

Spring 6-3-2015

Training and Application of Correct Information Unit Analysis to Structured and Unstructured Discourse

Audrey Bretthauer Cohen
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

 Part of the [Discourse and Text Linguistics Commons](#), [Other Communication Commons](#), and the [Speech and Hearing Science Commons](#)

Recommended Citation

Cohen, Audrey Bretthauer, "Training and Application of Correct Information Unit Analysis to Structured and Unstructured Discourse" (2015). *Dissertations and Theses*. Paper 2339.

10.15760/etd.2336

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Training and Application of Correct Information Unit Analysis to Structured and
Unstructured Discourse

by

Audrey Bretthauer Cohen

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Speech and Hearing Sciences

Thesis Committee:
Gerasimos Fergadiotis, Chair
Maria Kapantzoglou
Sarah Key-DeLyria

Portland State University
2015

Abstract

Correct Information Units (CIU) analysis is one of the few measures of discourse that attempts to quantify discourse as a function of communicating information efficiently. Though this analysis is used reliably as a research tool, most studies' apply CIUs to structured discourse tasks and do not specifically describe how raters are trained. If certified clinical speech-language pathologists can likewise reliably apply CIU analysis within clinical settings to unstructured discourse, such as the discourse of people with aphasia (PWA), it may allow clinicians to quantify the information communicated efficiently in clinical populations with discourse deficits. **Purpose:** The purpose of this study is to determine if using the outlined training module, clinicians are able to score CIUs with similar inter-rater reliability across both structured and unstructured discourse samples as researchers. **Method:** Four certified SLPs will undergo a two-hour training session in CIU analysis similar to that of a university research staffs' CIU training protocol. Each SLP will score CIUs in structured and unstructured language samples collected from individuals diagnosed with aphasia. The SLPs' scores within the structured and unstructured discourse samples will be compared to those of a university research lab staffs'. This will determine (1) whether SLPs can reliably code CIUs when compared with research raters in a lab setting when both using the same two-hour CIU training and resources allotted; (2) whether there is a significant difference in reliability when structured and unstructured discourse is analyzed.

To Sean, Mom, Dad, and Jefferson.

ACKNOWLEDGMENTS

This project would not have been possible without the support and guidance of many of my colleagues and mentors. I would like to thank my advisor, Gerasimos Fergadiotis for encouraging me to ask questions and design this study. I would also like to thank my committee members, Maria Kapantzoglou and Sarah Key-DeLyria for their optimism, supportive advice and thought provoking discussions. This study would not have been possible without the help of Jade Horton, fellow researcher, colleague and friend as well as the time and resources of the Aging and Adult Language Disorders Research Laboratory.

Thank you.

TABLE OF CONTENTS

	Page
ABSTRACT.....	i
DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES.....	v
CHAPTER	
1 LITERATURE REVIEW	
Correct Information Unit Analysis of Connected Discourse.....	1
Challenges of Current Measures of Discourse.....	2
Measures of Discourse as a Function of Information Transfer.....	7
Content Units.....	7
Information Units.....	9
Correct Information Unit Analysis.....	10
Purpose of Study.....	16
2 METHOD	
Participants.....	18
Procedures.....	18-20
3 RESULTS	
Analysis.....	21
4 DISCUSSION	
Limitations & Future Directions.....	27
References.....	31

LIST OF TABLES

TABLE

Page

TABLE 1: Pearson Product Moment Correlations & Confidence Intervals.....21

CHAPTER 1

Connected Information Unit Analysis of Connected Discourse

Evaluating the ability of people with aphasia (PWA) to convey information at the discourse level has been of interest to aphasiologists and clinicians alike. When considering the wide spectrum of conversational abilities of PWA, it is important to consider the communicative function of discourse as a tool for conveying information or ideas. In addition, the ideal language samples to measure are those that are elicited within contexts that are closely representative of PWAs' day-to-day conversations (e.g., using unstructured discourse tasks). Nicholas & Brookshire's Correct Information Unit Analysis or CIU analysis (1993) has been shown to be a potentially effective and reliable method of quantifying the transference of information within structured discourse tasks within research studies. In contrast, most other measures of discourse fail to measure the *functional* purpose of discourse-to communicate ideas or information effectively and efficiently. These measures either measure *structural* aspects of discourse (e.g., syntactic or phonological well-formedness) or in their attempts to measure *functional* aspects of discourse most measures inadvertently rely on narrow or vaguely defined discourse variables, and are therefore often subject to random error, or rater/scorer bias.

If CIU analysis can be applied to less structured language samples reliably, it has the potential to be a quantifiable tool for measuring PWAs' functional connected discourse. However, CIU analysis of structured discourse is more commonly used as a measure in research and not within therapeutic settings to measure treatment outcomes. Additionally, few studies describe their protocols for training their CIU raters, other than

adhering to Nicholas and Brookshire's "published rules for CIUs" (Doyle, Tsironas, Goda & Kalinyak, 1996, p. 55). Not only is there conflicting evidence regarding the ability to establish inter-rater reliability (Doyle, Goda & Spencer, 1995; Oelslaeger & Thorne, 1999) when applying CIUs to unstructured discourse but there has yet to be a replicable CIU training protocol. The main purpose of this study is to determine if certified SLPs are able to reliably code CIUs across discourse type (i.e., structured & unstructured) after receiving a CIU instructional protocol.

Challenges of Current Measures of Discourse

Many current assessments of discourse include measures of connected discourse either through evaluation of linguistic structure such as grammatical accuracy, fluency and complexity of clauses or, alternatively, by rating variables deemed representative of functional aspects of discourse such as the accuracy, relevance and/or informativeness (E.g., WAB-R, SSLA, Quantitative Assessment of Narrative Speech, Multi-Level Method of Discourse Quantification) (Kertesz, 2007; Shewan, 1988; Saffran et al., 1989; Marini, et al., 2011). However, each of these approaches often does not adequately quantify the primary function of discourse, which is to convey information and ideas. Those assessments which rate selected functional aspects of discourse often rely on vague definitions of these variables, possibly resulting in an overreliance on raters' subjective opinion. Contrastingly, structural analyses of discourse, such as rating a speaker's grammatical accuracy or syntactical complexity, do not systematically measure the *efficiency* of their connected discourse as a communicative tool. Despite grammatical flaws, paraphasias or circumlocutions, a person may very effectively and efficiently

convey relevant ideas and information. To illustrate this, current measures of discourse will be reviewed and evaluated as measures of discourse's primary purpose, conveying ideas and information.

In an attempt to measure both functional and structural aspects of discourse, the WAB-R's spontaneous speech subtest relies on rating vaguely defined functional variables or structural aspects of PWA's discourse instead of measuring discourse as a function of information transfer. PWA are asked a series of scripted interview questions such as "Have you been here today?" (Kertesz, 2007). The administrator records the accuracy of the information conveyed by the client's responses (i.e., whether the information they convey is deemed correct or incorrect by the rater) as well as the fluency and grammatical accuracy. The picture description subtest for the same battery includes asking the client to describe the Picnic Scene and then rating their responses based on accuracy, fluency, presence of circumlocution, and possible indicators of word-finding difficulties (i.e., paraphasias) (Kertesz, 2007). The PWA's responses to both the interview questions and picture are rated using an 11 point rating scale (0-10) where 1 represents "no information" and 10 represents the PWA having "correct" responses to all 6 questions. This method assumes that for a response to be informative, it must be correct. However, regardless of accuracy, there may still be a high proportion of information conveyed as a function of the amount of language used. Further, candidacy for a higher rating requires responses to be of "normal length and complexity" or "reasonably complete description(s) of the picture" (Saffran et al., 1989). These qualifying factors, such as relying on a rater's definition of "normal" are subjective,

making this method of evaluating structural and function aspects of discourse susceptible to random error and/or systematic bias on the part of the rater.

Similarly, the “Conversational and Expository Speech” subtest of the BDAE-3 evaluates simple social responses, free conversation, and picture description tasks also based on vague definitions of discourse function such as “appropriateness” as well as structural aspects of discourse syntax. A 7-point rating scale is used to measure the length and accuracy of linguistic units such as agrammatic deletions, or the complexity of clauses (Goodglass & Kaplan, 1983). The BDAE-3’s “Conversational and Expository Speech” subtest includes two subcategories: “Simple Social Responses” and “Free Conversation.” The “Simple Social Responses” subcategory includes evaluating each response as “appropriate” or not. Again, this relies on raters’ varying definitions of “appropriate.” Despite provision of two to three examples of “appropriate responses it is likely that even these examples are representative of demographic bias. Further, the BDAE-3’s “Free Conversation” subcategory includes the facilitator asking questions encouraging at least three minutes of conversation. Questions include asking about the PWA’s occupation, for example. Analysis of the resulting language sample includes quantifying the amount of simple to complex clauses, agrammatic deletions, and total number of utterances (Goodglass & Kaplan, 1983). While this method may evaluate the syntactic complexity of the PWA’s connected discourse it does not measure the amount of information communicated as a function of verbal output.

The CETI, in contrast to the WAB-R and BDAE-3 is much more focused on evaluating the *functional* characteristics of discourse, though, like most measures, the

CETI's rating scales are reliant on the subjective opinion of the rater's definitions of the discourse variables. Using a 0-10 point scale, the investigator rates the PWAs' functional language production within activities of daily living (Lomas, Pickard, Bester, Elbard, Finlayson, & Zoghaib, 1989). According to Lomas et al. "to ensure that the communication situations that would be rated were representative of patients' values and daily-living activities, we elicited situations from aphasic individuals themselves," (1989, p. 114). The authors had stroke survivors and their spouses select scenarios in which the caregivers felt was fundamental for the PWA to communicate their needs effectively, and where they must be able to understand the information communicated by their caregivers (Lomas et al., p. 115). Based on these selected scenarios, 16 items were created where caregivers rated the PWA's abilities in communication scenarios such as "getting somebody's attention," or "responding to or communicating anything," (Lomas et al., p. 123). The PWA's ability to perform these tasks is rated by a close family or caregiver, thereby taking into consideration the opinions of common communication partners (Lomas et al., 1989). In an attempt to capture the effectivity of PWA's connected discourse within contexts other than structured discourse tasks, methods of subjectively *rating* the efficiency of connected discourse, such as using the CETI, may be subject to random error and systematic bias. The rating scales used by caregivers to evaluate the PWA's communicative abilities range from "not at all able" to "as able as before stroke." Since the CETI relies on the subjective rating of caregivers who know the PWA, their ratings may reflect extraneous variables other than the efficiency of the PWA's verbal output. For example, the caregiver's ability to effectively remember the PWA's

communicative abilities before their CVA may vary from day-to-day. Such systematic and random bias can result in PWA's caregivers over or underestimating the PWA's abilities.

In an effort to define a procedure that evaluates both the structural and functional characteristics of narrative discourse, Marini, Andretta, Tin and Carlomagno (2011) developed a multi-level method of discourse quantification of two individuals diagnosed with fluent aphasia. However, like the previously mentioned measures of discourse Marini et al.'s multi-level approach relies on either measures of vaguely defined variables of functional discourse or structural features of syntax and phonology. They defined two different approaches to linguistic analysis; the "structuralist approach," where different constituents of processing such as lexical, phonological and grammatical are analyzed, and the "functionalist approach" which focused on the ability of PWA to convey information through discourse (Marini et al., 2011). Combining these approaches, Marini et al., designed a series of measures to analyze the four main aspects of linguistic processing: informativeness, narrative organization, grammatical processing and lexical productivity (Marini et al., 2011, p. 1379). Their measures of the informativeness of two PWA's discourse relied on the percentage of lexical informative speech units (LIUs) per minute (%LIUs/min) where LIUs are defined as language content words that are phonologically, and grammatically well-formed and appropriate as well as socially pragmatic (Marini et al., 2011, p. 1383). Marini et al.'s quantification of LIUs differs from previously mentioned measures in its attempt to highlight *functional* units of discourse. It is similar, however, in again quantifying units of lexical-semantic accuracy

and structural characteristics such as grammatical complexity instead of units of informativeness; Marini et al.'s definition of LIUs being "informative" relies on them being "phonologically well formed," as well as "appropriate from a grammatical and pragmatic point of view." Non-LIUs were further defined as any semantic or verbal paraphasias, paragrammatic errors, or words included in "tangential or conceptually incongruent utterances" (Marini, et al., 2011, p. 1389). This definition of informativeness has a few limitations; firstly, defining what is and is not an LIU are both subjective. A rater's opinion of which lexical units are deemed pragmatic, appropriate, tangential or conceptually incongruent may differ from one rater to another. More importantly, though not phonologically well-formed, a lexical unit may still convey ideas or information and therefore be representative of an effective discourse unit especially in languages that as opposed to Italian are not highly inflected like English. Specifically, not including any paraphasias in the LIU count is problematic; semantic and verbal paraphasias with enough linguistic context may still convey relevant and accurate information. Therefore, Marini et al.'s information unit method may disregard units of connected discourse produced by PWA that may effectively, and/or efficiently communicate information or ideas.

Measures of Discourse as a Function of Information Transfer

Content Units

Most of the previous methods of quantifying discourse rely on grammatical accuracy, fluency and/or linguistic complexity and/or vaguely defined rating variables of functional discourse. These methods do not measure discourse as a function of its

primary purpose to convey ideas and information. In contrast, Yorkston and Beukelman's "Content Unit" (CU) system, used by such measures as the Shewan Spontaneous Language Analysis (SSLA), quantifies units deemed informative based on the semantic content of the elicitation stimuli, however, this measure is subject to systematic bias due to its definition of a CU being based on a selected neuro-typical populations' utterances. Further, the samples produced and analyze for this method must be elicited through structured discourse tasks which may not be representative of PWA's connected discourse across other contexts. The extent of this semantic content is preemptively defined by "a grouping of information that was always expressed as a unit by normal speakers in structured discourse tasks" (Yorkston & Beukelman, 1980, p. 30). This method relies on picture description elicitation tasks where the content is both predictable and consistent based on previously compiled lists of words produced by neuro-typical adults to describe the pictures (Yorkston & Beukelman, 1980 p. 27). Firstly, this method assumes that the content units produced by the selected neuro-typical adults represent the *only* appropriate content units that can be used to describe the elements in the pictures; depending on the demographic characteristics of the selected participants in the neuro-typical group, variables such as age, gender, culture, education and dialect may result in their connected discourse differing from that of the PWAs being assessed. Secondly, like most of the previous methods, this method measures PWAs' discourse elicited through *structured* discourse tasks, which may not be as representative of their ability to effectively and/or efficiently communicate ideas or content within less structured contexts. Further, in using a picture description task, as is done by both the SSLA and

Yorkston and Beukelman, the CU method lacks expressive language parameters (Shewan, 1988, p. 113). In addition, since one cannot preemptively compile lists of relevant content units for free-speech, it is unlikely that this system could be used to quantify content within free-speech samples, or discourse with no shared references such as a picture.

Information Units

Another method that quantifies informative units of discourse is McNeil et al.'s Information Unit (IU) Scoring Metric which has the benefit of being applied to structured discourse tasks without relying on preemptively comprised lists like Yorkston & Beukelman's CU system (McNeil, Doyle, Fossett, Park and Goda, 2010). However, as with all of the previous methods, McNeil et al.'s IU system is limited in that it must be applied to structured discourse tasks, allowing for provision of a joint-referent during story-retells. As previously mentioned, discourse within a context with a joint referent may not represent day-to-day connected discourse. McNeil et al.'s study evaluated the reliability and concurrent validity of the IU Scoring Metric as applied to discourse elicited through the Story Retelling Procedure protocol defined by Doyle and McNeil et al. (2010). Participants retold stories after having listened to, and watching computerized images associated with events in four parallel stories (Doyle et al., 2000). McNeil et al., defined an IU as a "word, (or) phrase of acceptable alternative from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story" (2010, p. 994). Recorded language samples from 15 PWA and 31 normal individuals were analyzed using the IU metric. McNeil et al.'s IU scoring metric has the

benefit of being applied to structured discourse tasks without relying on preemptively comprised lists like Yorkston & Beukelman's CU system. This allows for "acceptable alternatives" to story content and prevents some bias in basing an information system on a particular population's utterances (McNeil et al., 2010, p. 994). Although McNeil and Doyle et al.'s story retelling procedure is *less* structured than the picture description discourse tasks used by Yorkston & Beukelman (1980), they still rely on having content units being previously identified based on the ideas and content of the selected stories (Doyle et al., 2000, McNeil et al., 2010). Further, the digital images used to tell the story appear on the computer screen for participants' to use as a reference during their re-tells (Doyle et al., 2000; McNeil et al., 2010). Having a relevant visual reference available is less common in day-to-day connected discourse. Again, use of the story retelling procedure is, therefore, less than representative of the unpredictable nature of conversation or free-speech.

In addition to relying on measuring syntactical or functional variables not representative of discourse as a function of the transference of ideas or information, all of the previous system measures of discourse rely on measuring discourse within *structured* discourse tasks. Measuring discourse within less structured discourse tasks has the benefit of allowing a clinician to evaluate numerous characteristics of both expressive and receptive language within a multitude of different structured communicative contexts.

Correct Information Unit Analysis

The above measures rely on measuring syntactical or functional variables not representative of discourse as a function of the transference of ideas or information.

Correct Information Unit (CIU) analysis, one of the few measures of discourse that reliably quantifies discourse as a function of information transfer, has been shown to be relatively reliable and stable across repeated measures as well as have high inter-rater reliability (Nicholas & Brookshire, 1993). As a result, CIU analysis has become a common means of quantifying the efficiency of *structured* discourse. In its advent, Nicholas and Brookshire defined CIUs as

Words that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or the topic. Words do not have to be used in a grammatically correct manner to be included in the correct information count. Each correct information unit consists of a single word and only words that have been included in the word count can be considered for inclusion in the correct information unit count (p. 357, Appendix B).

In their original study of the efficacy of CIU analysis, discourse elicitation stimuli included picture descriptions of two pictorial sequences and four single pictures, two requests for procedural information and two requests for personal information (Nicholas & Brookshire, p. 340). Two groups of participants' (20 healthy, and 20 adults diagnosed with aphasia) discourse were recorded and analyzed using the CIU system. Nicholas and Brookshire identified *words* as intelligible in context, but not necessarily relevant, and then calculated identified CIUs as percent CIU (%CIU), CIUs per minute (CIU/min), and words per minute (WPM) (Nicholas & Brookshire, p. 343). Nicholas & Brookshire's study found high inter-rater reliability for words and CIUs for participants with aphasia

(>95% for both) (1993, p. 343). Further, they found relative stability from session to session, with the standard error of measurement percentage of change varying no more than 3.2-8.2% (Nicholas & Brookshire, 1993, p. 345). These findings suggest that the CIU scoring system may be a reliable system both in repeated measures and when scored by different raters.

Since Nicholas & Brookshire's original study of CIUs, there has been evidence suggesting that CIU scores adequately reflect listeners' perceptions of functional speech (Doyle, Tsironas, Goda & Kalinyak, 1996). In their study, Doyle et al. compared the objective measures of 11 unfamiliar listeners of the informativeness of the connected discourse of 25 PWA in a story retelling task (1996). The 25 PWAs' connected discourse was elicited through Nicholas and Brookshire's four single pictured, two picture sequences, and two requests for procedural information (1993), the "cookie theft" picture from the BDAE-R, (Goodglass & Kaplan 1983), and the "picnic" picture from the Western Aphasia Battery (WAB Kertesz, 1982). All elicited discourses were audio-recorded. Using Steven's scaling theory (1946), Doyle et al. segmented the listeners' perceptual rating scale of informativeness into equal intervals that are linearly related to direct magnitude estimates (DME) (Doyle et al., 1996 p.55) Their results showed a moderately high correlation regression between the perceptual ratings of informativeness made by the judges and the %CIUs ($r=.81$), and CIUs/min ($r=.85$) (Doyle et al., 1995 p. 58-59, figure 5). These findings suggest that %CIUs and CIUs/min reflect objective listeners' opinion of PWAs' discourse informativeness. Assuming that these judges' ratings reflect the opinions of common discourse partners in PWAs' communities, this

may suggest that CIU analysis effectively measures the informativeness, and therefore, efficiency of PWAs' connected discourse.

Despite being a reliable metric, the CIU metric is restricted in that it was designed to be applied only to structured discourse tasks, as outlined by Nicholas & Brookshire (1993). Further, while most studies report using the same training procedure as Nicholas and Brookshire's study for CIU raters, most of them also report using SLPs who were co-authors, or research associates as raters rather than full-time clinicians (Nicholas & Brookshire, 1993; Doyle et al., 1996; Cameron et al., 2010). In fact, the only study that explicitly describes using clinical SLPs as raters was Oelschlaeger & Thorne's (1999) study, and their reported inter- and intra-rater reliability was less than acceptable. Therefore, a replicable CIU training program that includes collaboration and open discussion of CIU agreements/disagreements has yet to be designed and evaluated. This study will evaluate SLP's ability to established inter-rater reliability across structured and unstructured discourse after undergoing a structured CIU training program.

Studies suggest that CIU analysis be reliably used to measure structured discourse. However, evidence suggests that PWAs' discourse may be more informative and efficient within unstructured connected discourse. This suggests that CIU scores found using Nicholas & Brookshire's procedure may not adequately represent the informativeness of PWAs' discourse within less structured tasks. In Doyle et al.'s study in 1995, for example, twenty structured (following procedures of Nicholas and Brookshire, 1993) and twenty conversational speech samples were collected from subjects with aphasia. Conversational discourse was elicited through both "topic-open,"

and “topic-constrained” methods. “Topic-open” elicitation included recording a 7-minute conversation where participants were told that they could discuss any topic of their choosing (Doyle et al., 1995, p. 131). The “topic-constrained” method of discourse elicitation involved having the PWA and their caregivers watch a 4.5-minute news segment and then recording the subsequent discussion (Doyle, et al., 1995, p. 131). Doyle et al.’s results showed that while the number of words and CIUs did not differ between the structured and conversational elicitation tasks, subjects *did* produce a greater %CIU in the conversational discourse elicitation tasks (Doyle et al., 1995, p. 132). These findings also suggest that PWA may elicit more efficient connected discourse within unstructured elicitation tasks further, this evidence implies that the potential information transfer of connected discourse in PWA may not be adequately represented by discourse elicited within structured tasks.

There have been mixed findings with regard to establishing acceptable inter-rater agreement for CIU identification within both structured and unstructured discourse elicitation tasks. Considering the unpredictable nature of unstructured discourse, CIU identification may become more variable under less structured circumstances or without provision of a shared referent. However, Doyle et al.’s point-to-point inter-rater agreement for CIU identification in “topic-open” discourse tasks was largely acceptable (>85%) (Doyle et al., 1995 p. 132). This implies that CIU analysis may be applied to unstructured discourse reliably.

In contrast, Oelschlaeger and Thorne’s study in 1999 found inter-rater reliability to be more volatile when identifying CIUs within unstructured discourse contexts (1999).

Their investigation had a CIU count inter-reliability of less than 56% and an intra-rater reliability of less than 73% (p. 641, Oelschlaeger & Thorne, 1999). In contrast to Doyle et al.'s results, these findings suggest that the ability to establish acceptable inter- and intra-rater reliability when scoring CIUs is a more variable process when scoring unstructured discourse. However, Oelschlaeger et al.'s study's training of the CIU system differed from that of Nicholas and Brookshire in that raters did not have an opportunity to undergo an initial training together with a subsequent practice and discussion of CIU agreements and disagreements (1993, p. 5). Instead, Oelschlaeger et al.'s study included only provision of a packet outlining Nicholas and Brookshire's protocol to each rater as a resource while scoring the transcripts privately (1999). Likewise, it was not reported how much time the raters had to dedicate to learning to identify CIUs. Given the degree to which full-time *clinical* SLPs may have less time to dedicate to learning to score CIUs, this may be problematic when compared with SLPs dedicated to research investigations. Further, a comparison of raters' CIU analysis of structured discourse with the reported *unstructured* discourse was not described. However, Oelschlaeger and Thorne's study is one of the few studies of CIU analysis that described their CIU raters as being clinical certified SLPS versus co-authors or research associates or staff (Oelschlaeger & Thorne, 1999 p. 640). Oelschlaeger and Thorne's findings contrast those of Doyle et al.; however it is unclear whether this can be attributed to the amount and quality of the CIU training protocol participants received before scoring transcripts, or if CIU identification really is more variable when applied to less structured discourse. Further research is necessary to

better understand the ability to establish acceptable inter-rater reliability for CIU identification within unstructured discourse elicitation tasks.

Additional evidence suggests that CIU identification may be more variable between repeated measures, even when using Nicholas and Brookshire's structured discourse elicitation protocol (1993). Cameron et al.'s study found that %CIU and CIUs/min varied in repeated presentations and that the percent change for participants with aphasia was never lower than the reported 3% found by Nicholas & Brookshire (1993; Cameron et al., 2010). This may suggest that, even when using structured discourse tasks, CIU analysis is more variable between repeated measures.

Yet, measuring discourse within less structured discourse tasks has the benefit of allowing a clinician to evaluate numerous characteristics of connected discourse within a multitude of different structured communicative contexts. While structured discourse tasks allow for the content of the discourse to be predictable, these methods of quantifying different aspects of communication do not necessarily account for the efficiency of PWAs' ability to convey information efficiently in day-to-day connected discourse. Unstructured discourse samples would be more representative of free-form conversations within PWAs' homes and communities. Free-speech connected discourse is more representative of real-life conversations, where the content is not always predictable, there is not a consistent shared reference and the topic may shift at any time.

Purpose of the Study

The purpose of this investigation is to determine if certified SLP full-time clinicians with the same training as a university research staff can likewise reliably score CIUs

when analyzing structured (narratives elicited using pictorial stimuli) and free-structured (conversational speech elicited using open-ended questions). In contrast to previous studies, this study will select full-time *clinical* certified SLP CIU raters (instead of researchers) with the same training in CIU analysis that a research team uses to analyze data for discourse studies and as per the protocol outlined in Nicholas and Brookshire's study (1993). This study will determine (1) whether clinical SLPs can reliably code CIUs when compared with a gold standard after a two-hour CIU training; and (2) whether there is a significant difference in reliability between analyses of structured and unstructured discourse. Based on the current evidence, and that the definition of the informativeness of units of speech may be more finite when the content is consistent and predictable such as in structured discourse (Yorkston & Beukelman, 1980), we predict that the correlations will be lower for the SLP raters when scoring unstructured language samples when compared with the gold standard.

CHAPTER 2

Method

Participants

Raters. Four certified SLPs, experienced in providing therapy to individuals with aphasia (having worked with PWA for at least one year), from the Portland, Oregon, city area, were recruited to participate in this study. They were compensated for their time financially. Participants were trained to identify CIUs in an initial two-hour training session by a speech pathology graduate student assistant researcher accustomed to the CIU analysis system. Following this training, participants performed their analyses of the thirty transcripts over a period of five two-hour scoring sessions.

Procedures

Discourse Samples. Thirty discourse samples, fifteen structured and fifteen free-speech samples collected from 15 PWA were randomly selected from AphasiaBank, an online shared database of digital recordings of the discourse of PWA across a series of tasks (AphasiaBank; MacWhinney, Fromm, Forbes., Holland, 2011). All selected transcripts were from participants who have aphasia secondary to a left hemisphere stroke. Selected transcripts were from PWA who have previously met the following inclusion criteria: (a) English as their primary language; (b) corrected or uncorrected normal visual acuity; (c) aided or unaided hearing acuity; (d) no reported history of psychiatric or neurodegenerative disorders; and chronic aphasia (minimum=6 months post onset). All transcripts were elicited from PWA who were administered the Western Aphasia Battery-Revised (Kertesz, 2007), the Boston Naming Test (Goodglass & Kaplan,

1983) and several subtests from the Reading Comprehension Battery for Aphasia, Second Edition (LaPointe & Horner, 1998). Based on these assessments, the selected transcripts included eight diagnosed with fluent aphasia, and seven with non-fluent aphasia.

Discourse Elicitation. Language samples were based on four picture descriptions and one story retell designed to elicit narrative discourse including a four-panel picture sequence entitled “*Broken Window*” (Menn, 1995), a six-panel picture sequence entitled “*Refused Umbrella*,” a single picture called “*Cat Rescue*” (Nicholas & Brookshire, 1993), a single photograph depicting an emergency rescue of a girl in a flood (Rubin & Newton, 2001) and a request to tell the Cinderella fairytale after viewing a picture story book (Grimes, 2005). Elicitation will have included the investigator presenting the picture(s) and saying “Here is a picture. Look at everything that’s happening and then tell me a story about what you see. Tell me the story with a beginning, middle, and an end.” Fifteen unstructured discourse samples were collected that were elicited via minimally structured “free speech” conversational questions. These consisted of participants responding to two open-ended questions about participants’ experiences recovering from their stroke as well as life events both pre- and post-stroke.

Transcription. Orthographically transcribed transcripts from each PWA were parceled into structured and unstructured subcomponents. All codes for gestures (i.e., pointing, waving, nodding) were removed from the transcripts.

CIU Gold Standard. A research team (two graduate students), trained and experienced in scoring CIUs with acceptable inter-rater reliability, scored the discourse samples for CIUs. Their inter-rater reliability was determined to be 90% by one certified

SLP. The scorers then discussed and resolved all disagreements line-by-line in the 30 transcripts. Their resolutions were then reviewed for accuracy by the certified SLP. A final CIU score for each transcript was then established by the three researchers based on the research teams' discussions. These final CIU scores were used as the set standard number per transcript with which to compare the raters' scores.

CIU Training & Scoring. The four recruited participants underwent a two-hour training in CIU analysis based on the established CIU scoring guidelines outlined by Nicholas and Brookshire (1993) that consisted of a power-point presentation outlining rules for CIU identification followed by a guided scoring practice. As part of the guided scoring, participants scored two sample transcripts (one structured and the other unstructured) and then discussed their errors with the facilitating graduate student. Participants were also encouraged to ask questions after the initial training. After their initial training and guided practice, participants were asked to identify and quantify the #CIUs in each line of thirty (fifteen structured and fifteen unstructured) transcripts in a quiet environment (either their home or a quiet room located on the Portland State University campus near research lab) separately. They were each given five packets containing six transcripts per scoring session total. Each packet also included the picture stimuli used for elicitation of structured discourse samples. Type of discourses distributed for each session will be counterbalanced. Participants were given only one packet at a time, and instructed to spend no more than 120 minutes to score the entire packet. Once finished scoring a packet of transcripts the participants received the next packet. No less than two days break was permitted between scoring sessions.

CHAPTER 3

Results

Analysis

Pearson product moment correlations were estimated between the range of values of participants' and the Gold Standard CIU scores within structured and unstructured discourse samples. CIU scores were positively and significantly correlated with the CIU Gold Standard ($p < .05$) with averaged correlations of .91 for both CIU scores made within structured and unstructured discourse tasks (range = 0.88-0.95 for structured and range = 0.80-0.97 for unstructured), (Table 1). Table 1 also shows that 95% Confidence Intervals (CIs) around the point of estimates. These findings suggest that using this training raters were able to score CIUs across discourse type similarly and more importantly with high reliability. The overlapping CIs of correlations suggest that any difference in participants' CIU scores within structured and unstructured discourse samples was not statistically significant.

Table 1.
Pearson Product-Moment Correlations of the Gold Standard and Participants' CIU scores within structured and unstructured samples.

Participant ID	GS Structured CIUs (N=15)	GS Unstructured CIUs (N=15)	95% CI	
			Structured	Unstructured
Participant 1's CIU scores	0.91	0.94	0.73-0.97	0.83-0.98
Participant 2's CIU scores	0.95	0.97	0.86-0.99	0.91-0.99
Participant 3's CIU scores	0.91	0.91	0.74-0.97	0.74-0.97
Participant 4's CIU scores	0.88	0.80	0.66-0.96	0.48-0.93

Note. CI=Confidence Intervals
GS=Gold Standard
CIU= Correct Information Unit

A priori power analysis was run to determine the minimum sample size (number of scored transcripts) necessary to find a significant difference between the correlations found within structured and unstructured discourse scores. This was used to determine how large a sample size is necessary to ensure that any deviation from the null hypothesis ($H_0: \rho(\text{structured correlations}) = \rho(\text{unstructured correlations})$) is detected. Assuming that (1) the above correlations correspond to the true population correlations and (2) the differences estimated between CI are likewise accurate a two dependent Pearson r 's Correlation with no common index run in G Power was run to determine how large the sample must be to make sure that this deviation from the null is detected with a power of $1-\beta=.95$ using a two-tailed test and $\alpha=.05$. The power analysis indicated that $N=141$ is the minimum required sample size using the correlations from Participant 1 to corroborate the difference found between confidence intervals. Since the sample size is a function of the magnitude of the difference and absolute values of the correlations, we can infer that the sample sizes needed for the other participants will need to be larger than the $N=141$ computed. This suggests that the difference (or lack thereof) between confidence intervals for the correlations between structured and unstructured discourse sample scores may reflect the difference that might be found in a sample size of 141 or larger.

CHAPTER 4

Discussion

The purpose of this study was to evaluate the efficacy of a CIU training module through trained raters' ability to reliably score CIUs across structured and unstructured discourse samples. CIU analysis was originally designed for quantification of informativeness within structured and not unstructured discourse tasks (Nicholas & Brookshire, 1993). As previously mentioned, CIU analysis applied to unstructured discourse would be a valuable tool for evaluating PWAs' discourse efficiency within a wide variety of settings. This investigation attempts to establish a training protocol for CIU analysis that trains raters to reliably score CIUs within structured and unstructured discourse.

Our investigation intended (i) to determine whether, after using the same two-hour training, raters could reliably score CIUs within structured and unstructured discourse, and (ii) whether there was a significant difference in reliability between analyses of structured and unstructured discourse. Support for the effectiveness of the current CIU training module is reflected in the high degree to which raters' scores were correlated with the Gold Standard CIU scores across discourse type. Further, any difference between CIU correlations between structured and unstructured discourse samples were found to be statistically insignificant due to the overlapping confidence intervals (see Table 1). That is the raters scored CIUs reliably regardless of whether within a discourse sample elicited through structured or unstructured tasks. Given these

findings, this training may adequately prepare clinicians to score CIUs across discourse type.

It is well documented that the ability to produce connected discourse is most relevant to PWA's quality of life and thereby, generalization goals of treatment (Cruice, Worrall, Hickson & Murison, 2003). Cruice et al.'s study in 2003 found evidence suggesting that the assessment of discourse is closely representative of communication within activities of daily living where performance is closely related to the participation level of the International Classification of Functioning, Disability and Health (ICF) and subsequently clients' overall quality of life. Additionally, Mayer and Murray's findings, for instance, revealed that PWA had increased word retrieval and self-corrected errors in connected speech when compared with confrontational word naming tasks (2010). CIU analysis, therefore, captures aspects of communication at the activity level of the ICF and is more closely related to the participation level which typically is the ultimate generalization goal of treatment.

Further, in contrast to many current measures of discourse, CIU analysis has the benefit of quantifying discourse as a function of the information conveyed. While many other approaches quantify grammatical accuracy, fluency, complexity of clauses or, rely on obscure rating variables subject to systematic biases, CIU analysis may be one of the few methods of quantifying discourse based on its function of conveying ideas. Further, unlike other methods of quantifying informative units, CIU analysis does not require preemptively composing lists of informative units that may be included in a discourse sample (Yorkson & Beukelman, 1980). Though modern technological tools, such as

crowdsourcing, where such wordlists could be compiled and modified continuously in an online database for clinicians to use as a source, such tasks would still require discourse elicitation tasks based on somewhat predictable concrete stimuli versus unstructured discourse. Likewise, many other methods of quantifying discourse informative units such as McNeil's Information Unit (IU) analysis also rely on structured discourse such as story retells (McNeil et al., 2010). However, CIU analysis is more commonly applied to discourse elicited through structured discourse tasks. Since structured discourse is less representative of PWA's day-to-day speech, CIU analysis as applied to unstructured discourse would be a valuable research and clinical tool.

Since its advent in 1993, CIU analysis has been applied primarily within structured discourse and solely for research purposes (Nicholas & Brookshire, 1993). Further, when training raters, researchers are restricted in designing a training program based on their interpretations of the appendix of Nicholas and Brookshire's 1993 study. Therefore, a standardized CIU analysis training program that follows Nicholas and Brookshire's guidelines while also allowing for rater discussion would benefit researchers in preparing staff for CIU analysis. Further, the current findings suggest that with this particular training, CIU analysis may be reliably applied across structured and unstructured discourse tasks allowing researchers and clinicians alike to quantify and evaluate PWAs' functional discourse across elicitation tasks.

Our findings contrast that of previous evidence which suggests that CIU reliability becomes more volatile within conversational settings (Oelschlager and Thorne, 1999). Where Oelschlager and Thorne's case study of CIU analysis applied to the free-

speech samples of a PWA had low reliability, the correlations in our study between participants' scores and the gold standard CIU scores averaged at .91 (range = 0.88-0.95 for structured and range = 0.80-0.97 for unstructured) (Table 1). One possible reason for this contrasting reliability may be attributed to the difference in the rater training protocols. Where Oelschlager and Thorne gave participants a packet including a transcript and written instructions for CIU analysis, this study included a two-hour training in CIU analysis designed based on the protocol of a university research laboratory's protocol for training graduate students in CIU analysis. Further, instruction included guided practice where participants could discuss agreements and disagreements and the nature of their errors before scoring the selected language samples. This training session and opportunity to openly discuss the rules of CIU analysis may have attributed to the high correlations between participants' scores and the Gold Standard.

The results of this study and previous evidence suggest that not only is reliability in CIU scoring minimally affected by the structured nature of the discourse, but also that application of CIU analysis to unstructured discourse may provide more adequate information about PWA's functional discourse. The CIs of the correlations of participants' and the gold standard CIU scores overlapped, suggesting that any difference was statistically insignificant. This corroborates previous findings such as in Doyle et al.'s study where point-to-point inter-rater reliability between structured and conversational tasks for CIU scores was largely similar (1995). Further, when considering Doyle et al.'s finding that PWA produced CIUs more efficiently within conversational discourse, application of CIU analysis to unstructured discourse may more

adequately reveal the competency of PWA within day-to-day conversation (1995). Given the results of this study, as well as those of Doyle et al.'s, CIU analysis may not only be reliably applied to unstructured discourse, using an adequate training module, but this application has the potential to provide clinicians and researchers more information about the functional discourse of PWA.

Limitations & Future Directions

Though these findings suggest that CIU analysis can be reliably applied across structured and unstructured discourse using this outlined training, further evidence is necessary to determine the extent to which inter-rater reliability may vary across discourse type. One such limitation of this study was the nature of the CIU scoring. In an attempt to reflect the protocol of a research staff's, participants were instructed to score CIUs at the end of each transcript line. This provided less information than if participants had identified which specific words were and were not CIUs. In future investigations that latter approach would allow for analysis of the number of and type of errors (false positives vs false negatives). A correlation of the different types of errors and the type of discourse could help determine whether there is a tendency to over or under identify CIUs within structured versus unstructured discourse samples.

An additional limitation to this study was that when compared with a research staff, the participants had less time to continually discuss disagreements and understanding of CIU identification rules (Nicholas & Brookshire 1993). Where a research staff has the benefit of scoring CIUs within a setting where there are other researchers available for collaboration, the participants in this study scored their

transcripts off campus and at home. Indeed, the robust nature of the correlations found in this study suggests that reliability was high despite the lack of frequent collaboration between raters and facilitator. However, future investigations may consider the effect of these discussions on inter-rater reliability across discourse type.

This study had raters score CIUs within transcripts without listening to the accompanying audio or video recordings. This factor may likewise significantly influence raters' ability to identify CIUs and may also require additional discussion time to consider information derived not from the transcripts alone but from additional communicative aids evident in the recordings. Factors such as intonation, or environmental sounds may add context that influences the informative nature of the words within connected discourse. While the current investigation's protocol included removal of such influences from the raters' scoring, future investigations might determine the degree to which scoring CIUs with the accompanying audio and video recordings may influence CIU identification and inter-rater agreement.

Given that previous studies suggest that CIU identification may be more variable across discourse type, determination of the degree to which CIU scores vary across repeated measures and discourse is also needed (Cameron et al., 2010). Cameron et al.'s study found repeated measures of CIU scores to vary in repeated measures. Before future applications of this investigation's CIU training module, further evidence of raters' ability to score CIUs consistently across repeated measures is necessary.

Another consideration when reviewing these findings is the nature of the elicitation tasks used. All of the unstructured discourse transcripts were elicited through

questions about participants' experiences recovering from their stroke, or significant life experiences before their strokes. Previous evidence suggests that the nature of the communicative function influences discourse sampling (Wambaugh, Thompson, Doyle, Camarata, 1991). Wambaugh et al.'s study, for instance, found that the use of communicative functions varied depending on the discourse situation (1991). Since it is not currently known whether certain communicative function words are more or less frequently identified as CIUs versus others, the discourse elicitation task may affect the ability to establish CIU reliability. Different genres of discourse tasks have been found to in a different degree of lexical diversity; evidence suggests that procedural discourse elicitation tasks, for instance, result in less lexical diversity (Fergadiotis, Wright, & Capilouto, 2011). Similarly, CIU raters may be able to identify CIUs within some discourse genres more easily than others. Narratives, for example, may be easier to score for CIUs since the topic is known to the rater. This is particularly evident when scoring pronouns. When using a narrative that that rater knows, they will use their own knowledge of the narrative to identify a pronoun without a clear verbal referent as a CIU or not. However, within an expository discourse task, where the rater has less knowledge about the potential content of the content, pronouns without referents would be less likely to be identified as CIUs. Contrastingly, raters may over-identify CIUs within expository discourse samples as they try to interpret participants' meaning. Future investigations should determine the degree to which different types of unstructured discourse elicitation tasks affect communicative function words and possibly CIU score reliability.

This investigation's findings suggest that using the outlined training module to train raters in the CIU identification procedures developed by Nicholas and Brookshire to quantify structured discourse may also be used to train raters to reliably score CIUs within unstructured discourse samples. However, further research is necessary to determine the degree to which CIU identification is reliable within unstructured discourse circumstances after application of the outline training. Future investigatory goals include determining the degree to which CIU reliability varies across different unstructured discourse tasks, determining the nature of different types of errors (false positives versus negatives) made across discourse type, and finally, determining the degree to which CIU scores vary across repeated measures.

Despite the need for further research these findings are promising in that they suggest that using this training, CIU analysis of unstructured discourse may be a future possibility for researchers and clinicians alike. This would allow measurement of PWA's functional discourse providing SLPs with a more detailed understanding of the effects of their treatment approaches on day-to-day conversational skills. Further, application of CIU analysis to unstructured discourse could also provide quantifiable therapeutic goals.

References

- 1) AphasiaBank Consortium Group. (2014). *Transcript Database*. [Elman 01a-14a, BU1a]. Retrieved from <http://talkbank.org/APhasiaBank/>
- 2) Cameron, R. M., Wambaugh, J. L., Mauszycki, S. C., (2010) Individual Variability on Discourse Measures over Repeated Sampling Times in Persons with Aphasia. *Aphasiology*, 24 (6-8), 671-684.
- 3) Carlomagno, S., Giannotti, S., Vorano, L., & Marini, A., Discourse Information Content in Non-Aphasic Adults with Brain Injury: A Pilot Study. (2011) *Brain Injury*. 25(10); 1010-1018.
- 4) Cherney, L., et al., (2010) Updated Evidence-Based Systematic Review: Effects Of Intensity of Treatment and Constraint-induced Language Therapy for Individuals with Stroke-Induced Aphasia. Retrieved from <http://jslhr.asha.org/cgi/content/full/51/5/1282>.
- 5) 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.
- 6) Doyle, P. J., Thompson, C.K., Oleyar, K., Wambaugh, J., Jackson, A., (1994). The Effects of Setting Variables on Conversational Discourse in Normal and Aphasic Adults. *Clinical Aphasiology*. 22; 135-144.
- 7) Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology*, 4(4), 130-134.
- 8) Doyle, P. J., Tsironas, D., Goda, A. J., Kalinyak, M., (1996). The Relationship Between Objective Measures and Listeners' Judgments of the Communicative Informativeness of the Connected Discourse of Adults with Aphasia. *American Journal of Speech-Language Pathology*. 5, (53-55).
- 9) Doyle, P.J., McNeil, M.R., Park, G., Goda, A., Rubenstein, E., Spencer, K., Carroll, B., Lustig, A., Szwed, L., (2000) Linguistic Validation of Four Parallel Forms of a Story Retelling Procedure, *Aphasiology*, 14(5-6), 537-549, DOI: 10.1080/026870300401306.
- 10) Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, 25(10), 1261-1278.

- 11) Goodglass, H., and Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders* (3rd Edition). Philadelphia: Lea & Lebigier.
- 12) Goodglass, H., & Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Lea & Febiger.
- 13) Grimes N. Walt Disney's Cinderella. New York: Random House; 2005.
- 14) Hula, W. D., McNeil, M. R., Doyle, P., Rubinsky, H. J., Fossett, T.R.D., (2003) The Inter-rater Reliability of the Story Retell Procedure. *Aphasiology*, 17(5), 523-528.
- 15) Kertesz, A., (2007). *Western Aphasia Battery: Revised*. San Antonio: Harcourt Assessment, Inc.
- 16) Lapointe, L.L., & Horner, J. (1998). *Reading Comprehension Battery for Aphasia*. Austin, TX: Pro-Ed.
- 17) 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 of Adult Aphasia. *Journal of Speech and Hearing Disorders*, 54(1), 113-124.
- 18) MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25, 1286-1307.
- 19) Marini, A., Andreetta, S., Tin, S., Carlomagno, S., (2011) A Multi-Level Approach to the Analysis of Narrative Language in Aphasia. *Aphasiology*. 25(11), 1372-1392.
- 20) Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, 17(5), 481-497.
- 21) McNeil, M. R., Doyle, P.J., Fossett, T.R., Park, G.H., Goda, A.J., (2010). Reliability and Concurrent Validity of the Information Unit Scoring Metric for the Story Retelling Procedure. *Aphasiology*. 15:10-11, 991-1006, DOI: 10.1080/02687040143000348
- 22) Menn, L. (author). (1995). *Broken Window* [Picture story sequence], Retrieved

May, 2014, from:

<http://talkbank.org/APhasiaBank/protocol/pictures/window.jpg>.

- 23) Nicholas, L. E., & Brookshire, R. H. (1993). A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults. *Journal of Speech & Hearing Research*, 36(2), 338-350.
- 24) Oelschaeger, M. L., Thorne, J. C., Application of the Correct Information Unit Analysis to the Naturally Occurring Conversation of a Person with Aphasia, (1999) *Journal of Speech, Language, and Hearing Research*, 42, 636-648.
- 25) Rubin C, Newton E. Capture the moment: The Pulitzer Prize photographs. W. W. Norton; New York, NY: 2001.
- 26) Saffran, E.M., Berndt, R. S., &Schwartz, M. F., (1989) The Quantitative Analysis of Agrammatic Production: Procedure and Data. *Brain and Language*. 37, 440-479.
- 27) Shewan, C. M., The Shewan Spontaneous Language Analysis (SSLA) System for Aphasic Adults: Description, Reliability, and Validity (1988). *Journal of Communication Disorders*. 21, 103-138.
- 28) Stevens, S.S., On the Theory of Scales and Measurement (1946). *Science* 7. 103(2684), 677-680. DOI:10.1126/science.103.2684.677
- 29) World Health Organization. How to use the ICF: A practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment. October 2013. Geneva: WHO
- 30) Yorkston, K. M., Beukelman, D. R., An Analysis of Connected Speech Samples of Aphasic and Normal Speakers. *The Journal of Speech and Hearing Disorders*. 45(1) 27-36.