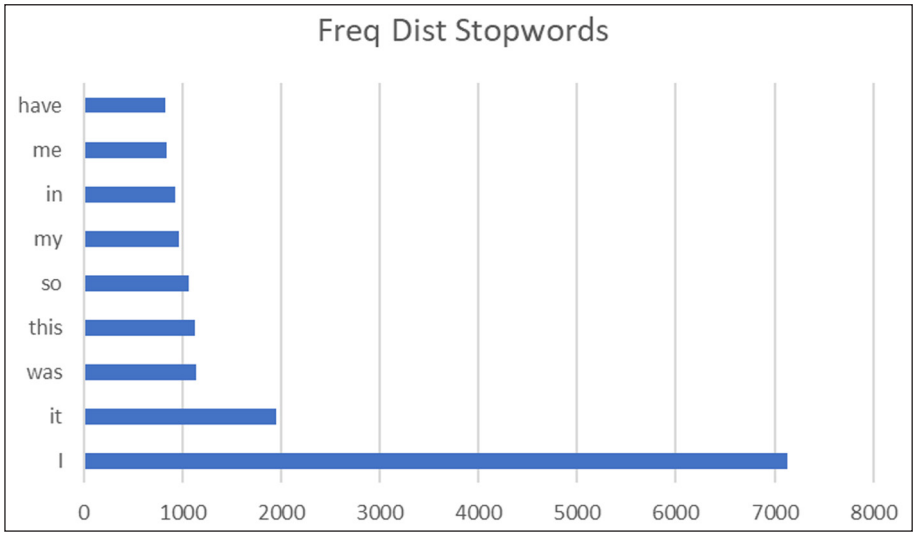

Figure 2: Full corpus stopwords frequency distribution. This figure excludes the words "the" and "and" with the goal of focusing on singular first-person pronouns.

org) to quickly analyze the appearance of the word "I" in the autoethnographic PhD dissertation of a colleague (Vossen 2018). The document contained 111,216 words, 1499 (1.35%) of which were "I". While a more formal analysis containing a significant number of samples would provide a stronger comparison, it would seem that at 5.56% of my word count, the word "I" does say something important that needs to be investigated further. I hypothesize that it has to do with the emotional intensity of many of my blog and video entries. I refer to my needs, my shame, and my trauma frequently. Recording the videos provided me with an outlet to say what was on my mind in what I perceived to be a safe environment engaged in the safe activity of video game play.

Figure 3 on the other hand shows a frequency distribution with all of the standard NLTK library stopwords removed. How much more or less do these words tell me about my "life-as-text" with the singular first-person pronouns removed? Out of context, the most frequent words make no obviously decipherable sense. With the exception of the word "game", none of the other words necessarily speak to the focus of my dissertation. There are no particularly traumatic, emotional, or alienation-related words and the most frequent word isn't really much of a word at all. That being

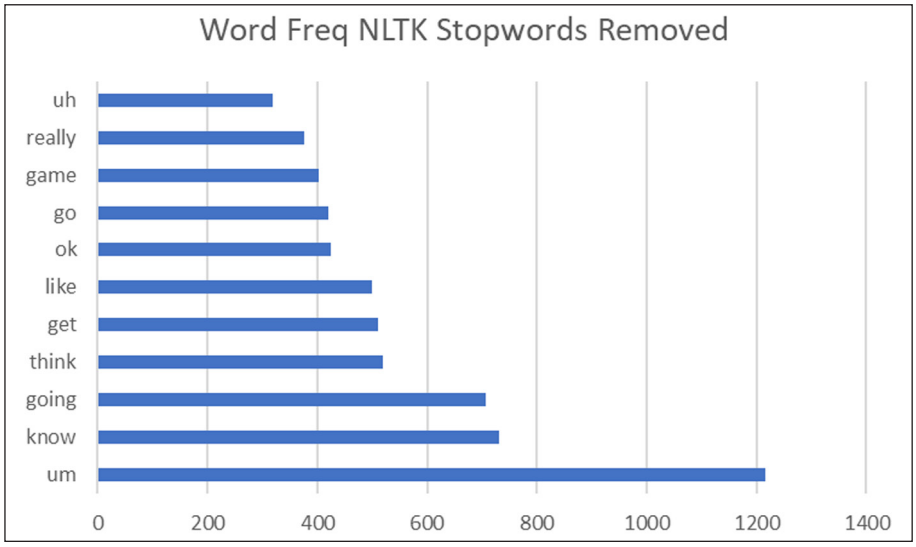

Figure 3: Corpus frequency distribution with NLTK standard stopwords removed.

said, I have stated above that a driving force behind my use of computer mediated text analysis was to seek out and acknowledge the breaks, lapses, and absences in my "life-as-text." Put into context, my XML tagging and qualitative coding indicated that the word "um" most often occurred before a significant pause, right around the time that I seem to begin experiencing some sort of emotion or connection to my trauma. These significant pauses are usually accompanied by a switch to an intense focus on whichever game I happen to be playing. XML tagging also seemed to suggest that I often switch to talking about in-game events after my "um's" and long pauses, indicating that I am using the game to re-focus and remove myself from the traumatic feelings associated with recollection. I am fascinated by the discovery of a pattern of behaviour that I was not consciously aware of prior to this research. My use of the word "um" as an indicator that I am becoming emotionally uncomfortable or unsafe is a result worthy of deeper analysis in my autoethnography.

Moving away from word frequencies, I was curious about the appearance of categories of words as found in the Harvard IV-4 and associated dictionaries (2020). Using *R*, I once again cleaned my corpus of data and divided it into two sections: blogs and games (representing my journal entries and video recording transcriptions

respectively). I then wrote basic code that would take one of the dictionaries which I had downloaded into a simple .txt format and run what basically represented a sentiment analysis, comparing the dictionary to my corpus and providing me with a mean score indicated the frequency with which words in the category appeared in each section of my corpus. The following code, for example, analyzes the pain category in comparison to my corpus:

```
#pain wordslist
# -> 254 words indicating suffering, lack of confidence, or commitment.
lexicon.pain <- scan('pain.txt', what='character', comment.char = ';')
pain <- lapply(bag, function(x) {sum(!is.na(match(x, lexicon.pain)))})
painscore <- unlist(pain)
mean(painscore)
```

I ran this process using 17 of the dictionaries that I felt were most relevant to my research. Again, as an exploratory project, I did this to see if I could find anything that stood out. I created boxplot visualizations in order to better understand the differences between my blogs and video transcripts for each of the dictionaries and three stood out to me. The first and second were the pain and pleasure categories. **Figure 4** shows the side-by-side boxplots demonstrating the frequency with which I used words related to pain in my blogs and video transcripts and **Figure 5** does the same with the pleasure category. On the surface, it makes sense that my game playing was more highly associated with the pleasure category while, with the exception of a few outliers, the blogs were more highly associated with pain. In fact, out of all 17 dictionaries, pain is the only category that had a mean score so close between the blogs and the video recordings. The third category that stood out was the one related to negativity shown in **Figure 6**. When thinking about trauma and C-PTSD, I expected myself to talk more about painful experiences during my unedited video recordings, particularly when triggered by playing games from childhood. On the other hand, I expected my video transcriptions to contain fewer negative words than my blogs as the majority of my blogs focused on negative experiences.

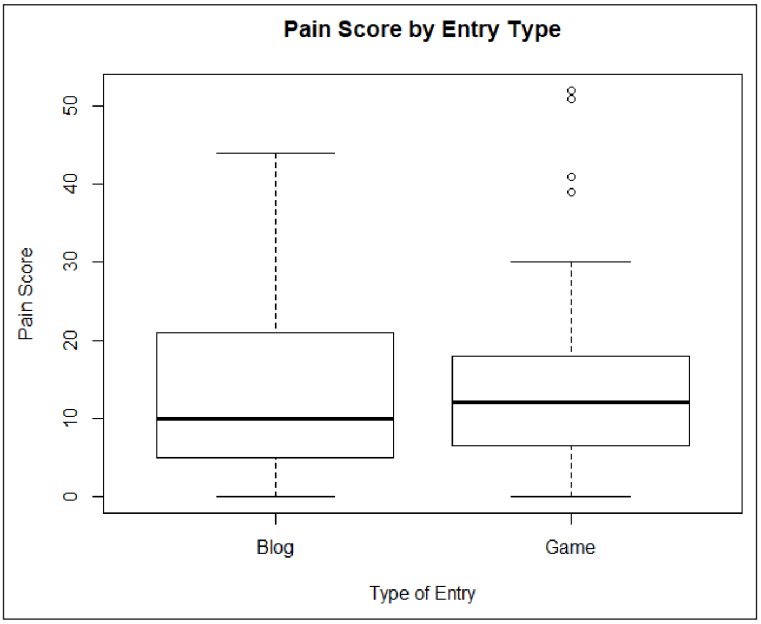

Figure 4: Pain category distribution comparison between blogsand video game play transcriptions.

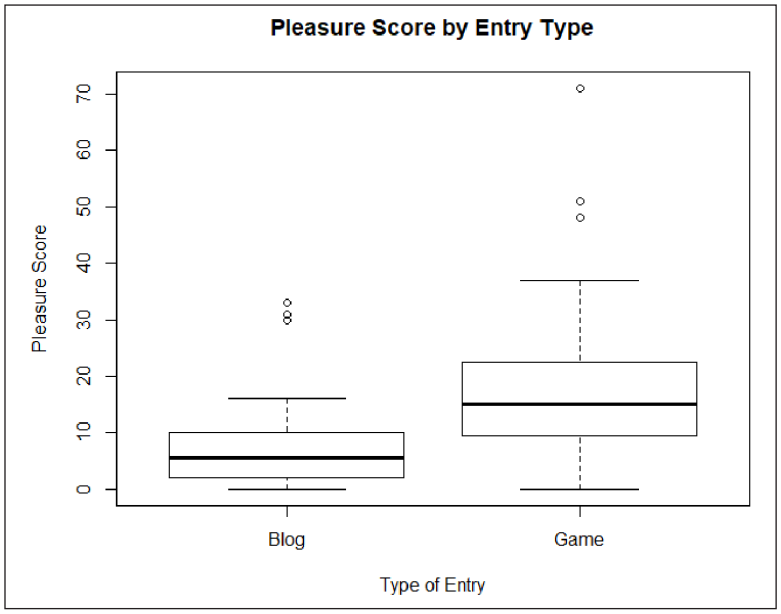

Figure 5: Pleasure category distribution comparison between blogs and video game play transcriptions.

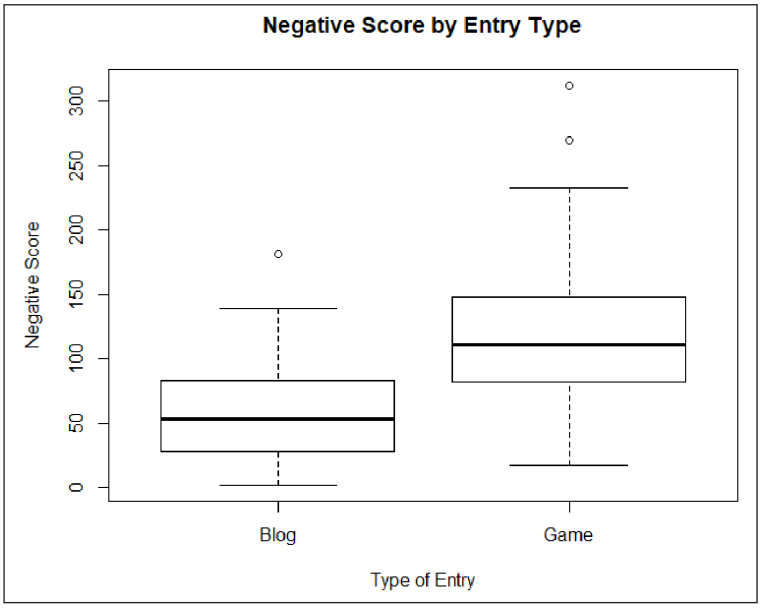

Figure 6: Negative sentiment category distribution comparison between blogs and video game play transcriptions.

At this point, I hope you are asking yourself what the point of this analysis was, especially considering that I could have cleaned things up by removing extreme outliers. Why the boxplots and the vague categories and explanations? For me, it was primarily about the therapeutic nature of the results. Throughout this discussion I have used the word "expected". I expected the word "I" to appear more frequently. I expected more mentions of pain during video game play. These expectations speak to the complex part of my C-PTSD and to the relation of relationlessness I have with myself. I tend to see myself through lenses that are not necessarily accurate. Taking the time to visualize my use of different categories of words allowed me to understand that I do, for example, express pleasure during pleasurable activities. That I do have pleasurable memories related to gaming during a traumatic childhood. That I can articulate pain through the written word when I am going through tough times like my grandmother's funeral. More than anything, this project has shown me that there is therapeutic value to taking the time to examine traumatic experiences through both micro and macro lenses. The computer, acting as a mediator, presenting me

with perspectives of my life narrative that I would otherwise miss or ignore. This leads me to a discussion of my final analysis technique, topic modelling.

Topic modelling

I decided to run topic modelling on my "life-as-text" primarily based on the following description by Jockers (2013):

> Aside from the researcher's somewhat arbitrary setting of the number of topics to be "discovered," the entire process is done in an unsupervised fashion. "Unsupervised" means that the machine does not know in advance what themes to look for. With no human input into what constitutes a theme, a motif, a topic, the model collects distributions of co-occurring words and then returns them in a manner that allows us to examine, assess, interpret, and intuit what they all have in common, that is, their shared "theme" (Jockers 2013, 123).

I wanted an unsupervised, unguided perspective on the themes that might be found in my life narrative. In the previously mentioned analysis techniques I controlled the code, the visualizations, and the selection of dictionaries. I wanted to try to obtain a digital analysis of my "life-as-text" with the least amount of interference from me as possible, and topic modelling seemed like the best fit based on my skillset.

Using *Mallet* I ran an initial set of topic models, 10, 20, and 50, on the complete corpus of all of my transcribed videos and blogs. I did remove the standard NLTK stopwords, though you may notice some of them in **Figure 7**, as I struggled to

weight	Topics
0.624	don't mom feel things life back people fear make time thing afraid didn't feeling fact scared head good understand
0.55	um don't game uh sigh good yeah I've thing back lot things play sort I'll stuff guess video talk
0.291	remember games memories game video dad time playing play memory room mom didn't house played christmas brent wasn't kid
0.283	brent dog work feel morrighan edmonton sleep pain puppy candidacy house sleeping worried hearthstone writing sesu ago training supposed
0.2	alienation pain trauma relation life people social research ways real culture write theory dissociation experiences enjoy alienated identity states
0.183	day birthday mom guilt dad mother feel gift shame year email call didn't mario christmas child safe made situation
0.137	light road back feel room follow hmmm find tinder story sanity accident dreams creepy door interesting car missing photos
0.11	dad family shame grandma pain john aunt real funeral dream felt remember died left sort part odd women donna
0.088	mumble good computer guys yeah I'll yay alright ah brent played quest die character kill dragon war fun playing
0.08	world truth people post fact science sonja real gonna play facts comfort passion identity social important sort emotion fantasy

Figure 7: 10-topic model created using Mallet. Note that due to a character set conversion error, some of the NLTK stopwords remain, though most were removed.

correctly remove the punctuation due to novice mistakes around character sets. For example, the word "don't" is included, but it actually read "don" in the output. I left it in however, as it carries certain contextual meaning related to negativity. The unintentional inclusion of the words "I'll" and "I've" also reinforce the importance of singular first-person pronouns in my corpus.

There are two "outliers" in this example that stand out like sore thumbs to me. The 7th one, "light, road, back, feel, room, follow, hmmmm, find, tinder, story, sanity, accident, dreams, creepy, door, interesting, car, missing, photos" speaks to my combined playthroughs of the games *Trauma* and *Amnesia: The Dark Descent*. Most of the words describe in-game objects and activities. The 10th one, "world, truth, people, post, facts, science, sonja, real, gonna, play, facts, comfort, passion, identity, social, important, sort, emotion, fantasy" comes from a single video that I recorded as part of another project. I have included the transcription of the video in my corpus as it does contain important autoethnographic points and was technically recorded while playing a game during my data collection timeline.

The remaining 8 topics are very interesting in that they make sense to me. Each topic reminds me of a specific aspect of my life. I see myself in my data. I see myself in ways that I cannot otherwise put into words. The topics provide an odd sense of stability in a seemingly chaotic assortment of words. To be honest, the actual statistical significance of each topic is relatively unimportant to me. How I use the topics is largely related to story-telling. As I pieced together my autoethnographic narrative and connect my "life-as-text" to various themes and theories, the topic models have helped me group things together that I would otherwise have a difficult time accessing and connecting due to dissociation and the emotional attachment I have to my data. These topics encouraged me to acknowledge my ingrained tears, fears, guilt, shame, and isolation. I have been able to access and examine things that have remained hidden due to my C-PTSD. It is through the computer mediated analysis of my data that I have been able to step back from my emotional attachment and ingrained assumptions, and view the data through a somewhat logical, orderly lens: a lens temporarily cleared of the fog of trauma. The topic modelling in particular has allowed me to step back from my dissociation and destabilization, to view my experiences as connected parts of a larger whole.

Conclusion

My "life-as-text" is more than just the corpus of words that I have gathered. It is a series of interconnected narratives, experiences, and emotions. My relation to my "life-as-text" can be seen as a deficient one, an alienated one. My life narrative comes from me, it is me, and yet I feel separated from it in many ways. My C-PTSD makes me afraid of my own story. My acceptance of my own destabilization and dissociation makes the process of writing an autoethnography a monumental undertaking. I have found, as described above, that the addition of digital text analysis to my autoethnographic process has allowed me to regain and rework my relationship to myself and my narrative. By allowing me to step back and view my "life-as-text" through a somewhat objective lens, methods like topic modelling are aiding in the process of appropriation that Jaeggi (2014) argues is so central to the resolution of alienation. Recall that appropriation needs to be an active and involving process of integration and transformation of meaning and experiences. Analyzing my life narrative using digital methods has allowed me to take an active role in first understanding, and then transforming the meaning of my trauma.

Shining a light on my breaks, lapses, and absences has allowed me to better understand myself despite my C-PTSD and related dissociation. Writing the autoethnography, which in some ways is more of an autobiography, has fostered the mediation of my own alienation. By allowing me to embrace and examine my "tears and fears" through an unsupervised and, in some ways impartial, digital lens, text-mining has provided me with a foundation upon which the reconstruction and acceptance of a haunting and painful past is possible.

Competing Interest

This article draws on research supported by the Social Sciences and Humanities Research Council.

Editorial Contributors

Congress 2018 Special Editor: Constance Crompton, University of Ottawa, Ottawa, Ontario, Canada.

Section/Copy Editor: Shahina Parvin, University of Lethbridge Journal Incubator.

References

Alexander, Jeffrey. 2012. *Trauma: A Social Theory*. Cambridge: Polity Press.

Bird, Steven, Ewan Klein, and **Edward Loper.** 2009. *Natural Language Processing with Python*. Beijing: O'Reilly.

Bynum, L., T. Griffin, D.L. Ridings, K.S. Wynkoop, R.F. Anda, V.J. Edwards, Y. Liu, L.R. McKnight-Eily, and **J.B. Croft.** 2010. "Adverse Childhood Experiences Reported By Adults — Five States, 2009." *Morbidity and Mortality Weekly Report, CDC* December 17 (2010): 1609–1613. Accessed May 26, 2020. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5949a1.htm. [The authors' full names were not available online]

Chu, James. 2011. *Rebuilding Shattered Lives: Treating Complex PTSD and Dissociative Disorders*. New Jersey: John Wiley & Sons, Inc. DOI: https://doi.org/10.1002/9781118093146

Denzin, Norman. 2014. *Interpretive Autoethnography*. California: Sage Publications, Inc. DOI: https://doi.org/10.4135/9781506374697

Ellis, Carolyn. 2004. *The Ethnographic I: A Methodological Novel about Autoethnography*. Walnut Creek, CA: AltaMira Press.

Eyerman, Ron. 2013. "Social Theory and Trauma." *Acta Sociologica* 56(1): 41–53. DOI: https://doi.org/10.1177/0001699312461035

Felitti, Vincent, Robert Anda, Dale Nordenberg, David Williamson, Alison Spitz, Valerie Edwards, Mary Koss, and **James Marks.** 1998. "Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study." *American Journal of Preventive Medicine* 14: 245–258. DOI: https://doi.org/10.1016/S0749-3797(98)00017-8

Giles, Audrey R., and **D.J. Williams.** 2007. "Are We Afraid of Our Selves? Self-Narrative Research in Leisure Studies." *World Leisure Journal* 49(4): 189–198. DOI: https://doi.org/10.1080/04419057.2007.9674511

Grip, Thomas, and **Jens Nilsson.** 2010. *Amnesia: The Dark Decent* [Video Game]. Frictional Games. Microsoft Windows.

Hartman, Geoffrey H. 1995. "On Traumatic knowledge and Literary Studies." *New Literary History* 26(3): 537–563. DOI: https://doi.org/10.1353/nlh.1995.0043

Harvard IV-4 and Associated Dictionaries. 2020. *Descriptions of Inquirer Categories and Use of Inquirer Dictionaries*. Accessed May 25. (http://www.wjh.harvard.edu/~inquirer/homecat.htm).

Holmes, Emily A., Ella L. James, Thomas Coode-Bate, and **Catherine Deeprose.** 2009. "Can Playing the Computer Game "Tetris" Reduce the Build-Up of Flashbacks for Trauma? A Proposal from Cognitive Science." *PLoS One* 4(1): 1–6. DOI: https://doi.org/10.1371/journal.pone.0004153

Jaeggi, Rahel. 2014. *Alienation: New Directions in Critical Theory*. New York: Columbia University Press.

Jockers, Matthew. 2013. *Macroanalysis: Digital Methods & Literary History*. Chicago: University of Illinois Press. DOI: https://doi.org/10.5406/illinois/9780252037528.001.0001

Kagan, Vadim, Edward Rossini, and **Demetrios Sapounas.** 2013. *Sentiment Analysis for PTSD Signals*. New York: Springer. DOI: https://doi.org/10.1007/978-1-4614-3097-1

Krippendorff, Klaus. 2019. *Content Analysis: An Introduction to Its Methodology*. 4th ed. Thousand Oaks: Sage.

Laub, Dori, and **Andreas Hamburger.** (eds.). 2017. *Psychoanalysis and Holocaust Testimony: Unwanted Memories of Social Trauma*. London: Routledge. DOI: https://doi.org/10.4324/9781315717456

Majewski, Krystian. 2011. *Trauma* [Video Game]. Krystian Majewski. Microsoft Windows.

McGonigal, Jane. 2015. *SuperBetter: The Power of Living Gamefully*. New York: Penguin Books.

Ohm, Jens-Rainer. 2016. *Multimedia Content Analysis*. Berlin: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-662-52828-0

Sapach, Sonja C. 2018. "Let's Play with Research Methodologies". *First Person Scholar*, January 03. Accessed May 26, 2020. http://www.firstpersonscholar.com/lets-play-with-research-methodologies/.

Smith, Sidonie, and **Julia Watson.** 2010. *Reading Autobiography: A Guide for Interpreting Life Narratives, Second Edition*. 2nd ed. Minneapolis: University Of Minnesota Press.

St. Laurent, Simon, and **Michael Fitzgerald.** 2005. *XML Pocket Reference*. 3rd ed. Sebastopol, California: O'Reilly.

Stevens, Jane Ellen. n.d. "Got Your ACE Score?" *Aces Too High News*. Accessed May 26, 2020. https://acestoohigh.com/got-your-ace-score/.

Vossen, Emma. 2018. *On the Cultural Inaccessibility of Gaming: Invading, Creating, and Reclaiming the Cultural Clubhouse*. PhD Diss., University of Waterloo. Accessed May 26, 2020. UWSpace. http://hdl.handle.net/10012/13649.

Waite, Roberta, and **Ruth Ann Ryan.** 2019. *Adverse Childhood Experiences: What Students and Health Professionals Need to Know*. Milton: Routledge. DOI: https://doi.org/10.4324/9780429261206

Wikipedia. 2018. *Jacksepticeye*. Accessed May 26, 2020. https://en.wikipedia.org/w/index.php?title=Jacksepticeye&oldid=864087282.

———**.** 2018. *Markiplier*. Accessed May 26, 2020. https://en.wikipedia.org/w/index.php?title=Markiplier&oldid=864339332.

How to cite this article: Sapach, Sonja. 2020. "Tagging My Tears and Fears: Text-Mining the Autoethnography." *Digital Studies/Le champ numérique* 10(1): 4, pp. 1–19. DOI: https://doi.org/10.16995/dscn.328

Submitted: 18 October 2018 **Accepted:** 10 June 2020 **Published:** 31 August 2020

Digital Studies/Le champ numérique is a peer-reviewed open access journal published by Open Library of Humanities. OPEN ACCESS