A NOVEL APPLICATION OF NATURAL LANGUAGE PROCESSING FOR THE ASSESSMENT OF COMMUNICATION OUTCOMES POST-STROKE

by

Sydney S. Cadogan

May, 2021

Director of Thesis: Kazlin Mason, PhD

Department: Human Services

Program: Communication Sciences & Disorders Program

ABSTRACT:

Aphasia is an acquired neurogenic communication disorder resulting in an impairment across one or many modalities of communication. When assessing aphasia, speech-language pathologists (SLPs) utilize a blend of quantitative and qualitative measures to aid in clinical diagnoses and treatment outcomes. Analysis of discourse is a particularly important component to document language recovery. Computer aided text analysis (CATA) utilizing natural language processing (NLP) is an intersection of quantitative and qualitative research that aims to draw the thoughts and emotional attitudes from individual narratives and written texts. Due to the advancement and accessibility of software programming and computational powers, CATA has the ability to both investigate the superficial and latent semantic attributes of language embedded in a text sample. Furthermore, sentiment analysis, the automated process of deriving positive, negative, or neutral opinions from text, is one specific application of CATA. Past studies have applied sentiment analysis towards consumer-driven and marketing research. Fewer studies have researched how sentiment analysis can be applied to healthcare domains. The purpose of this exploratory study is to apply a methodology for programmatic analysis of the sentiment of transcribed post-stroke speech samples (text) and assess change over time.

A NOVEL APPLICATION OF NATURAL LANGUAGE PROCESSING FOR THE ASSESSMENT OF COMMUNICATION OUTCOMES POST-STROKE

A Thesis Submitted to the Department of Human Services in Partial Fulfillment for the Master's Degree in Communication Sciences and Disorders

Submitted by

Sydney S. Cadogan

University of Virginia

Department of Human Services

A NOVEL APPLICATION OF NATURAL LANGUAGE PROCESSING FOR THE ASSESSMENT OF COMMUNICATION OUTCOMES POST-STROKE

by

Sydney S. Cadogan

APPROVED BY:

DIRECTOR OF THESIS: ______________________________
Kazlin N. Mason, PhD

COMMITTEE MEMBER: ______________________________
Filip Loncke, PhD

COMMITTEE MEMBER: ______________________________
Randall Robey, PhD

ACKNOWLEDGMENTS

Undertaking this thesis as a compliment to my master's degree would not have been possible without the unwavering support and guidance of several key individuals, who each contributed a wealth of assistance to make this project possible.

I would first like to thank my advisor, Dr. Kazlin Mason, Ph.D., who took the spark of research interest I personally had at the start of my degree and allowed for me to truly grow and develop as a researcher throughout this entire process. Her expertise was invaluable in steering the direction of this thesis and her patient, insightful feedback pushed me to bring my work to a higher caliber. I would additionally like to thank Dr. Randall Robey Ph.D. and Dr. Filip Loncke, Ph.D., for carving time out of their own schedules to serve as committee members, as well as for continuing to support my academic and research endeavors over the years during my time as both an undergraduate and graduate student. I would also like to acknowledge project collaborator Anna Godwin, MSDS for her valuable guidance in regards in the initial stages of data analysis. Additionally, I express my deepest gratitude to Dr. Thomas Broussard, Jr., Ph.D. for his continued support and ongoing collaboration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Aphasia is an acquired neurogenic communication disorder resulting in an impairment of the expression or comprehension of language across one or many modalities of communication (Brookshire & McNeil, 2014). This disruption in the language network system of the brain results from injury to the brain such as from a stroke, traumatic brain injury, or other catastrophic or progressive neurological disease. Aphasia is most often observed as a byproduct of a left-hemispheric stroke when regions of the brain responsible for language lose their source of blood supply due to an abrupt disruption (Gerstenecker & Lazar, 2019). It is estimated that upwards of a third (225,000) of people experience aphasia as a byproduct of stroke ("Aphasia Statistics," n.d.). Aphasia symptoms can range from minor word-finding challenges to profound impacts and impairments that disrupt one's underlying rule-structure of language. These speech and language deficits are known to be observed throughout the acute, subacute, and chronic phases of recovery post-stroke (Hills et al., 2018). Aphasia diagnosis can be further classified by subtype, typically characterized by deficits in speech fluency, comprehension of spoken messages, and repetition (Brust et al., 1976). Early and consistent intervention is the recommended course of action for aphasia recovery. Speech-language therapy, carried out by speech-language pathologists (SLPs), is a primary factor in this recovery process and is aimed at identifying and ameliorating marked communication difficulties (Pauranik et al. 2019). Therapeutic interventions often focus on a patient's primary language deficit (i.e., anomia and marked word-finding difficulties or agrammatism), patient and communication partner education regarding compensatory strategies for improved communication efficiency, and generalization of these skills into the greater community (Doesborgh et al., 2003). Further investigating interventions for individuals with agrammatic aphasia, Thompson and colleagues (2003) concluded that speech-language treatment that focused on syntactically complex structures within sentences actively promotes generalization to less complex material, demonstrating that syntactic complexity is an encompassing factor in the treatment of aphasia (Thompson et al., 2003).

When assessing aphasia, speech-language pathologists (SLPs) utilize a blend of quantitative and qualitative measures to aid in determining clinical diagnoses and treatment outcomes. Quantitative measures are typically collected using scores from standardized, formal assessments. Qualitative data, or data that is descriptive and conceptual, is derived from an array of informal assessments such as through interviews, observations, and conversational language analysis. Within clinical practice, an overarching goal of assessment is to use findings based on a patient's baseline form, content, and use of language to inform effective treatment and best predict recovery outcomes and trajectories for people with aphasia relative to patient-specific neural and behavioral presentations (Thompson, 2019).

Computer aided text analysis (CATA) is an intersection of quantitative and qualitative research that aims to quantitative analysis of unstructured data and extract the thoughts and emotional attitudes from individual narratives and written texts. Due to the advancement and accessibility of software programming and computational powers, this type of analysis has the ability to investigate the superficial (e.g., passage breaks or word-count frequency) and latent semantic attributes of language embedded in a text sample (Castelfranchi, 2017). Sentiment analysis, the automated process of deriving positive, negative, or neutral opinions from text, is one specific application of CATA and machine learning used to categorize text. Sentiment analysis processes a specific unit of text and generates an output of quantitative scores or classifications to identify whether the scripted algorithm considers the text to convey a positive or negative sentiment.

Past studies have applied sentiment analysis towards consumer-driven and marketing research. Fewer studies have researched how sentiment analysis can be applied to healthcare domains (Denecke, 2015). An objective, time-efficient (automated), and accessible health-related quality of life measure from the discourse of persons with aphasia (PWA) initially and throughout treatment, may allow for insights into the post-stroke recovery process and impact on one component of a patient's quality of life. Specifically, in the field of speech-language of pathology, it is unknown if sentiment analysis can reveal meaningful clinical information when applied to clinical language samples. If applicable, this may potentially serve as an enhanced metric in health-related quality of life assessment batteries and informed clinical decision making in post-stroke rehabilitation. While health related quality of life and the investigation of sentiment within the context of healthcare and allied health is a multidimensional construct, analysis of patient sentiment in healthcare settings may further add insight by systematically capturing a reflection of a patient's health status relative to a treatment or intervention over time (Denecke & Deng, 2015).

The purpose of this exploratory study is to apply a methodology for programmatic analysis of the sentiment of transcribed post-stroke speech samples (text) and assess change over time. It is anticipated that automated text analysis of speech transcriptions will identify a key health-related quality of life measure: sentiment of communicated messages. The following aims and hypotheses will be addressed in this thesis.

AIM 1: To apply an automated, programmatic analysis (Python) to post-stroke speech samples and identify changes in the form and content of language as time post-stroke increases.

Hypothesis 1: A decrease in the number of non-words and increase in grammatic categories will be seen across speech samples over time.

Rationale: Despite following a general course of improvement, the most significant gains in language recovery from post-stroke aphasia are typically observed in the first three months post-stroke (Bakheit et al., 2007; Holland et al., 2017; Laska et al., 2001; Pickersgill & Lincoln, 1983, Stockert et al., 2016). Studies have also demonstrated the benefit of incorporating computer-aided methods into the assessment process to document clinical progress over time (Price, 2010). Through the application of programmatic analyses, efficient, automated analysis of post-stroke language data will be completed. It is expected that the most significant decrease in non-words over the first three months post-stroke and grammatic units will expand as time post-stroke increases.

AIM 2: To apply programmatic natural language processing methods (sentiment analysis) to post-stroke speech data.

Hypothesis 2: Changes in the overall sentiment of speech transcriptions will be noted from initial onset to one-year post-stroke.

Rationale: Previous studies have outlined the dynamic and longitudinal nature of the post-stroke language recovery process (Denier et al., 2014; Gerstenecker & Lazar, 2019; Johnson et al., 2019). It is unknown if the underlying sentiment expressed in an individual's language changes as time post-stroke increases. Due to the advancement and accessibility of software programming and computational powers, computer-aided text analysis has the ability to both investigate both the superficial (e.g., word-count frequency) and latent semantic attributes of language embedded in a text sample (Castelfranchi, 2017). Using this methodology, is hypothesized that the speaker's overall sentiment will not remain longitudinally stagnant and instead vary in a quantifiable manner.

CHAPTER 2: REVIEW OF THE LITERATURE

This review of the literature presents pertinent information associated with aphasia, neuroplasticity, and outcomes for this population. Current research related to aphasia and post-stroke assessment and rehabilitation, as well as extant research on the direct applications and statistical underpinnings of sentiment analysis, is summarized.

Neuroplasticity & Aphasia

Aphasia is an acquired language disorder evoked by damage to the language network systems of the brain, most commonly due to a lesion in the left hemisphere. The language impairments brought by aphasia are not the result of a sensory motor disorder, an intellectual deficit, dementia, or other disorders psychotic in nature (McNeil & Pratt, 2001).

One of the most widely used categorization systems used to describe aphasia presentation in patients is not based on lesion location, but rather the specific language impairments and deficits in language-dependent cognitive processes (Brookshire & McNeil, 2014; Hillis, 2007; Hoffman & Chen, 2013). Under this system, aphasia is classified as “fluent” or “non-fluent” and further characterized by factors, such as an individual’s degree of comprehension of spoken messages and patterns of repetition. A patient’s symptoms may not align perfectly into a single subtype classification and may evolve over time. In addition to this dynamic presentation, recovery from aphasia is highly convoluted because it encompasses a broad cognitive domain (language), but with different functions, such as spoken language, writing, and other activities of daily living (Koenig-Bruhin et al., 2013).

The human brain is constructed for change and exquisitely adaptable. Neuroplasticity is the ability of the brain and central nervous system to change and adapt in response to environmental experience, injury, or disease (Ludlow et al., 2008). Given the brain’s lifelong plasticity, known as neuroplasticity, the nervous system adapts to these incidences of stress by reorganizing its cortical structure, function, and neural connections (Carey et al. 2019). Neuroplasticity is a mechanism by which the brain encodes and learns behaviors throughout a lifetime and is also the pathway the damaged brain relearns to function (Kleim & Jones, 2008).

Within the realm of stroke rehabilitation and recovery, the focus of neuroplasticity tends to be on how the brain adapts after a stroke impacts and alters premorbid functioning. After a stroke, the injured brain is challenged to coordinate and execute sense, movement, and communication with a suboptimal neural system. Different brain functions occur to promote neural and synaptic reconstruction after a stroke. Even in the absence of targeted rehabilitation efforts, individuals with brain damage post-stroke often develop compensatory strategies to perform daily functions. In response to this deviation from typical functioning, neural plastic changes occur immediately after in the months, weeks, and years following a stroke (Cramer et al. 2008).

These changes can be broken down as learning and experience dependent. Neuroscience research has highlighted that via learning, the brain, regardless of age, is continually remodeling its neural circuitry to process new experiences and evoke behavioral change (Allred et al., 2014; Kleim & Jones, 2019; Raymer et al. 2008). Following a stroke, learned neuroplastic changes can be shaped and driven by experiences occurring post-stroke, demonstrating that experiential learning is a core principle of neuroplasticity and is at the root of neurorehabilitation efforts. (Carey et al., 2019; Janssen et al., 2010). Stroke rehabilitation, in an attempt to establish restorative outcomes, capitalizes on how an individual's recovery experience can be best tailored to evoke positive plasticity.

Basic neuroscience research has highlighted several experience-dependent principles of neuroplasticity that hold relevance to rehabilitation outcomes (Kleim & Jones, 2019; Carey et al., 2019). The most general of these principles demonstrates that if a neural circuit or substrate is not actively engaged, functional degradation may result over time (Ludlow et al. 2008). First illustrated by Hubel and Wiesel in their 1960s visual deprivation experiments, the researchers showed that if a kitten is deprived of typical visual experiences early in life (achieved by suturing one eye shut), the neural circuitry in the visual cortex is irreversibly altered (Hubel & Wiesel, 1965). In primates, Merzenich and colleagues showed that amputation of the third digit in eight adult owl monkeys resulted in decreased cortical somatosensory representation for that body part (Merzenich et al., 1984; Vega-Bermudez & Johnson, 2002). Following a brain injury, a further cortical loss can occur without retraining, as movements formerly represented in a lesioned zone may not spontaneously reappear in neighboring cortical regions (Friel et al., 2000; Nudo & Milliken, 1996). Failing to engage a brain network due to lack of use may lead to further functional degradation. For instance, Robbins et al. (2007) suggest that tube feeding may promote the disuse of the swallowing mechanism, which in turn reduces its cortical representation in brain topography and result in diminished swallowing capability long-term.

The counterpart to the use it or lose it principle, various studies also demonstrate how neuroplasticity can be evoked with specific, repetitive, and intense training (Kleim & Jones, 2008). Capitalizing on key aspects of neuroplasticity is fundamental to resulting language recovery overtime. While early language therapy is beneficial in the acute phase, language treatment does benefit people with aphasia in the chronic phases of recovery, suggesting that neuroplasticity and activity-dependent intervention are mutually reinforcing (Allred et al., 2014; Hamilton et al., 2011; Moss & Nicholas, 2006). Imaging studies post-implementation of intense speech and language therapy in individuals with chronic Broca's aphasia have revealed structural changes in the arcuate fasciculus, showing that plastic changes can occur in areas that were relatively in-tact post-stroke to further aid in the brain's language rehabilitative efforts (Schlaug et al., 2009).

Post-Stroke Language Recovery

Aphasia is a prevalent phenomenon occurring post-stroke, affecting upwards of 21-38% of individual cases (Laska et al., 2001; Gerstenecker & Lazar, 2019). Upon admission in an acute care setting, aphasia symptoms can range from minor word-finding challenges to profoundly impacting the underlying rule structure of language. Speech and language deficits resulting from aphasia change throughout the acute, subacute, and chronic phases of recovery (Hillis et al., 2018; Gerstenecker & Lazar, 2019). Recovery from post-stroke aphasia is a longitudinal, effortful process aimed at ameliorating speech and communication difficulties. Speech-language therapy is a primary contributing factor in recovery and supports long-term recovery, even in the chronic stage (Engelter et al. 2006, Johnson et al., 2019; Robey, 1998).

Given the heterogeneity of aphasic patients, however, recovery from post-stroke aphasia is notoriously difficult to predict (Denier et al., 2014). Despite following a general course of improvement, the trajectory for language recovery from post-stroke aphasia is known to decelerate, with the most prominent gains occurring in the initial, early acute period following a stroke (Holland et al., 2017; Pickersgill & Lincoln, 1983; Stockert et al., 2016). This period typically encompasses the first three-months post-stroke (Holland et al., 2017; Laska et al., 2001). For instance, Bakheit et al. (2007) report that across a 24-week study period, the language function of aphasic patients steadily improved, but the fastest rate of improvement was noted in the first four weeks post-study recruitment immediately following the stroke.

Initial severity of aphasia, when measured soon after stroke occurrence, is reported to be predictive of patterns of long-term speech and language outcomes (Plowman et al., 2012; Harvey, 2015). For example, Wade et al. (1986) noted that the initial severity of aphasia was linked to the degree of improvement in aphasia as evidenced by improvement in assessment battery scores. An additional longitudinal study carried out by Laska and colleagues (2001) documenting aphasia recovery at 3-, 6-, and 18-months post-stroke found that initial severity of aphasia is negatively correlated with the degree of aphasia recovery. Similarly, Pedersen et al. (2004) also reported that initial severity of aphasia, as measured by WAB scores, was a relevant predictor of improved language outcomes one-year post-stroke, but not other secondary factors such as age, sex, or type of aphasia. While symptoms of aphasia remained in 61% of participants one-year post-stroke, the majority did, in fact, present with a reduction in the severity of speech and language deficits and a strong relationship was noted between initial aphasia severity and long-term language outcomes within the second to fourth-week post-stroke (Pedersen et al., 2004). On the contrary, looking at scores on the Western Aphasia Battery (WAB) or the Boston Diagnostic Aphasia Assessment (BDAE), Lazar and colleagues (2008) showed that patients with less severe scores 72 hours post-stroke were as likely to make gains from baseline naming scores 90 days, or three months later. The authors conclude that post-stroke language recovery is multidimensional and cannot be predicted solely on severity or socio-demographic factors in the acute stage of recovery (Lazar et al., 2008).

Further research on socio-demographic factors, such as age, gender, educational level, and handedness related to post-stroke language recovery is equivocal and establishes no clear link between outcomes (Croquelois & Bogusslavky, 2011; El Hachioui et al., 2013; Ellis & Urban, 2011; Kremer et al., 2013; Lazar et al., 2008; Plowman et al., 2012).

Effect of tPA on Language Recovery

Approved as a post-stroke intervention by the Federal Drug Administration (FDA) in 1996, tissue plasminogen activator (tPA) has proven to be a gold standard treatment for ischemic or thrombotic stroke. tPA is a naturally occurring protein found on endothelial cells, otherwise known as the cells that line the blood vessels. It works by preventing blood clots that obstruct blood flow to the brain by serving as a catalyst to activate the conversion of plasminogen to plasmin, the specific enzyme responsible for breaking down a clot (Klabunde, 2007). tPA destructs occluding clots and restores blood flow to the ischemic tissue in the brain, resulting in salvation of damaged tissue and potential restoration of function that would have been jeopardized with continued occlusion (Meiner et al., 2010). In one retrospective study, it was noted that compared to individuals who were not treated with intravenous tPA, tPA-treated patients had an overall lower length of stay in rehabilitation centers and showed continued improvements in functional outcomes up to one year after the acute phase of treatment. Additionally, the difference between tPA-treated patients' NIHSS scores between admission to inpatient neurology and admission to an inpatient rehabilitation unit during the acute period of recovery was larger compared to the counter group, a link between improved short-term neurological outcomes post-stroke as a result of tPA administration (Meiner et al., 2010). This aligns with other previous work that supports the use of tPA as an effective treatment for ischemic stroke due to a reduction of dependency due to functional impairment and mortality (Wardlaw et al., 2015).

While stroke patients receiving tPA in the acute phase have been noted to experience significant instances of general functional improvement, the direct impact of tPA on speech and language recovery post-stroke is less explicitly documented across sources of empirical literature

(Felberg et al., 2002). Another retrospective study of 228 participants who received tPA by Martins et al. (2016) analyzed NIHSS scores composite verbal scores as a measure of aphasia severity and language recovery. The authors concluded that tPA contributed to an early recovery of aphasia given that approximately a third of patients recovered completely and an additional 40% of participants documented some degree of language improvement one-week post tPA administration (Meiner et al., 2016).

Baseline aphasia severity and lesion size were also noted to serve as predictors of language progress. These findings support a previous observational cohort study comparing the Boston Diagnostic Aphasia Examination (BDAE) scores of patients who received and did not receive tPA after ischemic stroke. At one week and three months post-stroke, patients treated with tPA scored better on the BDAE, which the authors conclude is indicative of an upward language recovery trajectory pattern (Jacquin et al., 2014).

Assessment of Communication and Language Post-Stroke

The emphasis for many speech-language pathologists immediately after and in the time following a stroke is assessment. Assessment is generally defined as the collection of data for the purpose of decision-making in educational, healthcare, and other person-centered contexts (Pierangelo & Giuliani, 2016). After a stroke, assessment aims to determine the presence or absence of post-stroke aphasia, provide a level of severity, and differentiate post-stroke aphasia from other motor-speech or neurogenic communication disorders (Shultz, 2009**)**. Assessment in post-stroke aphasia also allows speech-language pathologists a window into a patient's communicative strengths and weaknesses at a particular point in their recovery, allowing for both therapists and members of a patient's interdisciplinary healthcare to make informed recommendations for future intervention and support services (American Speech-Language-Hearing Association [ASHA], n.d.; Thomson et al., 2018).

Broad assessment of functional outcome post-stroke, including the presence of post-stroke aphasia, most often starts with the National Institutes of Health Stroke Scale (NIHSS) (Payabvash et al., 2010). The NIHSS is a common, quantitative clinical diagnostic tool that was built to assess the cognitive deficits brought forth by a stroke and assist healthcare professionals in determining stroke severity (Kwah & Diong, 2014). Consisting of a 15-item impairment scale, which includes a subsection dedicated to language functioning, the NIHSS gained notoriety as a clinical assessment tool after its use as an outcome measure in clinical trials assessing the effect of recombinant tissue plasminogen activator (tPA) (Lyden, 2017; Kwah & Diong, 2014). Despite being an accessible metric for the recording of clinical progress, it is not a substitute for comprehensive neurological examinations or language-specific batteries in patients presenting with post-stroke functional and language impairments (Marsh et al., 2016; Payabvash et al., 2010).

Current assessment specifically related to post-stroke aphasia is traditionally a constellation of both standardized and non-standardized measures (ASHA). Standardized assessments, or assessments empirically designed with established statistical measures of reliability and validity, requires adherence to uniform administration and scoring protocols to make individual and cross-group comparison accurate. Within the field of speech-language pathology, standardized assessments are typically classified as being norm-referenced tests or criterion-referenced tests. Norm-referenced tests seek to compare a test taker to others in a statistically selected group of individuals, such as age (American Speech-Language-Hearing Association [ASHA], n.d.). Criterion-referenced assessments, such as the Western Aphasia

Battery (WAB), report how well an individual is performing relative to a predetermined performance level or expectations (Brookshire & McNeil 2014, Coelho et al., 2005; Kertesz & Poole, 1974). Presently, widely used English language assessment batteries for post-stroke aphasia include the Comprehensive Aphasia Test (CAT), the Western Aphasia Battery (WAB) and the Boston Diagnostic Aphasia Exam (BDAE) (Wilson et al., 2018).

Informal Clinical Assessments

Non-standardized, or informal assessments are more flexible. Rather than explicitly following a series of steps, informal assessment is instead based on a broader set of guidelines or a framework (Thomson et al., 2018). A process fueled by active and critical thinking, informal assessments, such as language sampling in the form of semi-structured interviews or conversations is more reflective of naturalistic, real-world scenarios (Armstrong & Morensten, 2006, Thomson et al., 2018; Murray & Coppens, 2013). For both standardized and non-standardized measures, it is imperative that clinicians administering such assessments are aware of potential cultural biases associated with test administration.

According to Wilson et al. (2018), for an aphasia battery to be effective in research contexts, the assessment must have an ease of administration, be psychometrically sound, and should produce a multifaceted description of an individual's language functioning strengths and weaknesses. While existing batteries, such as the CAT, the WAB, and the BDAE are comprehensive assessments and hold statistical validity; these assessments require lengthy administration periods and may not accurately represent true, everyday conversational competence. For example, while the CAT is a reliable measure of post-stroke aphasia in adults, Bruce and Edmundson (2009) found that when a sample of 56 adults with aphasia was administered the CAT, no individual subject completed the assessment in a single testing session. Instead, individuals were found to be more likely to stretch the language subtest section to an additional setting, despite the CAT's authors estimate that the test can be completed in its entirety during a 90–120-minute session and their recommendation that the cognitive and language subtests be completed in tandem (Bruce & Edmundson, 2009). Regarding outcome assessment specifically, Simmons-Mackie et al. (2005) reported that the most frequently identified barrier to outcome assessment in aphasia among practicing speech-language pathologists, across settings, is clinical time constraints.

Another underlying goal in stroke rehabilitation is the improvement of an individual's quality of life. Speech-language pathologists, however, rarely address the quality of life of patients with aphasia directly and explicitly due to the multidimensional nature of quality of life and a lack of available and sensitive assessments measuring quality of life in relation to communication (Borglin et al., 2005; Cruice et al., 2003; de Haan et al., 1995). Among persons with aphasia, as opposed to clinician-administered assessment measures, personal thoughts and perceptions of aphasia are best divulged through self-assessment measures (Babbit & Chenery, 2010). There is variability, however, among these current self-assessment rating scales in that there is no standard among measures. For instance, the Communicative Effectiveness Index (CETI) presents 16 functional situations and asks respondents to mark their responses on a 10-cm visual analog scale (VAS), with the designations "not at all able" and "as able as before the stroke" serving as anchor points on the line (Lomas et al., 1989). The ASHA Quality of Communication Life (ASHA-QCL), another self-report questionnaire, instead uses a five-point vertical line with graphic symbols on each end (Paul et al., 2004; Cherney et al., 2011). The Stroke and Aphasia Quality of Life Scale (SAQOL-39) is another metric with established reliability and validity designed specifically for the post-stroke population measuring health related quality of life across physical, psychosocial, and communication, and energy domains (Hilari et al., 2003).

Additionally, no one aphasia battery can capture the true nuances of conversation and may overestimate an individual's communication skills due to the inability of these assessments to narrowly pinpoint cognitive or linguistic deficits and the ceiling effect observed in aphasia batteries (Murray & Coppens, 2013; Coelho et al., 2005). Instead of a routine dependence on a standardized assessment, assessment of post-stroke aphasia is most comprehensive when blends of standardized and non-standardized assessments are used complementary to gauge speech function (Armstrong & Morensten, 2006; Murray & Coppens, 2013).

A crucial supplement or functional alternative to restrictive standardized assessment to gauge an individual's communication abilities is language sampling and analysis (LSA) (McCauley & Swisher, 1984; Price et al., 2010). LSA is a versatile tool that allows for the collection and interpretation of authentic spoken language productions in naturalistic or spontaneous contexts (Miller et al., 2016; Price et al., 2010). Standardized assessments, designed for efficiency, oftentimes provide insufficient opportunities for language production in a restrictive environment and are not useful in monitoring progress consistently over time (McCauley & Swisher, 1984). Free from highly regimented administration protocols, LSA provides documentation of spoken language competency within real-life contexts, can be elicited at frequent time intervals from speakers of any age, and can be modified to better acknowledge cultural diversity- allowing for a more comprehensive assessment of post-stroke language (Heilmann & Westerveld, 2013; Miller et al., 2016; Rojas & Iglesias, 2010; Stockman, 1996). While LSA is regarded as a gold standard assessment for assessing spoken language production by researchers and clinicians alike, three of the most popular computer-based LSA approaches including the Systematic Analysis of Language Transcripts (SALT), Child Language Data Exchange System (CHILDES) and Computerized Profiling are only a source of normative data for child language analysis (Pavelko & Owens, 2017; Miller et al., 2016). Sources of normative reference points for adult language transcript analysis are inaccessible besides those reference norms provided by formal, standardized assessments of speech and language.

Language Sampling, Sentiment Analysis, and its Applications

Current methods for the analysis of speech measure performance in naturalistic speaking contexts (speech transcriptions), are developmentally sensitive and measure the morphology (form of words), syntax (grammar), semantics (vocabulary used), and pragmatics (appropriateness) of the speaker. Thus, speech sampling is a powerful method of documenting language use across various speaking situations (Miller et al., 2016). Analysis of language samples allows for speech-language pathologists (SLPs) to describe a client's expressive language abilities at a given point in time, as well as to identify client strengths and weaknesses when assessing the extent or severity of speech impairments (Price et al., 2010). However, there is currently no methodology that has been applied to the language samples of individuals with communication disorders that objectively assess an individual's overall sentiment, topic, and message intent along with changes in sentiment over time. Objective information related to an individual's overall sentiment is an important clinical indicator and measure of quality of life (Borglin et al., 2005; Gabriel & Bowling, 2004; Leedham et al., 1995) that is not consistently or accurately being applied to the assessment process due to previous analytical limitations.

Saif et al. (2013) documented inconsistencies of manual inspection of sentiment in text and lack of clear criteria across raters for sentiment quality. Agreement on sentiment was found to be as low as 60% when individuals were asked to judge the sentiment of text (Saif et al., 2013). Thus, this aspect is currently heavily subjective and influenced by personal experiences, thoughts, and beliefs of raters. The use and application of a consistent sentiment analysis classification system has the potential to streamline criteria for analysis, reduce inter-rater discrepancies, and improve data consistency. Incorporating computer-aided methods into assessment procedures and research methodology allows for a more efficient and comprehensive language sample analysis. This also allows for both clinicians and researchers to track and document clinical progress over a longitudinal period (Price et al., 2010). This shows significant promise for the clinical utility of sentiment analysis for the analysis of the speech of individuals with communication disorders.

Sentiment analysis, also referred to as opinion mining, is a branch of natural language processing via text analysis that aims to detect, extract, and analyze opinionated text (Alharbi & de Doncker, 2019). Sentiment analysis is a sub-field of Natural Language Processing (NLP) that aims to identify and extract opinions within a given text. The objective of sentiment analysis is to gauge the attitude, sentiments, evaluations, attitudes, and emotions of a speaker or writer based on the computational treatment of subjectivity in unstructured text. With origins in the field of web mining, research in sentiment analysis is closely linked to the rapid growth of user-created content, such as online discussion forums, blogs, and product reviews, on the World Wide Web (Li & Wu, 2010). Rather than conducting traditional polls or focus groups to gain product feedback, organizations are shifting towards the utilization of sentiment analysis as a way to unveil how positively or negatively their specific entity is regarded (Alharbi & de Doncker, 2019; Paltoglou, 2016). Sentiment analysis allows for the content of natural language, the words written or spoken by individuals, to be examined and then stratified by polarity and emotion (Gohil et al., 2018, Greaves et al., 2013b).

Natural language processing of large datasets utilizing sentiment analysis has been crucial to understanding consumer attitudes and behaviors (Petz et al., 2013). Sentiment analysis involves the process of computationally identifying and categorizing opinions expressed in text to determine the speaker's attitude towards a particular topic. Sentiment analysis enables the content of natural language to be examined for positive, negative, or neutral opinions, emotion, and intent (Pang & Lee, 2008). Further, sentiment analysis systems allow unstructured information (text/speech transcriptions) to be automatically transformed into structured data for objective analysis. If applicable to SLP evaluations, these analytical methods could allow for the interpretation of speech content related to patient experience in an efficient and objective manner. For example, sentiment analysis may offer the potential to develop insights into the attitudes of a speaker during the language recovery process following a stroke or other brain injury. Conventional quantitative and analytical procedures for speech and language assessment have not previously been able to capture this aspect of communication analysis. Thus, an exploratory analysis is needed to determine the applicability of sentiment analysis to post-stroke language samples.

However, identified opinion information has been used to help people and organizations make sound decisions, leading to a breadth of real-world applications rooted in sentiment analysis (Alharbi & de Doncker, 2018). Within marketing and consumerism, sentiment analysis allows business analysts to unlock and gauge the speaker's attitude or opinion through that user-generated content widely available on the Internet. It is human nature to reference friends, colleagues, and specialists in search of an opinion towards the quality of a product. The Internet and the accessibility of user-generated content provide individuals with an expansion of their personal social network, allowing them to seek validation from a diverse crowd of people (Priya et al., 2019). A sentiment analytics model allows a business to quickly and efficiently gauge the emotional polarity of reviews pertaining to their company, such as whether a particular stream of text is negative, positive, neutral, or mixed in tone. Moving forward, businesses are then able to implement response measures, such as improving the quality of their products or introducing quality checks on areas most often flagged with negative reviews— eventually providing consumers with an overall improved experience in the future. The same may hold true in a clinical setting.

Branching off from business applications, sentiment analysis also holds useful applications within healthcare and medical domains. In clinical settings, it is not uncommon for unstructured data to play a central role in clinical decision-making and treatment plans. While formal examination and laboratory results are typically reported in a structured manner, other information sources that contribute to clinical decisions, such as physician or patient narratives, observations, and experiences, are communicated in an unstructured manner across multiple means of clinical documentation (Denecke & Deng, 2015). Incorporating techniques derived from natural language processing and sentiment analysis makes it possible for unstructured, prose statements to be quantified into usable measures within healthcare analytics (Greaves et al., 2013a).

Health social media forums are accessible and popular avenues for patients to access health information and share lived experiences, especially among patients diagnosed with chronic conditions (Carrillo-de-Albornoz et al., 2018). Pharmaceutical and medical technology companies are examples of stakeholders who might mine these types of forums as a way to track raw patient opinions on their products and services over time (James et al., 2017). Pharmaceutical companies, in particular, frequently utilize sentiment analysis as a way to track patient's reactions to their products and in turn personalize marketing and outreach efforts to better mirror consumers' lived experience (Grissette et al., 2017).

Additionally, sentiment analysis can be applied to publicly accessed online patient sources of patient commentary to complement quantitative and qualitative derived from traditional survey measures of patient satisfaction and perceptions of health care experience (Alemi et al., 2012; Greaves et al., 2013a, b). Denecke and Deng (2015) applied sentiment analysis to investigate the sentiment expressions present in medical texts, including discharge summaries, radiological reports, drug reviews, and online blogs. Results revealed that physicians and medical practitioners are more likely to express negative sentiments in clinical documentation directly tied to patient experience, as opposed to in drug reviews, inconspicuously. Due to the dynamic nature of health status among patients, it is also stressed that medical sentiment should be examined over time to account for these trajectories. In medical settings, sentiment is often classified as a direct reflection of patient health status, the presence or change of a medical condition, and judgment of treatments or of treatment outcomes (Denecke, 2015). For example, if health statuses fluctuate between "good," "bad," or "normal" at any point in time, sentiment aspects will reflect that change in health status, the outcome or response to treatment, or poignant events related to a patient's diagnosis.

Given the multifaceted nature of healthcare analytics and service quality measures, sentiment analysis techniques for the examination of unstructured feedback have also been reported to be effective when commentary is first segmented by specific quality elements and then processed in data mining algorithms (James et al., 2017).

A currently untapped use of sentiment analysis relates to post-stroke language recovery and changes over time. Sentiment analysis allows for objective quantification of health-related quality of life measures, such as emotion and objective quantification of the language used. These factors are tied to the speech individual's produce.

The Underlying Algorithm in Sentiment Analysis

Sentiment analysis is a classification task that can be performed utilizing a machine learning approach by training a classifier to determine positive, negative, or neutral sentiments in text (Pang & Lee, 2008; Zhang & Liu, 2017). Once a text dataset is acquired, a series of interconnected preprocessing steps allow for the dataset to be cleaned and transformed, allowing a classifier to then extract a maximal amount of accurate information from the text (Kobayashi et al., 2018).

Machine Learning Approach using naïve Bayes

A naïve Bayes classifier is a probabilistic model for text classification based around Bayes theorem, a method of examining conditional probabilities that allows for conditions to be reserved in a convenient, streamlined way (Troussas et al., 2013). A conditional probability is the measure of the probability that event A occurs, given the evidence or prior knowledge that event B has occurred, represented as P (A | B). Bayes Theorem revolves around relating conditional probabilities and is used to calculate the likelihood of an event based on its associations with another event. This allows for the determination of probability when the only information available is the opposite result of the two components individually:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

The reiteration of this is helpful when attempting to estimate the probability of an event occurring based on examples, also known as prior evidence, of its occurrence.

Text classification aims to assign text documents to one or to a variety of categories based on categories such as content, genre, language, or sentiment (Hirschberg & Manning, 2015; Kobayashi et al., 2018). In sentiment analysis, the objective is to determine the writer's point of view and emotional polarity about a topic or service from a piece of text. This methodology is known as a "naïve" Bayesian tangential because it makes a broad assumption of independence of events, despite this assumption being false in a great deal of real-world scenarios. This implies that there is no direct link between one word and another and despite this unrealistic assumption, naïve Bayes classifiers have been noted to run text classification well and have been applicable in various research efforts, such as clinical decisions in treatment processes or spam filtering in e-mails (Kazmierska & Malicki, 2008; McCallum & Nigam, 1998; Troussas et al., 2013).

Lexicon-Based Approach

An additional model known as a lexicon-based approach determines the sentiment or polarity of opinion by calculating the semantic orientation of a word or short phrase that appears in a text. Under this approach, a domain-specific dictionary, or sentiment lexicon, of positive and negative charged words is required, with positive or negative sentiment value assigned to each of these words. Both manual and automatic approaches exist for creating these dictionary databases. When an example of text is examined under a lexicon-based approach, sentiment values from that corresponding dictionary model are assigned to all positive and negative words or phrases within the message and a sum or average is then applied to make a final calculation regarding the overarching sentiment for the message (Jurek et al., 2015). Unlike under machine learning models, lexicon-based approaches are not strictly reliant on labeled data.

Hybrid Approach

A third approach, the hybrid approach, is a union of the previously described machine learning and lexicon-based approaches. The advantage of a hybrid approach is enhanced classification performance compared to a stand-alone machine learning and lexicon approach, yielding better results, however, its applicability to unstructured data is less clear (Ahmad et al., 2017).

CHAPTER 3: METHODOLOGY

De-identified, post-stroke speech transcriptions were used to develop and apply programmatic scripts to analyze the content and sentiment of written transcriptions of speech. The dataset used for analysis was provided to the research team by a well-known collaborator in the field of post-stroke rehabilitation who has independently recorded and transcribed speech samples post-stroke. The dataset included 196 speech transcriptions of brief language samples recorded across one-year post-stroke. Within this dataset, 106 samples were viable for transcription and focused, in-depth analysis. Language samples were transcribed by a professional verbatim transcriber located in Casco, Maine.

Approval to use the dataset in the present study was granted by the Institutional Review Board (IRB) committee at the University of Virginia (IRB Protocol #2653). The goal of the IRB approval process is to protect the rights and welfare of participating subjects and ensure the highest quality of ethically sound research. This study utilized a retrospective case study investigation, using existing data that has been previously recorded for reasons other than the direct purpose of research. Case studies allow for a focused, in depth analysis of a real-world subject example.

Subject Demographics

The participant was a 68-year-old right-handed man who suffered a left ischemic stroke the night of September 26, 2011 at the age of 59 years. The participant holds a Ph.D. in Social Policy and worked as an associate dean in a university setting before his stroke. The participant was admitted to the emergency room at 9:35 PM ET the night of his stroke. tPA was administered intravenously at 10:30 PM, within three hours of initial stroke onset. He was hospitalized for four days, including one day in the neuroscience intensive care unit (Neuro-

ICU). Initial assessment documented deficits in the ability to read, write, or speak fluently secondary to the stroke and a resulting Broca's aphasia.

Throughout his recovery, the participant kept a 193-page diary for 10 months comprised of drawings and other graphic representations. While this started as a method to document homework assignments from outpatient therapy sessions, this journaling exercise became a facet of the participant's everyday routine. Additionally, the participant produced the 196 voice recordings, comprising the dataset, from November 4, 2011, until July 31, 2012. All but 11 of these language samples were recorded by April 2012. These brief language samples averaged 6 minutes and 55 seconds in length and were narrative inner monologues primarily revolving around the speaker's current internal state of being, events of daily living, or recordings of the speaker reading a short passage. Samples were recorded primarily in the late morning and afternoon hours, with an average time of day of 1:08 PM.

In addition to journaling and independently producing voice recordings, the participant received weekly outpatient speech-language therapy services. Individual sessions were held twice weekly for 30 to 45-minute sessions beginning 28 days post-stroke. In total, the participant received 30 sessions over a five-month, five-day treatment period. Outside of individual speech-language therapy sessions, the participant attended 11 classes and three Saturday session clinics at the Boston University Aphasia Resource Center. He additionally self-reported consistent practice at home by utilizing flashcards, graphic organizers, reading Wiki notes (brief online articles, Wikipedia entries, and research articles), and composing emails.

Programmatic Analysis of Post-Stroke Speech Samples

Data analysis involved extracting, transforming, and loading (ETL) the de-identified text data (speech transcriptions) into a data management system. A general-purpose programming language (Python) was utilized for data cleaning, preprocessing, and descriptive text analysis. The Python Pandas library was used to clean and extract the data (McKinney, 2010). This included removing extraneous punctuation, capitalization, and formatting dates and times. Quantitative data on the number of words, non-words present as a whole and over time, number of filler words, and number of unintelligible instances was obtained. Filler words, were defined as short, meaningless non-words or sounds occurring in speech (i.e." or "uh"). These non-words do not semantically alter the content of the message. In addition, stop words, or commonly used words such as "the" or "a" that contribute no additional significant meaning, were removed with NLTK to support sentiment analyses described below. Grammatic units of words used was obtained using the Python Natural Language Toolkit (NLTK) Part of Speech (POS) tagging for classification of speech samples (i.e., nouns, verbs, adjectives, adverbs, prepositions, conjunctions, interjections, etc.) (Bird et al., 2009). Words in a specific language sample are first tokenized, or split from a string, into tokens. POS-tagging then attaches a likely part of speech tag to each token, or word. Tokenization and POS-tagging additionally serves as a prerequisite analysis for additional natural language processing analysis, including sentiment analysis. This process of data cleaning and preprocessing additionally facilitated a decrease in overall processing time. The general workflow utilized for data preprocessing and NLP analyses is provided in Figure 1.

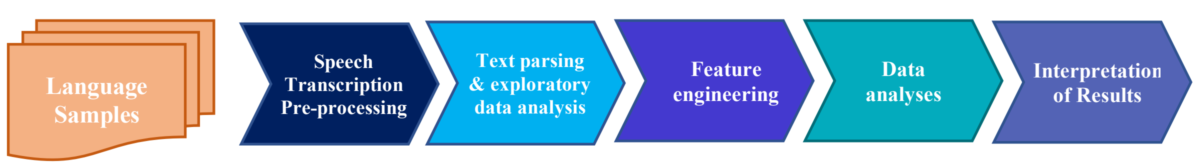

Figure 1. Workflow Utilized for Data Preprocessing & NLP Analyses

Sentiment Analysis of Post-Stroke Speech Samples

Programmatic methods for text analysis (natural language processing with sentiment analysis) were applied to identify feasibility of analysis for future studies. The Valence Aware Dictionary and sEntiment Reasoner (VADER) sentiment analysis package was adapted to assess the sentiment of speech transcriptions. VADER is an open-source, rule-based model for sentiment analysis of text. The corpus of reference in this package is specifically validated to accurately identify sentiment expressed in short text responses. VADER uses a combination of a sentiment lexicon (corpus of words) and a list of lexical features, which are labeled according to their semantic orientation as either positive or negative.

VADER has been found to be highly specific and reliable when dealing with unstructured text responses such as social media texts, news editorials, movie reviews, unstructured survey responses, and product reviews. Hutto and Gilbert (2014) found that VADER performed as well as individual human raters ($r = 0.881$ vs. $r = 0.888$, respectively) at accurately identifying the sentiment of the input data. Additionally, training data is not needed as the reference corpus is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon. VADER not only identifies the positivity and negativity of the text (via the compound score), but also indicates how positive or negative the sentiment of the response is (using positive, negative, and neutrality scores). This analytic method was be applied to the post-stroke speech transcriptions.

How Positive, Neutral, and Negative Lexicon Scores Are Obtained

The Positive, Negative, and Neutral scores represented the proportion speech sample text that fell in these categories for each response in reference to the corpus. The total score for positivity, negativity, and neutrality for each response equals 1.

How Compound Scores Are Obtained

The Compound score is a metric that calculates the sum of all the lexicon ratings, which have been normalized between -1 (most extreme negative) and +1 (most extreme positive). Compound scores greater than 0.40 indicate positive sentiment. Scores below -0.40 indicate negative sentiment. Scores within -0.39 – 0.39 indicate neutrality of the response. The scale ranges from negative one to positive one (Figure 2). Thus, categorical sentiment corresponds to scores falling between -1.00 and 1.00. Sentiment ratings and the associated categories are provided in Table 1.

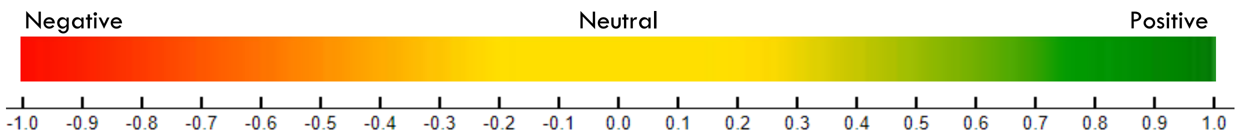

Figure 2. Representation of scale utilized in Sentiment Analysis

Table 1. Categorical Sentiment Ranges

Sentiment	**Score Range**
Positive	0.40 – 1.00
Neutral	-0.39 – 0.39
Negative	-1.00 – -0.40

Summary of Study Aims & Statistical Plan

AIM 1: To apply an automated, programmatic analysis (Python) to post-stroke speech samples and identify changes in the form and content of language as time post-stroke increases.

Hypothesis: A decrease in the number of non-words and increase in grammatic units will be seen across speech samples over time.

Variables: Total number of words, total number of non-words, total number of speech therapy sessions, number of non-words 1-3 months, 3-6 months, 6-9 months, and 9-12 months post-stroke, classification of words used [word/non-word, noun, verb, article, conjunction, adjective, adverb, etc.], and time post-onset.

Statistical Treatment: Descriptive statistics will be used to quantify means and standard deviations related to the form and content of language use over time.

AIM 2: To apply programmatic natural language processing methods (sentiment analysis) to post-stroke speech data.

Hypothesis: Changes in the overall sentiment of speech transcriptions will be noted from initial onset to one-year post-stroke.

Variables: Speech transcriptions from onset to 1-year post-stroke (input variable) and three target class labels to predict negative, neutral, or positive sentiment.

Statistical Treatment: Python programming language will be applied to a VADER Sentiment Analysis Model using multinomial naïve Bayes and logistic regression to determine changes in sentiment over time.

CHAPTER 4: RESULTS

A total of 106 recorded and transcribed language samples were cleaned, preprocessed, and utilized for quantitative and descriptive analysis. Target variables such as word count, frequency of non-words, frequency of unintelligible utterances, language sample length (in minutes), grammatic units, and sentiment were analyzed across time. Analysis was further stratified by months post-stroke as well as by time of day of the language sample recording.

Word Count & Length of Language Samples

The average number of words across all language samples utilized for analysis was 393.54 words. Average word count was highest in the initial 1-3 months post-stroke (M= 439.32, SD= 228.50) and lowest in the 9-12 month post-stroke period (M= 252.33, SD= 131.37) (Table 2). There was no significant correlation between mean word count per language samples across time post-stroke for this participant (*r(*104) = -0.083, p=0.395) (Figure 3). There was a strong, positive correlation between total word count and language sample recording length *r(*104) = 0.932, p<.05). As the length of recordings increased, total word count was also observed to increase (Figure 4).

Table 2. Descriptive Statistics: Word Count per Sample Over Time

	Language Samples (n)	Minimum	Maximum	Mean	Standard Deviation
1-3 Months	34	141	1244	429.32	228.50
3-6 months	49	38	963	387.82	212.19
6-9 months	20	137	759	377.65	159.07
9-12 Months	3	121	406	144.50	277.33

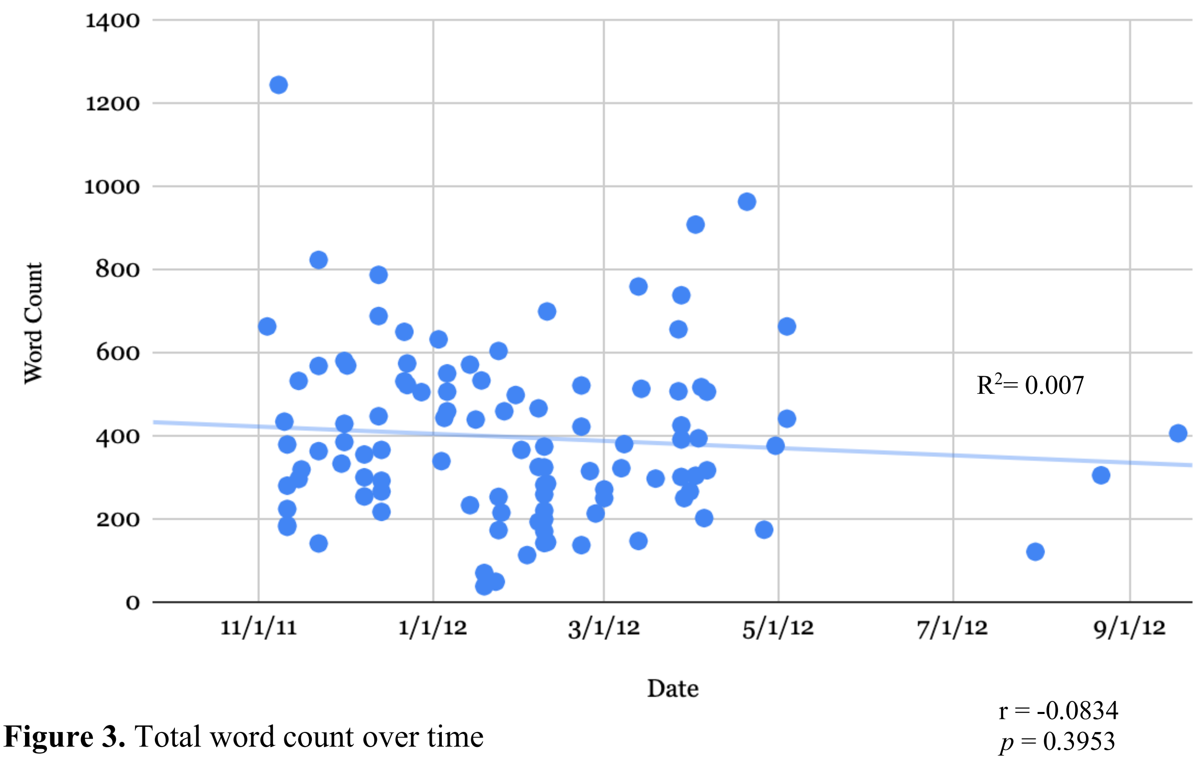

Figure 3. Total word count over time

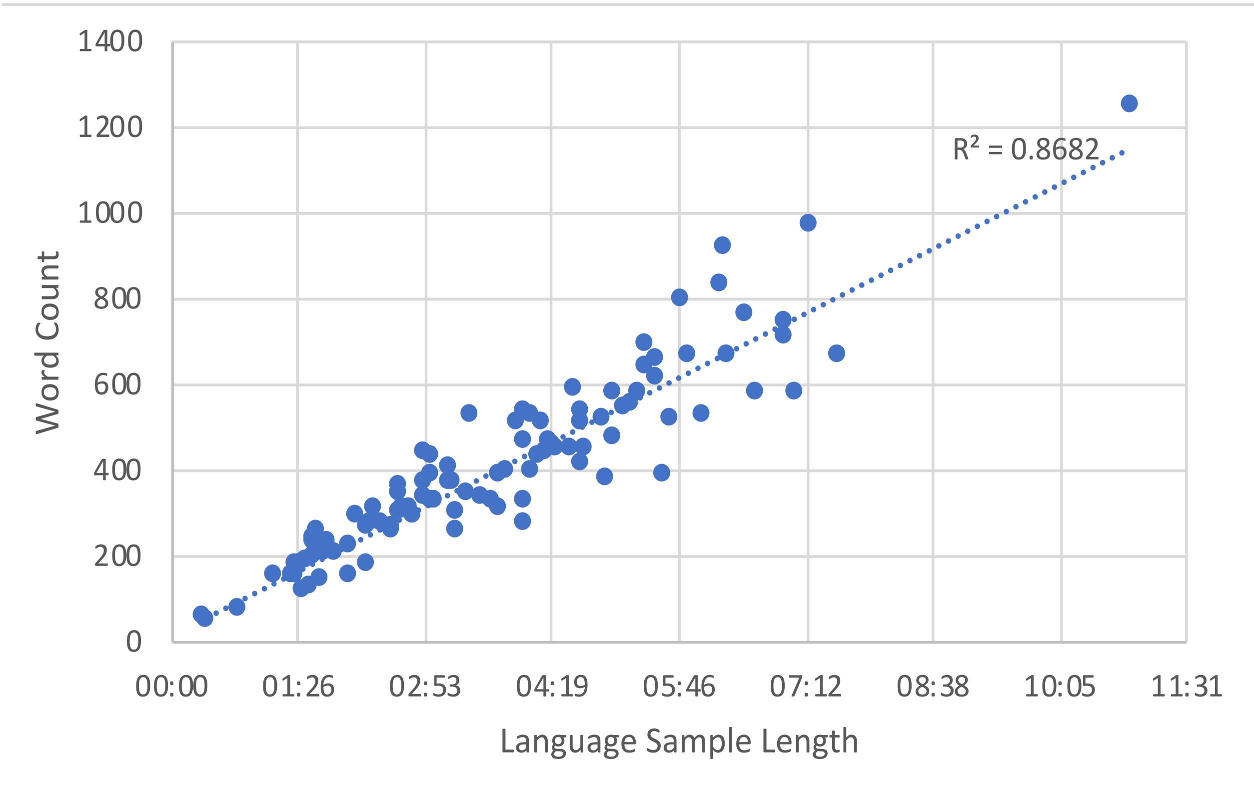

Figure 4. Correlation of total word count to language sample length

r = 0.9318
p = <0.001

Stratification of Word Count by Time of Day

More language samples were recorded in the afternoon compared to the morning (89 and 17, respectively) (Table 3). Word count differed between morning and afternoon recordings. The mean length of language sample recordings for morning recordings was 4 minutes, 5 seconds and 3 minutes, 35 seconds for afternoon recordings. There was no significant differences (t(104)= 0.996, p=0.321) between language sample length in the morning in seconds (M= 4min 5sec, SD= 1min 45sec) and language sample length in the afternoon (M= 3min, 35sec, SD= 1 min, 52 sec). A slightly greater, but non-significant, number of words per sample were used in the morning compared to the afternoon (t(104)= 0.4836, p= 0.6297) (Figure 5).

Table 3. Descriptive Statistics: Word Count per Language Sample by Time of Day

	Language Samples (n)	Minimum	Maximum	Mean	Standard Deviation
Morning	17	121	963	418.41	237.09
Afternoon	89	38	1244	391.82	201.93

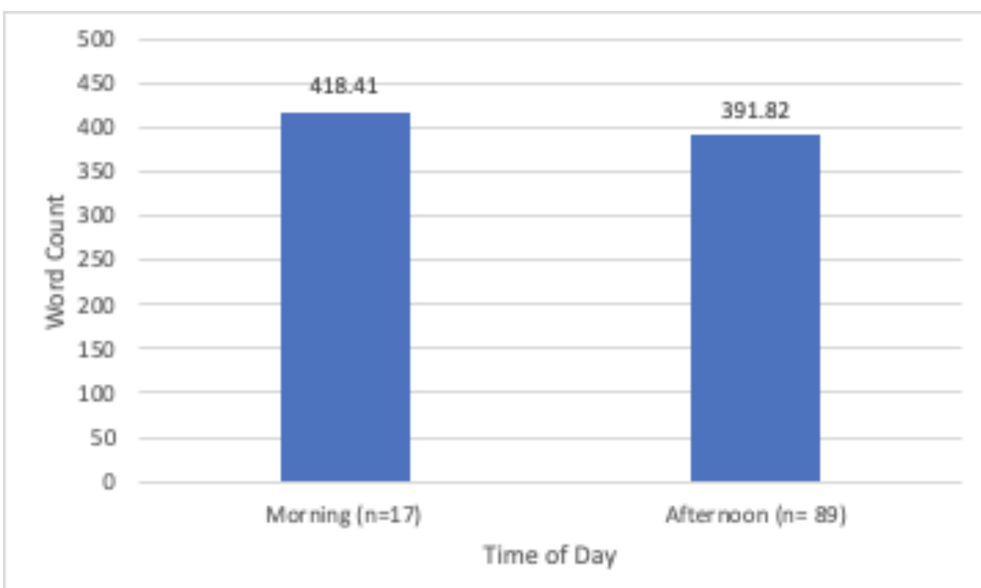

Figure 5. Average word count per language sample by time of day of language sample recording

Filler Words & Non-Words, & Unintelligible Utterances

Filler words, non-words, and unintelligible utterances were quantified over time. It was hypothesized that fillers, non-words, and unintelligible utterances would decrease as time post-stroke increased. Descriptive statistics are reported in Table 4. Fillers, non-words, and unintelligible utterances increased slightly during samples obtained at 3-6 months post-stroke onset (M= 60.55, SD= 36.48). An increase was also observed at 9-12 months post-stroke onset (M= 61.33, SD= 35.13) (Figure 6). However, when correlating the number of fillers, non-words, and unintelligible utterances to time post-stroke, a non-significant, weak positive correlation was noted ($r(104)= 0.058$, $p= 0.56$) (Figure 7). A strong positive correlation was observed between the total number of filler words, non-words, and unintelligible utterances and language sample length ($r(104)= 0.82$, $p <0.05$) (Figure 8).

Table 4. Descriptive Statistics: Number of Filler Words, Non-Words, Unintelligible Instances

	Language Samples (n)	Minimum	Maximum	Mean	Standard Deviation
1-3 Months	34	11	121	49.18	30.57
3-6 months	49	9	135	60.55	36.48
6-9 months	20	13	140	51.1	28.25
9-12 Months	3	23	92	61.33	35.13

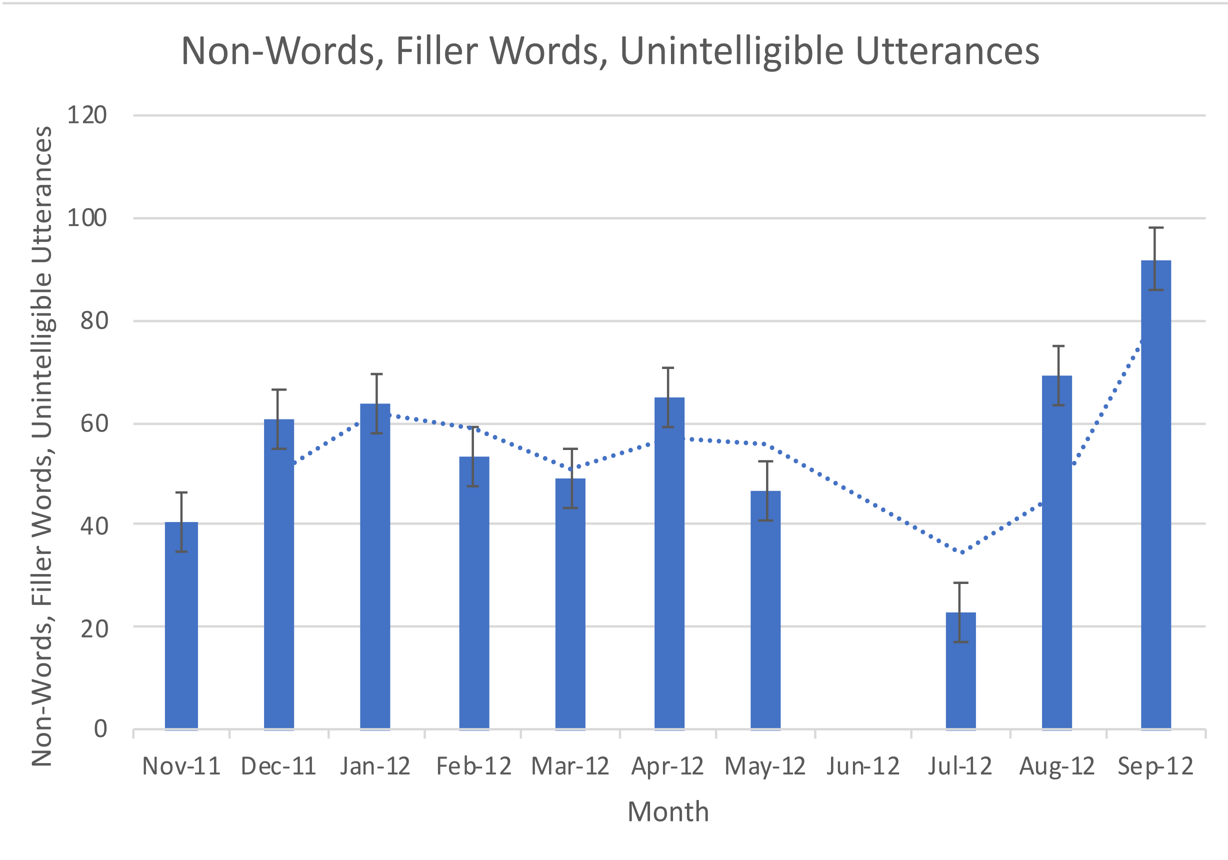

Figure 6. Total number of non-words, filler words, and unintelligible utterances over time

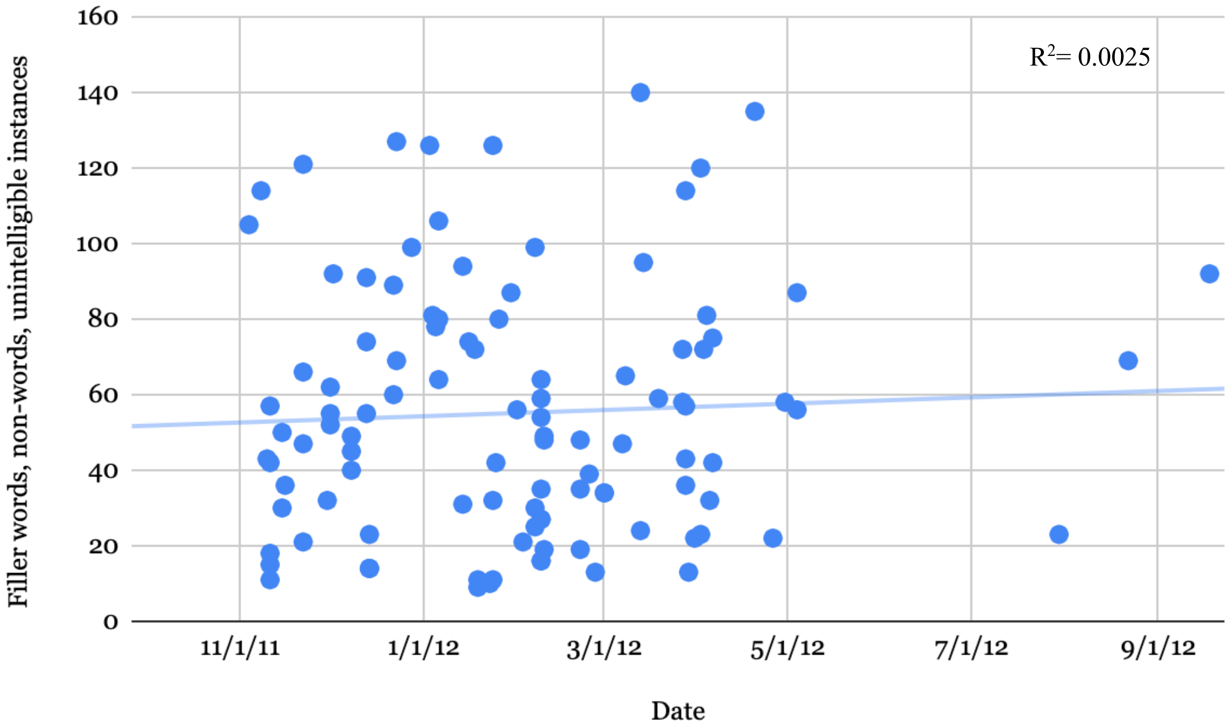

Figure 7. Correlation of filler words, non-words, and unintelligible utterances and time post-stroke onset

r=.0576
p= .55755

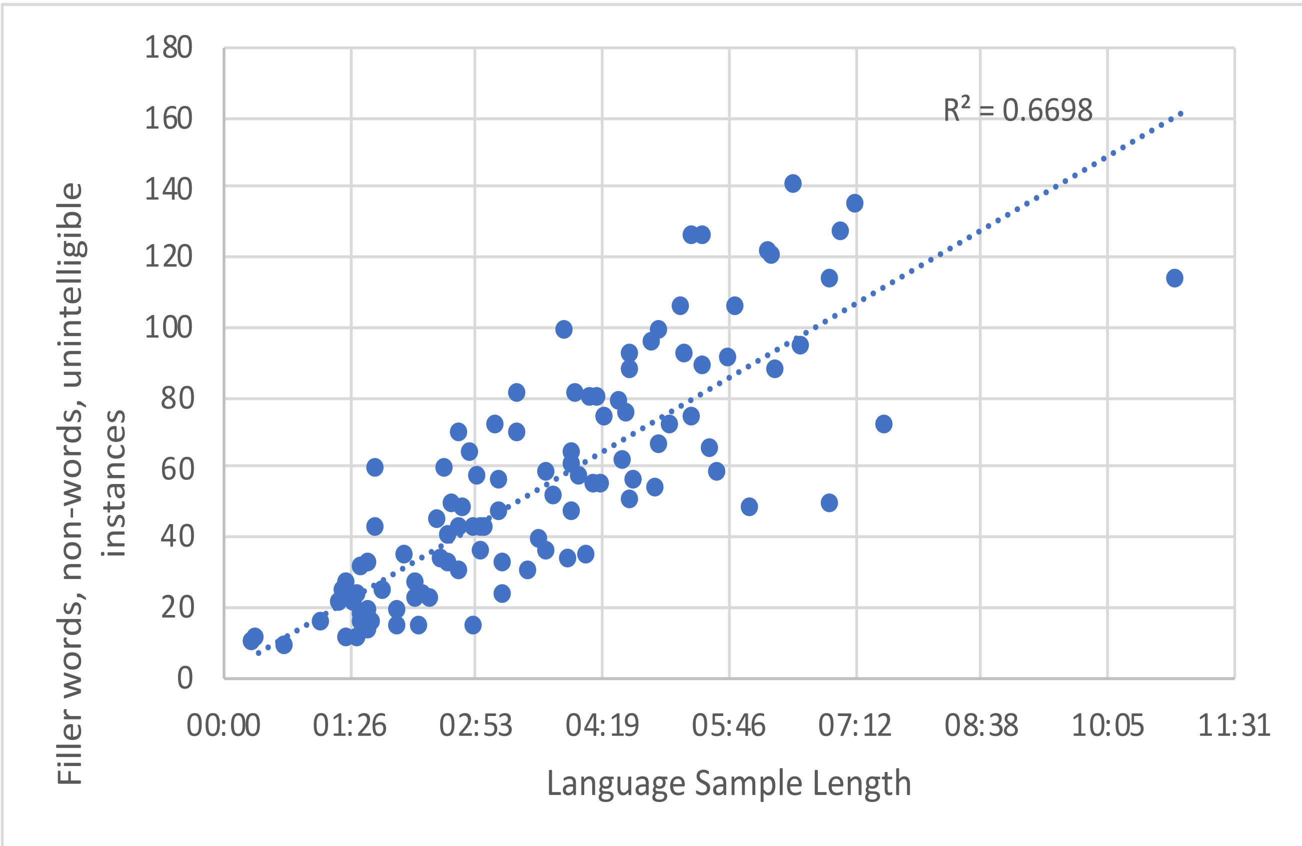

Figure 8. Correlation of filler words, non-words, and unintelligible utterances and language sample length

r=.8184
p= <.001

Grammatic Units

Analysis of speech transcripts utilizing the Python NLTK parts of speech tagging revealed the total number of grammatic units per language sample. Grammatic units analyzed in the script are described in Table 5. An average of 14.55 different grammatic units were used in each transcribed language sample after data preprocessing. The number of grammatic units used by the participant was fairly stable over time (Figure 9). The most frequently occurring grammatic units across language samples included adjectives (99% of samples), singular nouns (100% of samples), adverbs (99% of samples) and verbs in base form, gerund, or present tense (100%, 96.23%, and 95.28% of samples) (Table 6). The majority of samples consisted of present-tense syntax. Fewer samples demonstrated past-tense markers, comparatively. The frequency and percentage occurrence of grammatic units used within the speech samples is provided in Figure 10.

Table 5. Grammatic Units Assessed in the Python NLTK POS Tagging Script

CC	Coordinating conjunction	**RBR**	Adverb, comparative
CD	Cardinal digit	**RBS**	Adverb, superlative
DT	Determiner	**RP**	Particle
EX	Existential there	**SYM**	Symbol
FW	Foreign word	**UH**	Interjection
IN	Preposition or conjunction, subordinating	**VB**	Verb, base form
JJ	Adjective or numeral, ordinal	**VBD**	Verb, past tense
JJR	Adjective, comparative	**VBG**	Verb, gerund
JJS	Adjective, superlative	**VBN**	Verb, past participle
LS	List item marker	**VBP**	Verb, present tense, not 3rd person singular
MD	Modal auxiliary	**VBZ**	Verb, present tense, 3rd person singular
NN	Noun, common, singular or mass	**WDT**	WH-determiner
NNP	Noun, proper, singular	**WP**	WH-pronoun
NNS	Noun, common, plural	**WP$**	WH-possessive wh-pronoun
PRP	Personal pronoun	**WRB**	WH-adverb
RB	Adverb		

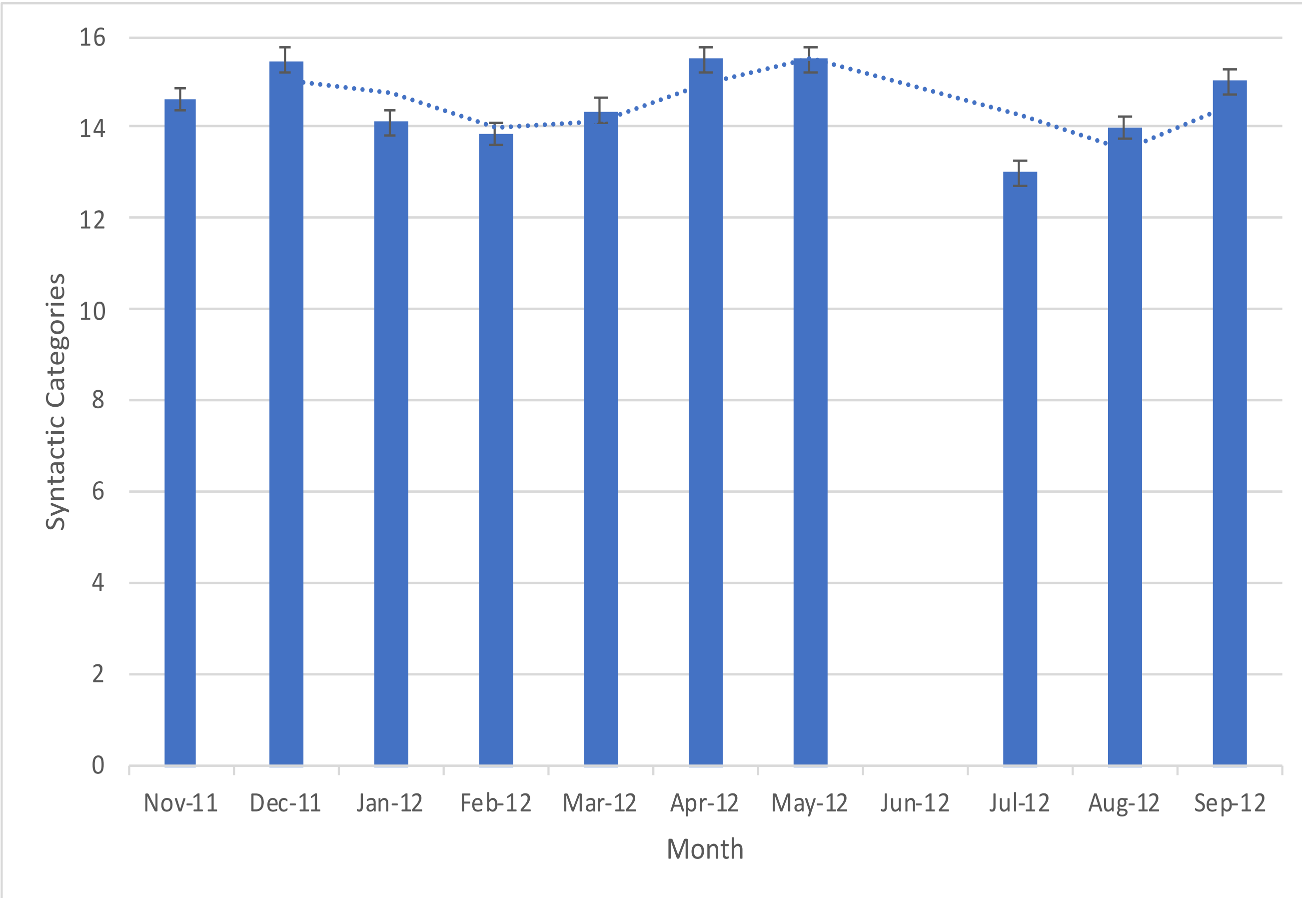

Figure 9. Average number of grammatic units used per language sample over time post-stroke

Table 6. Occurrence of Grammatic Units Used in Speech Samples

Marker	Frequency	Percent
NN	106/106	100%
JJ	105/106	99%
RB	105/106	99%
VBG	102/106	96.23%
VBP	101/106	95.28%
VB	100/106	94.33%
IN	98/106	92.45%
NNS	98/106	92.45%
VBN	92/106	86.79%
VBD	89/106	83.96%
CD	89/106	83.86%
VBZ	86/106	81.13%
MD	80/106	75.47%
DT	53/106	50%
JJS	43/106	40.57%
PRP	38/106	35.85%
JJR	38/106	35.84%
WDT	34/106	32.08%
RBR	30/106	28.3%
RP	15/106	14.15%
UH	9/106	8.49%
CC	8/106	7.54%
FW	7/106	6.6%
NNP	7/106	6.6%
WRB	4/106	3.77%
EX	2/106	1.89%
RBS	2/106	1.89%
WP$	2/106	1.89%
SYM	1/106	<1%
WP	1/106	<1%
LS	N/A	N/A

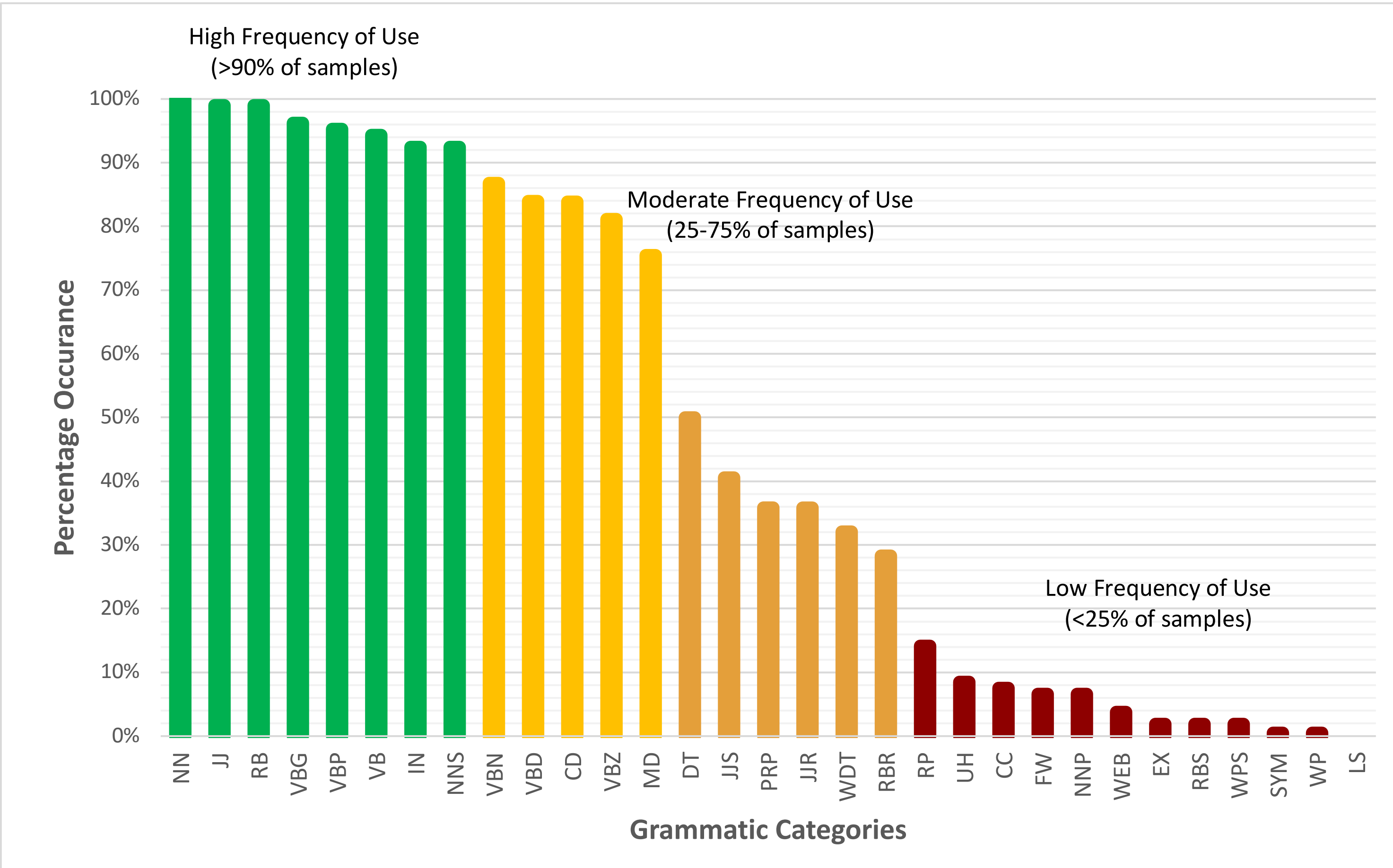

Figure 10. Percentage occurrence for each grammatic category across all language samples

Sentiment Analysis

Normalized compound scores calculating the sum of all lexicon ratings for each language sample were obtained by applying the VADER sentiment analysis and referencing the corpus library to obtain a single measure of sentiment across language samples. Of the 106 samples analyzed, 75% (80/106) obtained a positive compound score, indicating that the overall sentiment of speech across time was positive in nature (Figure 11). The remaining language samples in the dataset were detected to be neutral (16%, 17/106) or negative in sentiment rating (8.5% 9/106). Variance in compound scores for sentiment of transcribed messages was primarily observed in the 3-6 months post-stroke time period (Table 7). The participant demonstrated more negative sentiment in samples recorded during that initial three months post-stroke compared to six to nine months post-stroke. Additionally, positive sentiment was identified more often during the 3–6 month time period post-stroke (Figure 12).

Table 7. Sentiment of Language Samples Over Time

	Language Samples (n)	Positive	Neutral	Negative
1-3 Months	34	24 (70.6%)	2 (5.9%)	8 (23.5%)
3-6 months	49	36 (73.5%)	6 (12.2%)	7 (4.9%)
6-9 months	20	17 (85%)	1 (5%)	2 (10%)
9-12 Months	3	3 (100%)	0 (0%)	0 (0%)
TOTAL	106	80 (75.5%)	17 (16%)	9 (8.5%)

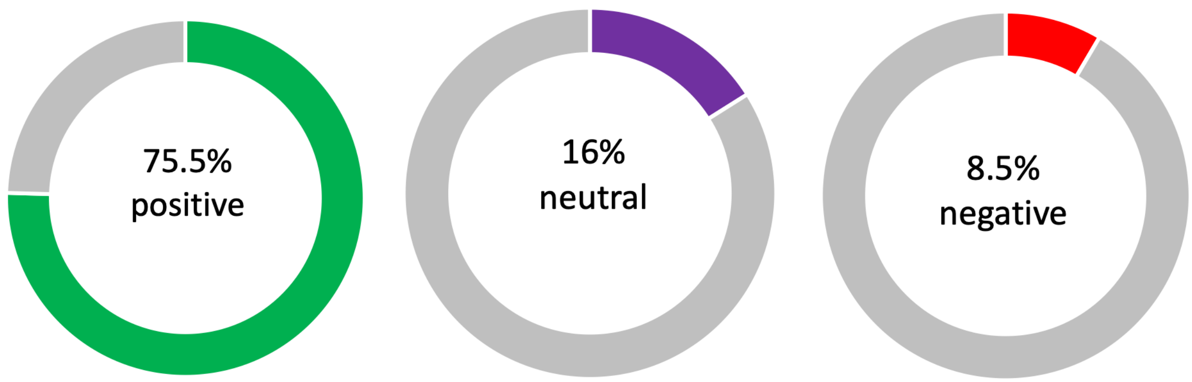

Figure 11. Percentage of positive, neutral, and negative sentiment of language samples based on compound scores

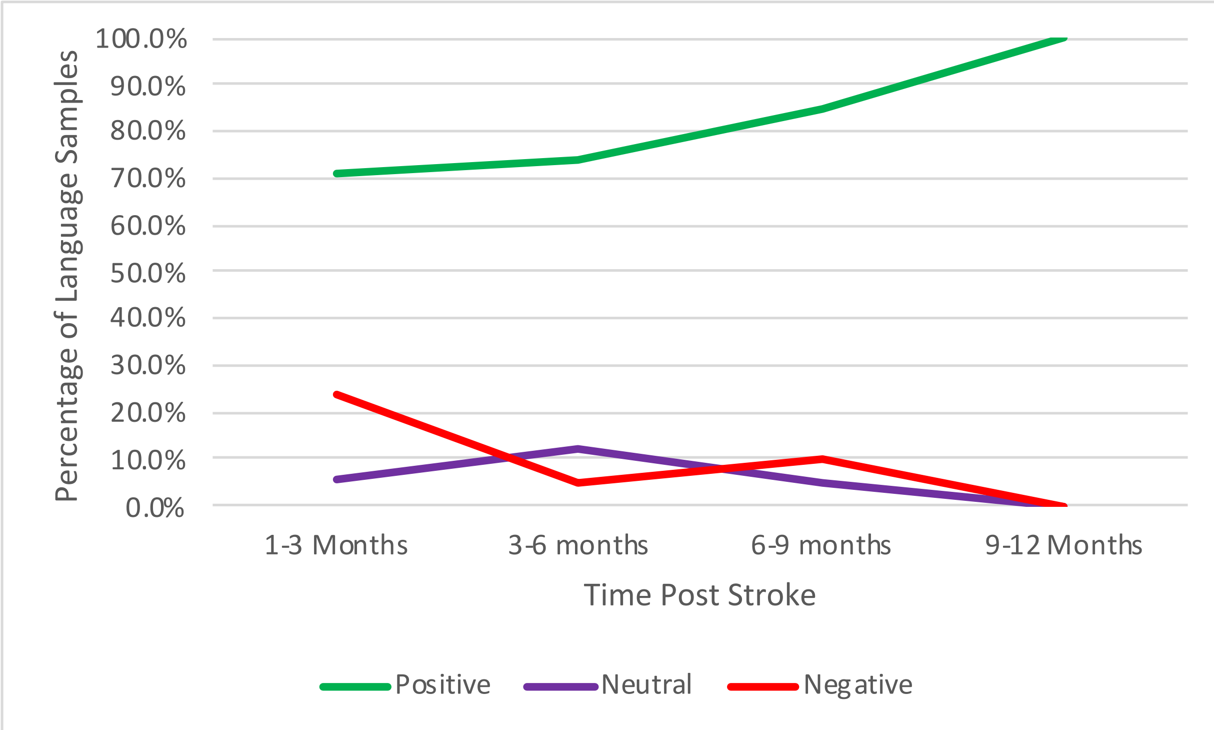

Figure 12. Percentage of positive, negative, and neutral sentiment in language samples based on compound scores at differing time points post-stroke.

CHAPTER 5: DISCUSSION

The purpose of this study was to conduct an exploratory study and apply a methodology for automated, programmatic analysis of transcribed post-stroke speech samples and assess change over time, addressing the following research aims: 1) to identify changes in the form and content of language as time post-stroke increases and 2) to apply programmatic natural language processing methods (sentiment analysis) to post-stroke speech data. No studies have previously examined the applicability of programmatic sentiment analysis on identifying change over time for post-stroke language samples. This preliminary investigation further aimed to utilize this methodology to assess feasibility and discuss future applications.

Current clinical recommendations cite that post-stroke language recovery is a longitudinal, dynamic process influenced by principles of neuroplasticity and driven by intervention and time (Allred et al., 2014; Gonzalez Rothi et al., 2008; Hier et al, 1983). While the most prominent of these neuroplastic changes are reported to occur in the immediate first days and weeks post-stroke, these changes continue to be documented past the acute phase of recovery into the subacute and chronic phases weeks, months and even years post-stroke (Kleim & Jones, 2008). Thus, analyzing changes that occur throughout the recovery timeline is important to support intervention processes that benefit language recovery as well as for measuring therapeutic outcomes (Moss et al., 2006).

Data from this study evaluated changes in form and content of language over time using Python NLTK POS tagging script and a preliminary methodology of VADER Sentiment Analysis to identify changes in sentiment over time in the immediate months post-stroke. Data from the present study provide insight into automated language processing of clinical datasets for discourse analyses and propose the applicability of using NLTK and VADER Sentiment Analysis to quantify aspects of language production with increased efficiency and accessibility.

Identifying Changes in Language Form & Content (Aim 1)

In this preliminary investigation, the average word count per language sample was highest in the 1-3 months post-stroke recovery period. This was an unexpected finding given the typical progression of change in language recovery that is often documented for patients following a stroke, in which noted gains in communication efficacy are typically observed in the immediate year post-stroke (Bakheit et al., 2007; Holland et al., 2017; Laska et al., 2001; Pickersgill & Lincoln, 1983, Stockert et al., 2016). For this participant, prompt, neuroprotective intervention through tPA administration may have mitigated the initial severity of language deficits and allowed for a recovery period with fewer significant deviations from his initial presenting baseline. It has additionally been discussed that the most critical period for spontaneous recovery of language is in the first weeks to month post-stroke (Gernstenecker & Lazar et al., 2019; Culton, 1969; Pederson et al., 2019). It is unclear whether this participant had a significant spontaneous language recovery in that initial window during his acute care admission, as the first language sample used for analysis was recorded 39 days post-stroke.

Furthermore, computer aided text analysis (CATA) allowed for the analysis of content words per sample and length of time for each sample. This measure was beneficial since language samples were of varying lengths of time. The number of words per language sample demonstrated a strong positive correlation to the length of time for the language sample, indicating some consistency in the verbal productions produced by the participant. Tracking the number of words per language sample and length of time needed to produce those words may be a relevant outcome measure for patients with reduced verbal expression post-stroke.

Interestingly, when stratifying language samples for analysis by time of day, the participant demonstrated a slightly increased, but non-significant, number of words per language sample in the morning compared to the afternoon. The increase in words during the morning recordings was correlated with length of speech sample, despite less samples being completed during the morning hours. This may be a relevant metric to consider for determining an ideal time of day to conduct therapy sessions. It has been acknowledged that intensive speech therapy is conducive to rehabilitative efforts post-stroke (Bhogal et al., 2003). Regarding patient-specific factors to treatment, patient motivation is an additional multifaceted construct that has been shown to influence therapeutic outcomes (Biel et al., 2018, Chapey et al., 2000; Shill, 1979). Following a patient-centered treatment approach, scheduling speech therapy sessions when verbal production is inherently higher or during a patient's preferred time of day could result in improved intrinsic motivation and overall willingness to actively engage in therapy, in turn potentially leading to increased verbal output.

The variability of non-words, filler words, and unintelligible utterances were minimal across time. Despite a minimal increase in fillers at 3-6 months post-stroke and 9-12 months post-stroke, the correlation between the number of fillers, non-words, and unintelligible utterances to time post-stroke, appeared non-significant. Further, the participant's use of fillers and non-words did not substantially decrease over time. The number of fillers were also directly correlated with language sample length. As language sample length increased, the number of fillers also increased at a similar rate. This indicates that fillers did not significantly increase or decrease over time but remained stable despite ongoing language rehabilitation. Any increase in observed number of fillers was likely related to increased language sample length. Thus, with more opportunities for verbal expression (i.e., as in a longer language sample recording), there was an increased likelihood of non-words, filler words, and unintelligible utterances to inundate this participant's speech, but the ratio at which these occurred as time post-stroke increased was consistent across language samples. Thus, no decrease in the use of fillers, non-words, or unintelligible utterance was overserved.

While past studies cite that the presence of neologisms or paraphasias, are more pronounced in earlier phases of recovery, it is not uncommon for these language impairments characteristic of aphasia to extend beyond 6-months post infarction (McKinnon et al., 2018). These are often indicative of some degree of word finding difficulty. Qualitative review of the language sample recordings and journal entries for this participant note some frustration with word finding and verbal output, which is consistent with the numeric data found. It remains unknown if fillers were higher in the immediately acute period following the stroke, as language sample collection did not begin until 30 days post-stroke.

In contrast, grammatic units utilized by the participant were diverse, including nouns, plurals, adjectives, adverbs, verbs in base form, gerund, and present tense, as well as prepositions. The frequency of grammatic units within each sample was additionally stable over time with minimal increases across language samples. The NLTK Python script successfully identified different grammatic units in the language samples for the participant with Broca's aphasia. Interestingly, this participant demonstrated a greater level of grammatic diversity than is typical for most patients with Broca's aphasia (Thompson et al., 2003). This may have been due to a relative strength in this area at baseline given the participant's age, educational background, tPA intervention, and/or severity of stroke, despite the Broca's aphasia diagnosis. Additionally, supports via therapeutic language intervention may have played a role in the diversity of grammatic units utilized in language samples. However, specific therapeutic goals were not available for analysis in the current dataset.

Within the grammatic analysis, however, tense markers, particularly those of past tense, were noted to be decreased compared to other grammatic units used by the participant within the dataset. In many languages, occurrence of events is expressed by marking tense of a verb. There is evidence that production of verb tense in sentences is more severely impaired than other functional categories in persons with agrammatic aphasia, such as Broca's aphasia (Miceli et al., 1984). Yet, the cause of verb-tense impairment remains unclear particularly related to semantic and processing demands (Faroqi-Shah & Friedman, 2015). Faroqi-Shah and Friedman (2015) found that verb tense impairment is exacerbated by processing demands of the elicitation task. This finding is consistent with data from the current study, given reductions in the use of past tense and past participle verb use noted in the samples and the inherent complexity of spontaneous speech production. Other grammatic units, such as WH-pronouns and possessive pronouns, were additionally markedly reduced in comparison to other grammatic units. This is likely due to the nature and context of the speech recordings (e.g., monologue recordings of events, activities, and feelings rather than multi-person conversational discourse).

According to Bryant et al. (2016a), two of the most frequently utilized computer-based language analysis software applications for discourse analysis in post-stroke populations include, Systematic Analysis of Language (SALT) and Computerized Language Analysis (CLAN), a part of the CHILDES (CHIld Language Data Exchange System), despite these programs original intent to be utilized in the investigation of child language disorders. Similar to this study's initiative to investigate the applicability of natural language processing methodologies to post-stroke language samples, the Northwestern Narrative Language Analysis System (NNLA), was developed to examine language deficits in Broca's aphasia (Thompson et al., 1995). Further aiming to quantify key, relevant aspects of connected speech in adult language disorders, the Quantitative Production Analysis (QPA) and subsequent C-QPA approach allow for faster analysis, with QPA being the most frequent measure of multiple linguistic structures for discourse analysis in aphasia (Bryant et al., 2016a; Fromm et al., 2021). A measure of grammatic diversity such as that derived from the Python NLTK POS script may be a complement to other means of language sample analyses and provide a means for development of normative measures for adult language outcomes.

From a clinical perspective, a measure of grammatic diversity could provide a quantitative baseline metric to determine progress for certain patients with aphasia to supplement analytic methods currently used. For example, patients with Broca's aphasia may benefit from this measure due to agrammatism frequently associated with this aphasia subtype (Kean, 1977). Additionally, it is known that sentence production difficulty in Broca's aphasia is characterized by difficulty producing distinct types of morphosyntactic structures, for example tense marking, relative to other structures such as agreement and mood marking, consistent with the data used for this analysis (Arabatzi & Edwards 2002; Clahsen, 2009). Analyzing baseline grammatic units for patients with global aphasia may also be beneficial due to the limited language output often seen with this aphasia subtype. Tracking grammatic diversity can allow any increases or improvements in the complexity and diversity of the grammatical structure of language samples to be documented after the initial, acute 0–3-month post-stroke period. Production of grammatical morphology is typically impaired in agrammatic aphasic individuals, as is their capacity to produce grammatic structures (Dickey & Thompson, 2007). Thus, this quantitative benchmark may support outcomes assessment for linguistically motivated treatment targeting grammatic deficits. Use of the Python NLTK script provides one automated and objective method for quantifying this area.

Analysis of Sentiment in Language Samples (Aim 2)

Sentiment of language samples was able to be objectively quantified over time. Sentiment additionally varied over time, in a quantifiable manner, suggesting that sentiment analysis using automated natural language processing methods could be considered in future prospective studies assessing patient-centered assessment and treatment developments for post-stroke rehabilitation. While all of the language samples of this subject were primarily positive in nature, a higher percentage of recordings at the 6-9 month post-stroke period achieved a positive compound sentiment score (85%) compared to the sentiment score of language samples during the initial 1-3 months post-stroke (70.6%). This is suggestive that, for this individual subject, the sentiment of recorded language samples increased as time post-stroke increased. These results suggest that sentiment of recorded language samples was varied during the rehabilitative process for this individual. This seems plausible given the dynamic trajectory of post-stroke language recovery profiles often seen among individuals (Denier et al., 2014; Gerstenecker & Lazar, 2019; Grefkes & Fink, 2020; Johnson et al., 2019; Lazar et al., 2008). Furthermore, sentiment is likely to wax and wane across time in a variant manner for the average person. In this single-subject investigation, sentiment showed a trend towards increased positivity as time post-stroke increased. It remains unknown if a positive sentiment is correlated with improvements language recovery measures. However, positivity has been previously associated with improved quality of life among individuals with adverse or chronic health conditions (Caprara et al., 2019; Higginson & Gao, 2008).

With improvements in health care, more people survive stroke, but many have to cope with the physical, psychological, social and functional sequelae due to language loss. Cerebral vascular accidents, such as stroke, have been reported to result in a significant deterioration of a patient's functioning and worsening of quality of life (Jaracz & Kozubski, 2003; Opara & Jaracz, 2010). The assessment of sentiment as a proxy for quality of life could facilitate identification of functional gains in the sequelae of language recovery as an indicator of the effectiveness of the post–stroke rehabilitation. As a positive psychological trait, individuals with higher resilience tend to hold a positive attitude, have a higher the level of adaptation, respond positively to disease, adjust to emotional distresses in a timely manner, and have improved quality of life (Zhang & Liu, 2019).

Furthermore, the ability to objectively and automatically quantify sentiment from transcribed language samples is novel and demonstrates potential for future clinical application. The use of NLP and VADER for post-stroke language analyses provided insights for change in language content and sentiment over time. These programmatic methods have the potential to address a number of current limitations in language analyses. A particular challenge often arises related to objectively quantifying elements in connected speech (Kintz & Wright, 2018). The challenge is multi-layered and not only involves skills necessary to transcribe a language sample, but also time to analyze it objectively, accurately, and consistently (Bryant et al., 2016a, b; Fromm et al., 2021). The use of VADER has the potential to solve one challenge related to objective and time-efficient analyses of sentiment of language samples. Comparisons between automated assessment of sentiment using VADER and manual assessment of sentiment demonstrated that VADER performed reliably in assigning sentiment to unstructured text data (e.g., social media posts), outperformed individual human raters, and generalized more favorably across contexts than other methods (Hutto & Gilbert, 2018; Watkins et al., 2020). Furthermore, analysis of sentiment using a validated, open-source library (VADER) as opposed to constructing and training a model and corpus exclusively for individual datasets may allow for increased accessibility and generalizability for future applications and interpretations.

Limitations

While a retrospective single-subject investigation allows for holistic, in-depth exploratory data analysis, case studies are by design limited in their generalizability to a given population. A prospective investigation that utilizes a larger sample size and is reflective of a wider spectrum of age, overall severity of deficits, and cultural background would be beneficial to enhance generalization of results and further define changes in sentiment and language content post-stroke. Future research on targeted demographics could add to the findings of this study.

Due to the design of data preprocessing and cleaning it is unknown if the non-words in the script were removed because those occurrences were in fact true neologisms, the production of an unintended sound within a word (paraphasia), or typographical transcription errors. In those instances, it is unclear if the transcription was related to how the word was pronounced or a transcription error. Semantic paraphasias, or substitutions of an actual word for an intended word, are another common presentation in aphasic speech (Kurowski & Blumstein, 2016). Since these substitutions are actual words and not neologisms, they were not included as non-words in the dataset despite them being unintended target productions. For bilingual or multilingual individuals with aphasia, code-switching is a typically observed behavior often to bypass word retrieval difficulty (Goral et al., 2019). While the subject in this study is a native English speaker, careful consideration should be given to future work with subjects from linguistically diverse backgrounds to ensure instances of code-switching are not coded as true nonwords or instances of neologisms.

Accuracy in reliably identifying paraphasias is another limitation resulting from this method of automated language analysis. Unlike neologisms, semantic or verbal paraphasias are incorrect words substituted for an intended word. Since semantic paraphasias for an English speaker are true English words, these are unlikely to be flagged using the automated preprocessing script as a true production error. It is unknown how the presence of semantic paraphasia impacted the subject's expressive language. Sarcasm or instances of verbal irony that are directly dependent on social context may go undetected or be falsely attributed to a "positive" sentiment utilizing automated machine learning models for sentiment analysis (Sykora et al., 2020; Farias & Rosso, 2017).

For this dataset in particular, the frequency of recordings was reduced overtime and was longitudinally inconsistent with no predetermined length of time between recordings. Best practice for future prospective studies would be to systematically record both structured and unstructured language samples at designated time points in a study period to avoid excessive gaps in data collection for more precise interpretation.

Additionally, the language samples utilized in this analysis were derived from voice recordings originally recorded not for the intention of this investigation. Therefore, utterances were not elicited in a systematic way using specific stimuli and instead were comprised of personal recounts of the subject's recovery and rehabilitative efforts. However, Brookshire and Nicholas (1994) suggested that personal narrative recounts have the ability to generate language samples that better represent actual language use. Therefore, the data derived from this study, having been exclusively personal narratives and recounts of events from the participant's daily life, are likely to be representative of typical language use for the participant. Future studies may benefit from including a combination of multiple structured and unstructured language samples including expositional picture descriptions, procedures, picture-supported narratives, and personal narrative recounts (Brookshire & Nicholas, 1994). Semi-structured interviews or predetermined, guided conversational topics could additionally be considered to elicit utterances for analysis and serve as a framework for reproducibility.

Clinical Implications & Future Directions

While the required time for data preprocessing utilizing NLTK is a current barrier to direct use in clinical populations, advancements and emerging research studying the efficacy of automated voice-to-text software with clinical data would significantly assist in streamlining this level of analysis into acute inpatient and outpatient rehabilitative settings (Bryant et al., 2016a). Additionally, this proposed method of analysis is unique in that it relies exclusively on open-source libraries, eliminating financial and additional barriers to accessibility. Sentiment analysis and POS tagging utilizing NLTK and VADER is not done by any individuals, but instead by the valid and reliable algorithm further adding to this preliminary investigation's replicability. Future research building on the principles of these data analyses will continue to add to the emerging body of evidence supporting the clinical utility of accessible approaches for adult, neurogenic populations.

Analyses utilized in this specific research methodology may further compliment advancements in linguistic discourse analysis for adults by providing an objective, supplementary form of assessment to enhance patient-centered service delivery. Across the scope of practice in speech-language pathology, discourse analysis is widely regarded as the preferred means of naturalistic assessment for communication efficacy, however the majority of well-regarded computer-based LSA tools are only a source of normative data for child language analysis (Ballard & Thompson, 1999; Bryant et al., 2016a; Pavelko & Owens, 2017; Miller et al., 2016). This present analysis has the potential to add to the growing means of comprehensive, multi-measure systems for discourse analysis and linguistic assessment for adults, such as QPA (Bryant et al., 2016a).

Application of sentiment analysis, in conjunction with other computer-assisted measures is one potential clinical application to allow for the identification of a patient-specific variable (sentiment of messages) that may otherwise remain overlooked on traditionally utilized psychometric and qualitative assessments. Objective measures that allow for understanding general trends in a patient's sentiment and outlook may benefit clinicians by identifying when additional psychosocial supports may be necessary in the recovery process. Future studies should assess the relationships between sentiment of verbal discourse, additional health-related quality of life metrics, and clinical outcome measures for language recovery for patients in the months to years post-stroke.

CHAPTER 6: CONCLUSION

Key findings from this study illustrate the potential utility of applying natural language processing methods, particularly sentiment analysis, to clinical datasets. Validation and application of natural language processing methods may ultimately improve discourse assessment procedures and identification of clinical outcome measures for patients with communication disorders. Specifically, in the field of speech language of pathology, findings from this exploratory study documented that sentiment analysis can be applied to clinically based language samples, which, with further rigorous study design and analyses, could serve as an enhanced metric for informed clinical decision making in post-stroke rehabilitation, given that sentiment of communicated methods may be correlated to key health-related quality of life measures. An objective, time-efficient, and accessible quality of life measure of the discourse of persons with aphasia at baseline and throughout treatment may provide significant insights for therapeutic outcomes assessment.

CHAPTER 7: REFERENCES

Ahmad, M., Aftab, S., Ali, I., & Hameed, N. (2017). Hybrid tools and techniques for sentiment analysis: A review. *International Journal of Multidisciplinary Sciences and Engineering, 8*(4), 28-33.

Alemi, F., Torii, M., Clementz, L., & Aron, D. C. (2012). Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. *Quality Management in Health Care*, *21*(1), 9–19. https://doi.org/10.1097/QMH.0b013e3182417fc4

Alharbi, A. S. M., & de Doncker, E. (2019). Sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cognitive Systems Research, 54(1)*, 60-61. https://doi.org/10.1016/j.cogsys.2018.10.001

Allred, R. P., Kim, S. Y., & Jones, T. A. (2014). Use it and/or lose it experience effects on brain remodeling across time after stroke. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00379

American Speech-Language-Hearing Association (n.d.). Aphasia (Practice Portal). Retrieved from www.asha.org/Practice-Portal/Clinical-Topics/Aphasia/.

Aphasia Statistics. (n.d.). *National Aphasia Association*. Retrieved July 21, 2019 from https://www.aphasia.org/aphasia-resources/aphasia-statistics/

Arabatzi, M., & Edwards, S. (2002). Tense and syntactic processes in agrammatic speech. *Brain and Language*, *80*(3), 314–327. https://doi.org/10.1006/brln.2001.2591

Armstrong, E., & Mortensen, L. (2006). Everyday Talk: its role in assessment and treatment for individuals with aphasia. *Brain Impairment*, *7*(3), 175–189. https://doi.org/10.1375/brim.7.3.175

Babbitt, E. M., & Cherney, L. R. (2010). Communication confidence in persons with aphasia. *Topics in Stroke Rehabilitation*, *17*(3), 214–223. https://doi.org/10.1310/tsr1703-214

Bakheit, A. M. O., Shaw, S., Carrington, S., & Griffiths, S. (2007). The rate and extent of improvement with therapy from the different types of aphasia in the first year after stroke. *Clinical Rehabilitation, 21(1)*, 941-949. https://doi.org/10.1177/0269215507078452

Ballard, K. J., & Thompson, C. K. (1999). Treatment and generalization of complex sentence production in agrammatism. Journal of Speech Language & Hearing Research, 42(3), 690–707.

Bhogal, S. K., Teasell, R. W., Foley, N. C., & Speechley, M. R. (2003). Rehabilitation of Aphasia: More Is Better. *Topics in Stroke Rehabilitation*, *10*(2), 66–76. https://doi.org/10.1310/RCM8-5TUL-NC5D-BX58

Biel, M., Nitta, L., & Jackson, C. (2018). Understanding motivation in aphasia rehabilitation. In P. Coppens & J. Patterson (Eds.), *Aphasia rehabilitation* (pp. 393-437). Jones & Bartlett Learning.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit.* O'Reilly Media.

Borglin, G., Edberg, A. K., & Hallberg, I. R. (2005). The experience of quality of life among older people. *Journal of aging studies*, *19*(2), 201-220.

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample-size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, 37(2), 399–407.

Brookshire, R. H., & McNeil, M. R. (2014). Introduction to Neurogenic Communication

Disorders (8th Ed.). Elsevier Health Sciences.

Bruce, C., & Edmndson, A. (2009). Letting the CAT out of the bag: A review of the Comprehensive Aphasia Test. Commentary on Howard, Swinburn, and Porter, "Putting the CAT out: What the Comprehensive Aphasia Test has to offer." *Aphasiology, 24(1)*, 79-93.

Brust, J. C., Shafer, S. Q., Richter, R. W., Brunn, B. (1976). Aphasia in acute stroke. *Stroke, 7*(2), 167-174. https://doi.org/10.1161/01.STR.7.2.167

Bryant, L., Ferguson, A., & Spencer, E. (2016a). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, *30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., & Ferguson, A. (2016b). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, *31*(10), 1105–1126. https://doi.org/10.1080/02687038.2016.1239013

Caprara, G. V., Alessandri, G., & Caprara, M. G. (2019). A response to commentaries on positivity. *Asian Journal of Social Psychology*, *22*(2), 146–150. https://doi.org/10.1111/ajsp.12361

Carey, L., Walsh, A., Adikari, A., Goodin, P., Alahakoon, D., De Silva, D., Ong, K.-L., Nilsson, M., & Boyd, L. (2019). Finding the Intersection of Neuroplasticity, Stroke Recovery, and Learning: Scope and Contributions to Stroke Rehabilitation. *Neural Plasticity*, *2019*, 1–15. https://doi.org/10.1155/2019/5232374

Carrillo-de-Albornoz, J., Rodríguez Vidal, J., & Plaza, L. (2018). Feature engineering for sentiment analysis in e-health forums. *PLOS ONE*, *13*(11), e0207996. https://doi.org/10.1371/journal.pone.0207996

Castelfranchi, Y. (2017). Computer-aided text analysis: an open-aired laboratory for social sciences. *Journal of Science Communication, 16*(2), 1-11. https://doi.org/10.22323/2.16020304

Chapey, R., Duchan, J. F., Elman, R. J., Garcia, L. J., Kagan, A., Lyon, J. G., & Mackie, N. S. (2000) Life participation approach to aphasia: A statement of values for the future. *The ASHA Leader. 5*(3), 4–6. https://doi.org/10.1044/leader.FTR.05032000.4.

Cherney, L. R., Babbitt, E. M., Semik, P., & Heinemann, A. W. (2011). Psychometric properties of the Communication Confidence Rating Scale for Aphasia (CCRSA): Phase 1. *Topics in Stroke Rehabilitation*, *18*(4), 352–360. https://doi.org/10.1310/tsr1804-352

Clahsen, H., & Ali, M. (2009). Formal features in aphasia: Tense, agreement, and mood in English agrammatism. *Journal of Neurolinguistics*, *22*(5), 436–450. https://doi.org/10.1016/j.jneuroling.2009.02.003

Coelho, C., Vlvisaker, M., & Turkstra, L. S. (2005). Nonstandardized assessment approaches for individuals with traumatic brain injuries. *Seminars in Speech and Language*, *26(4)*, 223-241. doi: 10.1055/s-2005-922102

Cramer, S. C., & Riley, J. D. (2008). Neuroplasticity and brain repair after stroke: *Current Opinion in Neurology*, *21*(1), 76–82. https://doi.org/10.1097/WCO.0b013e3282f36cb6

Croquelois, A., & Bogousslavsky, J. (2011). Stroke aphasia: 1,500 consecutive cases. *Cerebrovascular Diseases*, *31*(4), 392–399. https://doi.org/10.1159/000323217

Cruice, M., Worrall, L., Hickson, L., & Murison, R. (2003). Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being. *Aphasiology*, *17*(4), 333–353. https://doi.org/10.1080/02687030244000707

Culton, G. L. (1969). Spontaneous recovery from aphasia. *Journal of Speech and Hearing Research*, *12*(4), 825–832. https://doi.org/10.1044/jshr.1204.825

de Haan, R. J., Limburg, M., Van der Meulen, J. H. P., Jacobs, H. M., & Aaronson, N. K. (1995). Quality of life after stroke: Impact of stroke type and lesion location. *Stroke*, *26*(3), 402–408. https://doi.org/10.1161/01.STR.26.3.402

Denecke, K. (2015). *Health web science: Social media data in healthcare*. Springer International Publishing.

Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and Challenges. *Artificial Intelligence in Medicine, 64*(1), 17-27. https://doi.org/10.1016/j.artmed.2015.03.006

Denier, C., Flamand-Roze, C., Dib, F., Yeung, J., Solignac, M., de la Tour, L. B., Sarov-Rivière, M., Roze, E., Falissard, B., & Pico, F. (2015). Aphasia in stroke patients: Early outcome following thrombolysis. *Aphasiology, 29*(4), 442–456.

Dickey, M. W., & Thompson, C. K. (2007). The relation between syntactic and morphological recovery in agrammatic aphasia: A case study. *Aphasiology*, *21*(6–8), 604–616. https://doi.org/10.1080/02687030701192059

Doesborgh Suzanne J.C., van de Sandt-Koenderman Mieke W.E., Dippel Diederik W.J., van Harskamp Frans, Koudstaal Peter J., & Visch-Brink Evy G. (2004). Effects of Semantic Treatment on Verbal Communication and Linguistic Processing in Aphasia After Stroke. *Stroke*, *35*(1), 141–146. https://doi.org/10.1161/01.STR.0000105460.52928.A6

El Hachioui, H., Lingsma, H. F., van de Sandt-Koenderman, M. W. M. E., Dippel, D. W. J., Koudstaal, P. J., & Visch-Brink, E. G. (2013). Long-term prognosis of aphasia after stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(3), 310–315. https://doi.org/10.1136/jnnp-2012-302596

Ellis, C., & Urban, S. (2016). Age and aphasia: A review of presence, type, recovery and clinical outcomes. *Topics in Stroke Rehabilitation*, *23*(6), 430–439. https://doi.org/10.1080/10749357.2016.1150412

Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., & Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, *37*(6), 1379–1384. https://doi.org/10.1161/01.STR.0000221815.64093.8c

Farias, D. I. H., & Rosso, P. (2017). Irony, Sarcasm, and Sentiment Analysis. In *Sentiment Analysis in Social Networks* (pp. 113–128). Elsevier. https://doi.org/10.1016/B978-0-12-804412-4.00007-3

Faroqi-Shah, Y., & Friedman, L. (2015). Production of verb tense in agrammatic aphasia: A meta-analysis and further data. *Behavioural neurology*, *2015*.

Friel, K. M., Heddings, A. A., & Nudo, R. J. (2000). Effects of Postlesion Experience on Behavioral Recovery and Neurophysiologic Reorganization after Cortical Injury in Primates. *Neurorehabilitation and Neural Repair*, *14*(3), 187–198. https://doi.org/10.1177/154596830001400304

Fromm, D., Katta, S., Paccione, M., Hecht, S., Greenhouse, J., MacWhinney, B., & Schnur, T. T. (2021). A Comparison of Manual Versus Automated Quantitative Production Analysis of Connected Speech. *Journal of Speech, Language, and Hearing Research*, 1–12. https://doi.org/10.1044/2020_JSLHR-20-00561

Felberg, R. A., Okon, N. J., El-Mitwalli, A., Burgin, W. S., Grotta, J. C., & Alexandrov, A. V. (2002). Early dramatic recovery during intravenous tissue plasminogen activator infusion: Clinical pattern and outcome in acute middle cerebral artery stroke. *Stroke*, *33*(5), 1301–1307. https://doi.org/10.1161/01.STR.0000015556.48283.74

Gabriel, Z., & Bowling, A. (2004). Quality of life from the perspectives of older people. *Ageing and Society*, *24*(5), 675–691. https://doi.org/10.1017/S0144686X03001582

Gerstenecker, A. & Lazar, R. M. (2019). Language recovery following stroke. *The Clinical Neuropsychologist, 33*(5), 928-947. https://10.1080/13854046.2018.1562093

Gohil, S., Vuik, S., & Darzi, A. (2018). Sentiment analysis of health care tweets: Review of the methods used. *JMIR Public Health and Surveillance, 4*(2), e43. https://doi.org/10.2196/publichealth.5789

Goral, M., Norvik, M., & Jensen, B. U. (2019). Variation in language mixing in multilingual aphasia. *Clinical Linguistics & Phonetics*, *33*(10–11), 915–929. https://doi.org/10.1080/02699206.2019.1584646

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013a). Harnessing the cloud of patient experience: Using social media to detect poor quality healthcare: Table 1. *BMJ Quality & Safety*, *22*(3), 251–255. https://doi.org/10.1136/bmjqs-2012-001527

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013b). Use of sentiment analysis for capturing patient experience from free-text comments posted Online. *Journal of Medical Internet Research*, *15*(11), e239. https://doi.org/10.2196/jmir.2721

Grefkes, C., & Fink, G. R. (2020). Recovery from stroke: Current concepts and future perspectives. *Neurological Research and Practice*, *2*(1), 17. https://doi.org/10.1186/s42466-020-00060-6

Grissette, H., Nfaoui, E. H., & Bahir, A. (2017). Sentiment analysis tool for pharmaceutical industry & healthcare. *Transactions on Machine Learning and Artificial Intelligence*, *5*(4). 746-760. https://doi.org/10.14738/tmlai.54.3339

Gonzalez Rothi, L. J., Musson, N., Rosenbek, J. C., & Sapienza, C. M. (2008). Neuroplasticity and Rehabilitation Research for Speech, Language, and Swallowing Disorders. *Journal of Speech, Language, and Hearing Research*, *51*(1). https://doi.org/10.1044/1092-4388(2008/017)

Hamilton, R. H. (2016). Neuroplasticity in the language system: Reorganization in post-stroke aphasia and in neuromodulation interventions. *Restorative Neurology and Neuroscience*, *34*(4), 467–471. https://doi.org/10.3233/RNN-169002

Harvey, R. L. (2015). Predictors of functional outcome following stroke. *Physical Medicine and Rehabilitation Clinics, 26(4)*, 583-598. https://doi.org/10.1016/j.pmr.2015.07.002

Heilmann, J., & Westerveld, M. (2013). Bilingual language sample analysis: Considerations and technological advances. *Journal of Clinical Practice in Speech-Language Pathology, 15*(2), 87–93.

Hersh, D., Wood, P., & Armstrong, E. (2017). Informal aphasia assessment, interaction and the development of the therapeutic relationship in the early period after stroke. *Aphasiology, 32(8)*, 876-901

Hier, D. B., Mondlock, J., & Caplan, L. R. (1983). Recovery of behavioral abnormalities after right hemisphere stroke. *Neurology*, *33*(3), 345–345. https://doi.org/10.1212/WNL.33.3.345

Higginson, I. J., & Gao, W. (2008). Caregiver assessment of patients with advanced cancer: Concordance with patients, effect of burden and positivity. *Health and Quality of Life Outcomes*, *6*(1), 42. https://doi.org/10.1186/1477-7525-6-42

Hilari, K., Byng, S., Lamping, D. L. & Smith, S. C. (2003). Stroke and aphasia quality of life scale-39 (SAQOL-39) - Evaluation of acceptability, reliability, and validity. Stroke, 34(8), pp. 1944-1950. doi: 10.1161/01.STR.0000081987.46660.ED

Hillis, A. E., Beh, Y. Y., Sebastian, R., Breining, B., Tippett, D. C., Wright, A., Saxena, S., Rorden, C., Bonilha, L., Basilakos, A., Yourganov, G., & Fridriksson, J. (2018). Predicting recovery in acute poststroke aphasia: Aphasia recovery. *Annals of Neurology*, *83*(3), 612–622. https://doi.org/10.1002/ana.25184

Hillis, A. E. (2007). Aphasia: progress in the last quarter of a century. *Neurology, 69*(2), 200-213. https://doi.org/ 10.1212/01.wnl.0000265600.69385.6f

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science, 349*(6245). 261-266. doi: 10.1126/science.aaa8685

Hoffman, M., & Chen, R. (2013). The spectrum of aphasia subtypes and etiology in subacute stroke. *Journal of Stroke and Cerebrovascular Disease, 22*(8), 1385-1392. http://dx.doi.org/10.1016/j.jstrokecerebrovasdis.2013.04.017

Holland, A., Fromm, D., Forbes, M, & MacWhinney, B. (2017). Long-term recovery in stroke accompanied by aphasia: A reconsideration. *Aphasiology, 31(2)*, 152-165. https://doi.org/10.1080/02687038.2016.1184221

Hubel, D. H., & Wiesel, T. N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology, 28*, (6), 1041–1059. doi: 10.1152/jn.1965.28.6.1041

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Jacquin, A., Virat-Brassaud, M.-E., Rouaud, O., Osseby, G.-V., Aboa-Eboulé, C., Hervieu, M., Ménassa, M., Ricolfi, F., Giroud, M., & Béjot, Y. (2014). Vascular aphasia outcome after untravenous recombinant tissue plasminogen activator thrombolysis for ischemic stroke. *European Neurology*, *71*(5–6), 288–295. https://doi.org/10.1159/000357428

James, T. L., Villacis Calderon, E. D., & Cook, D. F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications*, *71*, 479–492. https://doi.org/10.1016/j.eswa.2016.11.004

Janssen, H., Bernhardt, J., Collier, J. M., Sena, E. S., McElduff, P., Attia, J., Pollack, M., Howells, D. W., Nilsson, M., Calford, M. B., & Spratt, N. J. (2010). An Enriched Environment Improves Sensorimotor Function Post-Ischemic Stroke. *Neurorehabilitation and Neural Repair*, *24*(9), 802–813. https://doi.org/10.1177/1545968310372092

Jaracz, K., & Kozubski, W. (2003). Quality of life in stroke patients. *Acta Neurologica Scandinavica*, *107*(5), 324–329. https://doi.org/10.1034/j.1600-0404.2003.02078.x

Johnson, L., Basilakos, A., Yourganov, G., Cai, B., Bonilha, L., Rorden, C., & Fridriksson, J. (2019). Progression of aphasia severity in the chronic stages of stroke. *American Journal of Speech-Language Pathology, 28*(2), 639-649. https://doi.org/10.1044/2018_AJSLP-18-0123

Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, *4*(1), 9. https://doi.org/10.1186/s13388-015-0024-x

Kazmierska, J., & Malicki, J. (2008). Application of the naïve byesian classifier to optimize treatment decisions. *Radiotherapy and Oncology*, *86*(2), 211–216. https://doi.org/10.1016/j.radonc.2007.10.019

Kean, M.-L. (1977). The linguistic interpretation of aphasic syndromes: Agrammatism in Broca's aphasia, an example. *Cognition, 5*(1), 9–46. https://doi.org/10.1016/0010-0277(77)90015-4

Kertesz, A., & Poole, E. (1974). The aphasia quotient: the taxonomic approach to measurement of aphasic disability. *Canadian Journal of Neurological Sciences, 1(1),* 7-16. doi: 10.1017/S031716710001951X

Kintz, S., & Wright, H. H. (2018). Discourse measurement in aphasia research. *Aphasiology*, *32*(4), 472–474. https://doi.org/10.1080/02687038.2017.1398807

Klabunde, R. (2007). Cardiovascular Pharmacology Concepts. Retrieved from http://www.cvpharmacology.com/thrombolytic/thrombolytic.

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research*, *51*(1). https://doi.org/10.1044/1092-4388(2008/018)

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text classification for organizational researchers: A tutorial. *Organizational Research Methods*, *21*(3), 766–799. https://doi.org/10.1177/1094428117719322

Koenig-Bruhin, M., Kolonko, B., At, A., Annoni, J., & Hunziker, E. (2013). Aphasia following a stroke: recovery and recommendations for rehabilitation. *Swiss Archives of Neurology and Psychiatry*, *164*(8), 292-298. doi:10.4414/sanp.2013.00209

Kremer, C., Perren, F., Kappelin, J., Selariu, E., & Abul-Kasim, K. (2013). Prognosis of aphasia in stroke patients early after iv thrombolysis. *Clinical Neurology and Neurosurgery*, *115*(3), 289–292. https://doi.org/10.1016/j.clineuro.2012.05.019

Kurowski, K., & Blumstein, S. E. (2016). Phonetic basis of phonemic paraphasias in aphasia: Evidence for cascading activation. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *75*, 193–203. https://doi.org/10.1016/j.cortex.2015.12.005

Kwah, L. K., & Diong, J. (2014). National institutes of health stroke scale (NIHSS). *Journal of Physiotherapy, 60(1)*, 61. http://dx.doi.org/10.1016/j.jphys.2013.12.012

Laska, A. C., Hellblom, A., Murray, V., Kahan, T., & Von Arbin, M. (2001). Aphasia in acute stroke and relation to outcome. *Journal of Internal Medicine, 249(5)*, 413-422. https://doi.org/10.1046/j.1365-2796.2001.00812.x

Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery, and Psychiatry, 79(5)*, 530-534. doi: 10.1136/jnnp.2007.122457

Leedham, B., Meyerowitz, B. E., Muirhead, J., & Frist, W. H. (1995). Positive expectations predict health after heart transplantation. *Health Psychology*, *14*(1), 74.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems 48(1)*, 354-368.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*(1), 113–124. https://doi.org/10.1044/jshd.5401.113

Ludlow, C. L., Hoit, J., Kent, R., Ramig, L. O., Shrivastav, R., Strand, E., Yorkston, K., & Sapienza, C. M. (2008). Translating principles of neural plasticity into research on speech motor control recovery and rehabilitation. Journal of Speech, Language, and Hearing Research, *51*(1), S240-S258. doi: 10.1044/1092-4388(2008/019)

Lyden, P. (2017). Using the national institutes of health stroke scale: a cautionary tale. *Stroke. 48*(2), 513-419. doi: 10.1161/STROKEAHA.116.015434

Marsh, E. B., Lawrence, E., Gottesman, R. F., & Llinas, R. H. (2016). The NIH stroke scale has limited utility in accurate daily monitoring of neurologic status. *Neurohospitalist, 6(3)*, 97-101. doi: 10.1177/1941874415619964

Martins, I. P., Fonseca, J., Morgado, J., Leal, G., Farrajota, L., Fonseca, A. C., & Melo, T. P. (2017). Language improvement one week after thrombolysis in acute stroke. *Acta Neurologica Scandinavica*, *135*(3), 339–345. https://doi.org/10.1111/ane.12604

McCallum, A., & Nigam, K. 1998. A comparison of event models for Naive Bayes text classification. In *Proc. AAAI-98 Workshop on Learning for Text Categorization*, pages 41-48. AAAI Press.

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced test in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders*, *49*(4), 338–348. https://doi.org/10.1044/jshd.4904.338

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

McKinnon, E. T., Fridriksson, J., Basilakos, A., Hickok, G., Hillis, A. E., Spampinato, M. V., Gleichgerrcht, E., Rorden, C., Jensen, J. H., Helpern, J. A., & Bonilha, L. (2018). Types of naming errors in chronic post-stroke aphasia are dissociated by dual stream axonal loss. *Scientific Reports*, *8*(1), 14352. https://doi.org/10.1038/s41598-018-32457-4

McNeil, M. R., & Pratt, S. R. (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. *Aphasiology, 15*(10/11), 901-911. https://doi.org/ 10.1080/02687040143000276

Meiner, Z. et al (2010). Rehabilitation outcomes of stroke patients treated with tissue plasminogen activator. *Physical Medicine and Rehabilitation, 2*(8), 698-702.

Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M. S., Schoppmann, A., & Zook, J. M. (1984). Somatosensory cortical map changes following digit amputation in adult monkeys. *The Journal of Comparative Neurology*, *224*(4), 591–605. https://doi.org/10.1002/cne.902240408

Miceli G., Silver M. C., Villa G., Caramazza A. On the basis for the agrammatic's difficulty in producing main verbs. Cortex. 1984;20(2):207–220. doi: 10.1016/S0010-9452(84)80038-6.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2016). Using language sample analysis to assess spoken language production in adolescents. *Language, Speech, and Hearing Services in Schools*, *47*(2), 99–112. https://doi.org/10.1044/2015_LSHSS-15-0051

Moss, A., & Nicholas, M. (2006). Language rehabilitation in chronic aphasia and time post onset: A review of single-subject data. *Stroke*, *37*(12), 3043–3051. https://doi.org/10.1161/01.STR.0000249427.74970.15

Murray, L. & Coppens, P. (2013, Formal and informal assessment of aphasia. In I. Papathanasiou, P. Coppens and C. Potagas (Eds), *Aphasia and Related Neurogenic Communication Disorders* (pp. 67–91). Jones and Bartlett Learning.

Nudo, R. J., Milliken, G. W. (1996). Reorganization of movement representations in primary motor cortex following focal ischemic infarcts in adult squirrel monkeys. *Journal of Neurophysiology, 75*(5), 2144-2149. https://doi.org/10.1152/jn.1996.75.5.2144

Opara, J., & Jaracz, K. (2010). Quality of life of post–stroke patients and their caregivers. *Journal of Medicine and Life*, *3*(3), 216–220.

Paul, D., Frattali, C., Holland, A., Thompson, C., Caperton, C., & Slater, S. (2004).

Quality of Communication Life Scale. American Speech-Language Hearing Association. Rockville, MD.

Pauranik, A., George, A., Sahu, A., Nehra, A., Paplikar, A., Bhat, C., Krishnan, G., Kaur, H., Saini, J., Suresh, P. A., Ojha P., Singh, P., Sancheti, P., Karanth, P., Mathuranath, P. S., Goswami, S., Chitnis S., Sundar, N., Alladi, S., Faroqi-Shah, Y. Expert group meeting on aphasia: A report. *Annals of Indian Academy of Neurology, 22*(2), 137-146. https://doi.org/10.4103/aian.AIAN_330_18

Paltoglou, G. (2016). Sentiment-based event detection in Twitter. *Journal of the Association for Information Science and Technology*, *67*(7), 1576-1587. doi: 10.1002/asi.23465

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1-135.

Pavelko, S. L., & Owens, R. E. (2017). Sampling utterances and grammatical analysis revised (SUGAR): New normative values for language sample analysis measures. *Language,* *Speech, and Hearing Services in Schools*, *48*(3), 197–215. https://doi.org/10.1044/2017_LSHSS-17-0022

Payabvash, S., Kamalian, S., Fung, S., Wang, Y., Passanese, J., Kamalian, S., Souza, L. C. S., Kemmling, A., Harris, G. J., Halpern, E. F., González, R. G., Furie, K. L., & Lev, M. H. (2010). Predicting language improvement in acute stroke patients presenting with aphasia: A multivariate logistic model using location-weighted atlas-based analysis of admission CT perfusion scans. *American Journal of Neuroradiology*, *31*(9), 1661–1668. https://doi.org/10.3174/ajnr.A2125

Pedersen, P., Vinter, K., & Olsen, T. (2004). Aphasia after stroke: type, severity, and prognosis. *Cerebrovascular Diseases, 17(1)*, 35-43. doi: 10.1159/000073896

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V., & Holzinger, A. (2013, July). Opinion mining on the web 2.0–characteristics of user generated content and their impacts. In *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 35-46). Springer, Berlin, Heidelberg.

Pierangelo, R., & Giuliani, G. A. (2016). Assessment in special education—A practical approach (5th Ed.). Pearson.

Pickersgill, M. J., & Lincoln, N. B. (1983). Prognostic indicators and the pattern of recovery of communication in aphasic stroke patients. *Journal of Neurology, Neurosurgery & Psychiatry*, *46*(2), 130–139. https://doi.org/10.1136/jnnp.46.2.130

Plowman, E., Hentz, B., & Ellis Jr., C. (2012). Post-stroke aphasia prognosis: a review of patient-related factors. *Journal of Evaluation in Clinical Practice, 18(1)*, 689-694. doi: 10.1111/j.1365-2753.2011.01650.x

Price, L. H., Hendricks, S., & Cook, C. (2010). Incorporating computer-aided language sample analysis into clinical Practice. *Language, Speech, and Hearing Services in Schools*, *41*(2), 206–222. https://doi.org/10.1044/0161-1461(2009/08-0054)

Priya, C., Santhi, K., & Durairaj Vincent, P. M. (2019). Provision of efficient sentiment analysis for unstructured data. In S. C. Satapathy, V. Bhateja, R. Somanah, X.-S. Yang, & R. Senkerik (Eds.), *Information Systems Design and Intelligent Applications* (pp. 199–207). Springer Singapore.

Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., Murray, L., Rose, M., Thompson, C. K., Turkstra, L., Altmann, L., Boyle, M., Conway, T., Hula, W., Kearns, K., Rapp, B., Simmons-Mackie, N., & Gonzalez Rothi, L. J. (2008). Translational research in aphasia: From neuroscience to neurorehabilitation. *Journal of Speech, Language, and Hearing Research*, *51*(1). https://doi.org/10.1044/1092-4388(2008/020)

Robey, R. R. (1998). A meta-analysis of clinical Outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research*, *41*(1), 172–187. https://doi.org/10.1044/jslhr.4101.172

Robbins, J., Butler, S. G., Daniels, S. K., Diez Gross, R., Langmore, S., Lazarus, C. L., Martin-Harris, B., McCabe, D., Musson, N., & Rosenbek, J. (2008). Swallowing and dysphagia rehabilitation: Translating principles of neural plasticity into clinically oriented evidence. *Journal of Speech, Language, and Hearing Research*, *51*(1). https://doi.org/10.1044/1092-4388(2008/021)

Rojas, R., & Iglesias, A. (2010). Using language sampling to measure language growth. *Perspectives on Language Learning and Education, 17*(1), 24–31. https://doi.org/10.1044/lle17.1.24

Saif, H., Fernandez, M., He, Y., & Alani, H. (2013). Evaluation datasets for twitter sentiment analysis. *Emotion and Sentiment in Social and Expressive Media*, 9.

Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for Plasticity in White-Matter Tracts of Patients with Chronic Broca's Aphasia Undergoing Intense Intonation-based Speech Therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 385–394. https://doi.org/10.1111/j.1749-6632.2009.04587.x

Shill, M. A. (1979). Motivational factors in aphasia therapy: Research suggestions. *Journal of Communication Disorders*, *12*(6), 503–517. https://doi.org/10.1016/0021-9924(79)90013-3

Shultz, J. R. (2009). Aphasia classification and assessment [PowerPoint slides]. Retrieved from: http://neurology.mcgill.ca/neurodocs/AHD%20Robilllard%20Shultz%20SLP%20lecture%20Apr%2015.pdf

Simmons-Mackie, N., Threats, T. T., & Kagan, A. (2005). Outcome assessment in aphasia: A survey. *Journal of Communication Disorders*, *38*(1), 1–27. https://doi.org/10.1016/j.jcomdis.2004.03.007

Stockert, A., Kümmerer, D., & Saur, D. (2016). Insights into early language recovery: From basic principles to practical applications. *Aphasiology*, *30*(5), 517–541. https://doi.org/10.1080/02687038.2015.1119796

Stockman, I. J. (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools*, *27*(4), 355–366. https://doi.org/10.1044/0161-1461.2704.355

Sykora, M., Elayan, S., & Jackson, T. W. (2020). A qualitative analysis of sarcasm, irony and related #hashtags on Twitter. *Big Data & Society*, *7*(2), 2053951720972735. https://doi.org/10.1177/2053951720972735

Thomson, J., Gee, M., Sage, K., & Walker, T. (2018). What "form" does informal assessment take? A scoping review of the informal assessment literature for aphasia. *International Journal of Language & Communication Disorders*, *53*(4), 659-675. doi: 10.1111/1460-6984.12382

Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L., & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language, 51*, 124–129.

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research: JSLHR*, *46*(3), 591–607.

Thompson, C.K. (2019). Neurocognitive recovery of sentence processing in aphasia. *Journal of Speech, Language, and Hearing Research, 62*(11), 3947-3972. https://doi.org/10.1044/2019_JSLHR-L-RSNP-19-0219

Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. *IISA 2013*, 1–6. https://doi.org/10.1109/IISA.2013.6623713

Vega-Bermudez, F., & Johnson, K. O. (2002). Spatial acuity after digit amputation. *Brain*, *125*(6), 1256–1264. https://doi.org/10.1093/brain/awf129

Wade, D. T., Hewer, R. L., David, R. M., & Enderby, P. M. (1986). Aphasia after stroke: Natural history and associated deficits. *Journal of Neurology, Neurosurgery & Psychiatry*, *49*(1), 11–16. https://doi.org/10.1136/jnnp.49.1.11

Wardlaw, J.M., Murray, V., Berge, E., & del Zoppo, G.J. (2014). Thrombolysis for acute ischemic stroke (Review). Cochrane Database of Systematic Reviews, 7. doi: 10.1002/14651858.CD000213.pub3

Watkins, J., Fabielli, M., & Mahmud, M. (2020). SENSE: A Student performance quantifier using sentiment analysis. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–6. https://doi.org/10.1109/IJCNN48605.2020.9207721

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLOS ONE*, *13*(2), e0192773. https://doi.org/10.1371/journal.pone.0192773

Zhang, L., & Liu, B. (2017) Sentiment analysis and opinion mining. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*, (pp.1152-1161). Springer.