JSLHR

Research Article

Manual Versus Automated Narrative Analysis of Agrammatic Production Patterns: The Northwestern Narrative Language Analysis and Computerized Language Analysis

Chien-Ju Hsu[a] and Cynthia K. Thompson[a,b,c]

Purpose: The purpose of this study is to compare the outcomes of the manually coded Northwestern Narrative Language Analysis (NNLA) system, which was developed for characterizing agrammatic production patterns, and the automated Computerized Language Analysis (CLAN) system, which has recently been adopted to analyze speech samples of individuals with aphasia (a) for reliability purposes to ascertain whether they yield similar results and (b) to evaluate CLAN for its ability to automatically identify language variables important for detailing agrammatic production patterns.
Method: The same set of Cinderella narrative samples from 8 participants with a clinical diagnosis of agrammatic aphasia and 10 cognitively healthy control participants were transcribed and coded using NNLA and CLAN. Both coding systems were utilized to quantify and characterize speech production patterns across several microsyntactic levels: utterance, sentence, lexical, morphological, and verb argument structure levels. Agreement between the 2 coding systems was computed for variables coded by both.
Results: Comparison of the 2 systems revealed high agreement for most, but not all, lexical-level and morphological-level variables. However, NNLA elucidated utterance-level, sentence-level, and verb argument structure–level impairments, important for assessment and treatment of agrammatism, which are not automatically coded by CLAN.
Conclusions: CLAN automatically and reliably codes most lexical and morphological variables but does not automatically quantify variables important for detailing production deficits in agrammatic aphasia, although conventions for manually coding some of these variables in Codes for the Human Analysis of Transcripts are possible. Suggestions for combining automated programs and manual coding to capture these variables or revising CLAN to automate coding of these variables are discussed.

Agrammatic aphasia, a language disorder resulting from damage to the neural networks supporting language, is commonly characterized by production of syntactically impoverished speech where production errors manifest on multiple levels (Rochon, Saffran, Berndt, & Schwartz, 2000; Saffran, Berndt, & Schwartz, 1989; Thompson, Shapiro, Li, & Schendel, 1995; Thompson, Shapiro, Tait et al., 1995; Webster, Franklin, & Howard, 2007). On the utterance and sentence levels, studies indicate that speakers with agrammatic aphasia produce short, grammatically ill-formed utterances with reduced syntactic complexity (Miceli, Silveri, Villa, & Caramazza, 1984; Rochon et al., 2000; Saffran et al., 1989; Thompson, Shapiro, Li et al., 1995; Thompson, Shapiro, Tait et al., 1995). In addition, lexical-level errors have been identified in speakers with agrammatic aphasia. For example, omissions and/or misuses of free-standing grammatical morphemes lead to the production of a greater proportion of open-class words than closed-class words (Rochon et al., 2000; Saffran et al., 1989). Difficulties with verb production are also common, yielding lower-than-normal verb-to-noun ratios in spontaneous speech (Bastiaanse & Jonkers, 1998; Bird & Franklin, 1996; Saffran et al., 1989; Thompson, Shapiro, Li et al., 1995). Relatedly, on the verb argument structure level, complex verbs, that is, those that select a greater number of

[a]Neurolinguistics and Aphasia Research Laboratory, Center for the Neurobiology of Language Recovery, The Roxelyn & Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL
[b]Cognitive Neurology and Alzheimer's Disease Center, Northwestern Feinberg School of Medicine, Chicago, IL
[c]The Ken & Ruth Davee Department of Neurology, Northwestern Feinberg School of Medicine, Chicago, IL

Correspondence to Chien-Ju Hsu: chienjuhsu@u.northwestern.edu

Editor-in-Chief: Sean Redmond
Editor: Carl Coelho

Received May 17, 2017
Revision received September 25, 2017
Accepted October 11, 2017
https://doi.org/10.1044/2017_JSLHR-L-17-0185

Disclosure: The authors have declared that no competing interests existed at the time of publication.

arguments or syntactic frames, are more difficult to produce and prone to errors where verb arguments may be omitted or misplaced (Dragoy & Bastiaanse, 2010; Jonkers & Bastiaanse, 1996; Kim & Thompson, 2000, 2004; Thompson, Shapiro, Li et al., 1995; Thompson, 2003; Thompson, Shapiro, Tait et al., 1995; Webster et al., 2007). Finally, on the morphological level, individuals with agrammatic aphasia often show omissions, additions, and/or substitutions of grammatical affixes, particularly past tense forms (Bastiaanse et al., 2011; Miceli et al., 1984; Rochon et al., 2000; Saffran et al., 1989; Yarbay Duman & Bastiaanse, 2009).

Such production difficulties are important to quantify for both clinical and research purposes (MacWhinney, Fromm, Forbes, & Holland, 2011; Prins & Bastiaanse, 2004; Rochon et al., 2000; Saffran et al., 1989; Thompson, Shapiro, Li et al., 1995; Thompson, Shapiro, Tait et al., 1995). Clinically, careful delineation of production deficit patterns can help characterize language breakdown patterns, guide intervention planning and treatment strategies, and document language changes associated with treatment. For research, detailing language production patterns is important for participant selection purposes and in treatment studies to address changes in production associated with treatment.

One common method used to quantify production deficits is spontaneous speech analysis. Although several methods for this are available, the Northwestern Narrative Language Analysis (NNLA) system was developed explicitly for evaluation of aphasic production patterns, with a focus on identification of agrammatic production deficits (Ballard & Thompson, 1999; Faroqi-Shah & Thompson, 2007; Kim & Thompson, 2004; Thompson et al., 2012; Thompson, Shapiro, Li et al., 1995; Thompson, Shapiro, Tait et al., 1995). The system includes procedures for eliciting and transcribing language samples, as well as coding and analyzing a variety of linguistically relevant variables in aphasic language production. The NNLA includes analysis of production at five microstructural levels: utterance, sentence, lexical, morphological, and verb argument structures. In addition, the system renders analyses of production errors, with codes for omission, substitution, addition, and misuse of lexical items, morphological affixes, and verb arguments, as well as different types of paraphasic errors. Several of these codes are not included in other spontaneous language analysis systems. For example, the Quantitative Production Analysis (Rochon et al., 2000; Saffran et al., 1989) does not provide methods for quantification of sentence types or verb argument structure.

The NNLA has been shown to reliably differentiate speakers with agrammatic aphasia from unimpaired speakers, characterize agrammatic production patterns, describe aspects of verbs used in sentence (Thompson, Shapiro, Li et al., 1995; Thompson, Shapiro, Tait et al., 1995), and document changes in narrative production occurring in individuals with aphasia over time, such as improved production from pretreatment to posttreatment and declining production ability in individuals with primary progressive aphasia (Ballard & Thompson, 1999; Jacobs & Thompson, 2000; Thompson, Ballard, Tait, Weintraub, & Mesulam, 1997). The NNLA has also been used to study language production patterns in patients with Alzheimer's disease (Kim & Thompson, 2004) and characterize narrative production in three variants of primary progressive aphasia (Thompson et al., 2012). However, like other traditional quantitative analysis systems, the NNLA is labor intensive and requires a high level of linguistic knowledge and extensive training to manually code each linguistic variable. As pointed out by previous researchers (Edwards, 1995; Prins & Bastiaanse, 2004), because of the difficulty associated with analysis of spontaneous language, it often is not undertaken as part of clinical practice and rarely included as part of research participants' language profiles.

AphasiaBank (MacWhinney et al., 2011), the world's largest database of aphasic language samples, provides a set of automated analysis tools, originally developed by MacWhinney et al. to analyze linguistic and discourse structures of child language, second language learning, and classroom conversation (MacWhinney, 2000). This system includes Codes for the Human Analysis of Transcripts (CHAT; MacWhinney, Fromm, Holland, Forbes, & Wright, 2010), a transcription format, and Computerized Language Analysis (CLAN; MacWhinney et al., 2011), a set of programs for automated analyses of CHAT transcripts. This system also allows for coding and analyses of gestures (Fromm et al., 2011; Kong, Law, Kwan, Lai, & Lam, 2015; Kong, Law, Wat, & Lai, 2015). Unlike manual quantitative analysis methods that require extensive training to learn and time for coding, CLAN performs automatic language analyses on utterance, lexical, and morphological levels when given appropriate commands. However, manual transcription and error coding on these levels are still required in CHAT (see MacWhinney, 2000; MacWhinney et al., 2011 for details). Additionally, automatic coding of sentence and verb argument structure variables is currently not available through CLAN programs, although some may be manually coded.

In recent years, several studies have used CHAT and CLAN to study aphasic language production patterns, reporting aspects of discourse, for example, information structure, discourse structure, and global coherence (Fergadiotis & Wright, 2016; Fridriksson et al., 2012; Richardson & Dalton, 2016), as well as lexical deficit patterns (Boyle, 2015; Fergadiotis & Wright, 2011; Fergadiotis, Wright, & Capilouto, 2011; Fridriksson et al., 2012; Forbes, Fromm, & MacWhinney, 2012; MacWhinney et al., 2011; MacWhinney et al., 2010) or both (Wright & Capilouto, 2012). However, no study to date has used automated systems to evaluate agrammatic speech production, and no study has compared this automated system with other traditional systems like NNLA to explore how well this system quantifies production deficits commonly seen in agrammatic aphasia.

The purpose of this study was to compare the outcomes of the two systems (a) for reliability purposes to ascertain whether they yield similar results and (b) to evaluate CLAN for its ability to automatically identify language variables important for detailing agrammatic production patterns. Importantly, because our purpose was to evaluate the accuracy of automated coding, comparisons between the two systems were made only for fully automated CLAN codes that require no hand coding (i.e., mean length of utterance [MLU], numbers of utterances and words, lexical categories, and grammatical morphemes). We used NNLA coding to detail additional aspects of language that are important for characterizing agrammatic speech but rely on hand coding, particularly utterance-level, sentence-level, and verb argument structure–level variables, and error analyses and point out that, although there are methods for analysis of some of these variables within CLAN, like the NNLA, hand coding is required. We used both systems to analyze the same language samples (Cinderella story) from individuals with agrammatic aphasia, as well as age-matched unimpaired speakers.

Method

Participants

Eight monolingual English-speaking individuals (five men, three women) with a clinical diagnosis of agrammatic aphasia and 10 age-matched, nonbrain-damaged healthy speakers (five men, five women) participated in the study. The mean age of the group with agrammatic aphasia was 58;2 (years;months) and that of the healthy controls was 57;4, with no significant difference between groups, $t(17) = 0.25$, $p = .81$. All participants were recruited from the Greater Chicago area and the Aphasia and Neurolinguistics Research Laboratory, Center for the Neurobiology of Language Recovery, Northwestern University. Participants with agrammatic aphasia suffered a single, left hemisphere stroke (except for P05, whose stroke affected the right hemisphere), were premorbidly right-handed, and at least 19 months post onset of aphasia at the time of testing (ranging from 19 to 216 months). All participants had completed at least 14 years of formal education. In addition, all exhibited adequate visual and hearing acuity to perform the task.

Table 1 presents demographic and language testing data for the participants with aphasia. Language testing included the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2006), subtests from the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes & Thompson, 2012), and for all but one participant with aphasia, the Northwestern Assessment of Verb Inflection (Thompson, 2017). It is important to note that the WAB-R scores were only used to determine aphasia severity but not subtypes because the WAB-R does not classify agrammatic aphasia. Aphasia quotients derived from WAB-R ranged from 65.9 to 93.0 with a mean of 78.88, indicating mild–moderate aphasia severity across participants. Results from various measures, including subtests of the NAVS and the Northwestern Assessment of Verb Inflection, indicated agrammatic production patterns for all participants, including production of grammatically impoverished sentences of both simple sentences, in which verbs and or obligatory arguments are missing, and complex noncanonical sentences, in which thematic role reversals are common, an overall reduction in the proportion of verbs produced compared with nouns, production of fewer closed-class compared with open-class words, and impaired production of grammatical morphemes (i.e., verb inflection). Finally, clinical judgment confirmed that no participant with agrammatic aphasia presented with more than mild apraxia of speech, and none presented with dysarthria of any type.

Procedure

Speech samples used for analyses were derived by asking participants to tell the Cinderella story using the methods described by Saffran et al. (1989). Specifically, participants viewed a wordless storybook of Cinderella and were then asked to tell the story without the book. Interjections by the examiner were kept to a minimum. However, when there were long pauses between utterances, the examiner asked questions, such as "What happened next?" to encourage participants to continue their narration.

Language samples were obtained in a quiet room by trained graduate students or research staff in the Northwestern University Aphasia and Neurolinguistics Research Laboratory with data collection completed within one session. All samples were digitally recorded using the Praat software program (Boersma & Weenink, 2008), which automatically recorded the start and end times of each narration. All aphasic and healthy control participants' samples were then independently transcribed and coded using both the NNLA and AphasiaBank systems (MacWhinney, 2000; Thompson et al., 2012). The NNLA-coded samples were transcribed and coded by highly trained researchers in the Aphasia and Neurolinguistics Research Laboratory. The CLAN-coded patient samples on the basis of the same recordings were transcribed by expert AphasiaBank personnel and were provided through the AphasiaBank database. The samples analyzed are identifiable on the database as Thompson01a–06a, 07b, and 12a. Narrative samples from healthy controls were transcribed and coded by the first author of this article, who was trained to use the CHAT and CLAN systems, following procedures described by MacWhinney (2000).

Language Sample Analyses

NNLA

Transcription. Transcription of language samples involved two steps. First, all identifiable words, including paraphasias and neologisms, were transcribed verbatim with documentation of pause durations. In addition, repetitions, revisions, interjections, false starts, comments, perseverations, and unintelligible words were transcribed but excluded from the analysis. Importantly, speakers tend to use conjunctions like "and" or "but" as fillers at the beginning of an utterance during narration, rather than for conjoining utterances. Sometimes, speakers also tend to add comments like "I think so" at the end of utterance. Instances like this were transcribed but excluded from analysis according

Table 1. Demographic variables, WAB-R scores, and language test scores of the individual participant with agrammatic aphasia (and control group means, where appropriate).

Participant		P01	P02	P03	P04	P05	P06	P07	P08	Group with aphasia, *M*	Control group, *M*
Age at testing		45	47	68	80	54	70	57	45	58;2	57;4
Gender		M	F	M	F	M	M	M	F		
Handedness		R	R	R	R	R	R	R	R		
Education (years;months)		17	16	18	16	21	18	14	16	17	17;1
Time post onset (years;months)		3;0	1;7	14;3	3;0	2;6	9;11	18;0	2;3		
WAB-R	Fluency	9	9	4	4	5	5	4	5	4.5	
	Auditory Comprehension	9.8	10	7.65	9.7	8.95	10	7.4	7.8	8.5	
	Repetition	9	10	7.6	7.6	9.8	9.2	8.9	9.4	3.6	
	Naming	9.7	8.3	8.8	8.5	9.6	9.5	8.6	7.6	6.3	
		93	86.5	73	77.6	82.7	85.4	75.8	77.6	77.2	
NAVS-SPPT	Canonical	100%	47%	47%	53%	100%	100%	67%	53%	61%	
	Noncanonical	67%	0%	40%	47%	73%	53%	60%	20%	37%	
NAVS-SCT	Canonical	87%	87%	100%	100%	100%	100%	100%	73%	93%	
	Noncanonical	68%	53%	100%	100%	87%	93%	40%	67%	76%	
NAVS-ASPT	All arguments	80%	88%	94%	98%	97%	100%	96%	88%	93%	
	All words	98%	92%	97%	100%	100%	100%	86%	94%	96%	
NAVS	VNT	38%	29%	86%	91%	100%	100%	88%	82%	77%	
	VCT	100%	97%	100%	100%	100%	100%	100%	100%	100%	
NAVI	All inflection	95%	65%	56%	N/A	84%	60%	92%	68%	74%	
	Regular past	80%	80%	0%	N/A	80%	20%	60%	20%	49%	
	Irregular past	100%	100%	0%	N/A	100%	0%	100%	40%	63%	

Note. WAB-R = Western Aphasia Battery–Revised; M = male; F = female; R = right; NAVS = Northwestern Assessment of Verbs and Sentences; SPPT = Sentence Production Priming Test; SCT = Sentence Comprehension Test; ASPT = Argument Structure Production Test; VNT = Verb Naming Test; VCT = Verb Comprehension Test; NAVI = Northwestern Assessment of Verb Inflection; N/A = not applicable.

to the NNLA guidelines. The samples were then segmented into utterances using syntactic, prosodic, and semantic criteria. That is, sentence boundaries, prosodic indicators, such as falling intonation and pauses, and semantic content were all considered when marking the end of utterances. Only when utterance markers were unclear, boundaries were placed to create shorter rather than longer utterances (e.g., strings of noun phrases or unintelligible words with no discernable syntactic structure and prosodic indicators). Importantly, when segmenting agrammatic samples, pauses and semantic content were never used as the only definitive utterance markers because pauses were frequent and shifts in content are difficult to detect.

Coding. Utterance, sentence, lexical, bound morpheme, and verb argument structure levels of the NNLA were coded for each narrative sample. At the first coding level, the utterance level, codes were assigned to each utterance to indicate its status as a sentence, a nonsentence, or a sentence fragment. In order to be coded as a sentence, the utterance was required to contain a verb. Utterances coded as sentences then were coded either as being *flawed* or *unflawed* and, if *flawed,* marked as *syntactically* and/or *semantically flawed.* The second coding level, the sentence level, involved further elaboration of utterances coded as sentences. Sentence-level codes denoted sentence complexity (e.g., *syntactically simple* or *complex*), the structure of the sentence (e.g., active, *wh*-question, and cleft sentence), the number of embedded elements in the sentence, and the type of embedding (e.g., complement clause and relative clause). At the third level, the lexical level, all lexical items within each utterance were coded by grammatical category (e.g., open class, closed class, or noun, verb, auxiliary, and preposition). Lexical-level errors, such as omissions, additions, and substitutions, also were each identified by specific error codes. In addition, paraphasias and neologisms were coded by type (i.e., semantic, phonological, mixed, and neologisms). At the fourth coding level, the bound morphological level, all bound morphemes were marked, and specific codes were assigned to each regularly inflected morpheme (e.g., tense, agreement, and plural markers) and irregularly inflected forms. Morphological-level errors, such as omissions, additions, and substitutions, and misuses were also differentiated by codes. Finally, at the fifth coding level, the verb argument structure level, each verb was coded by verb type (e.g., obligatory 1-place verb and optional 2-place verb), and thematic and grammatical roles of noun phrases around the verb (e.g., subject/object and agent/ theme, respectively) were identified. Incorrectly produced verb arguments also were marked by specific codes. The complete NNLA manual is downloadable from the website of the Northwestern University Aphasia and Neurolinguistcs Research Laboratory (http://anr.northwestern.edu/research/clinical-and-research-tools/#NNLA).

CHAT/CLAN

Transcription. CHAT transcription, used by the AphasiaBank, was undertaken for all language samples.

All were transcribed verbatim directly from video (samples of the group with agrammatic aphasia) or audio files (control group's samples). Transcribers segmented all samples into utterances on the basis of the following criteria: syntax, intonation, pauses, and semantics, with primary weight given to the first two criteria (MacWhinney et al., 2011). Repetitions, revisions, word fragments, fillers, and pauses were also marked using CHAT transcription conventions and later excluded in CLAN analyses. Hand-entered error codes were not included (e.g., codes denoting incomplete utterances, paraphasias, and inflectional errors) because this study aimed to compare automated CLAN components with NNLA. It is important to note, however, that AphasiaBank provides guidelines for error coding. Users can also create their own codes for other variables using CLAN's code editor. However, these codes must be entered manually. The guidelines of CHAT transcription, error coding, and the complete CHAT manuals can be obtained from the AphasiaBank website: *talkbank.org/aphasiabank*.

Coding. The transcribed samples were coded automatically using the MOR and POST programs by entering two command lines into the CLAN interface (MacWhinney et al., 2011; Parisse & Le Normand, 2000; Sagae, Davis, Lavie, MacWhinney, & Wintner, 2007). The MOR program automatically assigned all possible part-of-speech tags and morphological tags to each lexical item in the transcribed samples. For example, in (1), the first lexical item, *all* received multiple part-of-speech tags, because *all* can be a quantifier, an indefinite pronoun, or an adverb. Similarly, the morphological marking *-ed,* affixed to the verb *help,* was assigned multiple morphological tags because *-ed* can be a past tense marker or a perfective aspect marker.

(1) …<u>all</u> of the creatures help<u>ed</u> her.

Once the MOR program generated all possible part-of-speech and morphological tags for each of the lexical items produced by the participant, the POST program was used to automatically identify the appropriate tags for each lexical item and morphological affixes within each utterance context. In doing so, *all* in (1) was identified as a pronoun, and the *-ed* ending affixed to *help* was specified as a past tense marker. Detailed descriptions of how to use the MOR and POST programs for automatic coding of transcribed samples can be found in MacWhinney et al. (2011), as well as the complete CLAN manual downloadable on the AphasiaBank website.[1]

Reliability

All NNLA samples were checked for transcription and utterance segmentation accuracy by two independent highly trained transcribers and judged for agreement on a point-to-point basis. The total number of agreements divided by the sum of both the number of agreements and disagreements resulted in an interrater reliability of 96.79% for transcription and 98.37% for segmentation. All samples then were coded by a primary coder, and to examine reliability of coding, 50% of the samples were recoded by a second coder. Transcription coding was judged for consensus, and point-to-point agreement for the utterance-level, sentence-level, lexical-level, bound morphological-level, and verb argument structure–level codes was computed. The total number of agreements divided by the sum of both the number of agreements and disagreements resulted in an interrater reliability of 94.16%.

Reliability data for CHAT transcription were not obtained, but all samples were checked twice by two highly trained speech-language pathologists from AphasiaBank to ensure transcription accuracy before samples were analyzed. The assigned CLAN codes also were not checked for reliability because all samples were coded automatically. However, according to previous reports, CLAN achieves greater than 98% coding accuracy (MacWhinney et al., 2010).

Data Analysis

For the NNLA, the transcribed and coded samples were entered into the Systematic Analysis of Language Transcript (Miller & Chapman, 2000) software program, in which frequency counts of each manually entered code were calculated. The frequency data for each coded variable then were entered into a preprogrammed Microsoft Excel workbook, which quantified and computed general language measures, including MLU, number of utterances, and number of words produced in the selected samples, as well as all coded linguistic variables at the utterance, sentence, lexical, morphological, and verb argument structure levels.

To quantify variables in CLAN, a series of CLAN automatic analysis programs, including EVAL and MORtable were used. The MORtable analysis uses the morphological tagging created by the MOR program to compute frequency counts of all lexical items (open-class and closed-class words) and morphological affixes. The EVAL program also uses the same morphological information to calculate additional information, including the abovementioned general language measures, as well as the percentage of words produced by word class. These CLAN analysis programs were operated by entering a command line in the CLAN command box. Detailed descriptions of each analysis program and relevant commands can be found in MacWhinney (2000) and Forbes et al. (2012).

Dependent Measures

General language measures, including MLU, number of utterances and number of words produced were computed by both systems, as were lexical- and morphological-level variables. At the lexical level, frequency counts of lexical items were calculated by both systems, including open-class words (nouns, verbs, adjectives, and adverbs) and

[1]Note that in the most current version of AphasiaBank, MOR and POST are no longer two-step processes. Please check the latest manual downloadable from AphasiaBank website for details.

closed-class words (determiners, pronouns, auxiliaries, conjunctions, modals, prepositions, negation markers, infinitival markers, quantifiers, *wh*-words, and particles). Proportional data of each open-class and closed-class lexical item over total words produced in the selected samples were computed by both systems. While NNLA provides open-to-closed-class word (O:C) ratios, as well as noun-to-verb (N:V) ratios in the output, these two measures were manually calculated from CLAN's MORtable output. At the morphological level, bound morphemes, calculated by both NNLA and CLAN by type, were quantified. These included comparative, superlative, and possessive markers; regular and irregular plural forms; and verb tense markers, including third person present, regular, and irregular past, regular perfect aspect, irregular perfect participles, and progressive aspect markers. The NNLA also computed values for the proportion of correct regular and irregular inflections used, but CLAN does not (although manual error coding in CHAT is possible and can be analyzed by CLAN).

Where the two coding systems differ primarily pertain to utterance-level, sentence-level, and verb argument structure–level codes. At the utterance level, the NNLA computed the proportion of sentences produced (i.e., utterances containing at least a verb), as well as the proportion of sentences with correct and/or flawed syntax and/or semantics. At the sentence level, NNLA computed the number of embedded clauses produced per sentence and a sentence complexity ratio. *Complex sentences* were defined as sentences containing an embedded clause and/or sentences involving syntactic movement. CLAN does not include a sentence code, does not provide automatic analysis of the number of embedding per sentence, and does not compute a sentence complexity ratio.

Finally, at the verb argument structure level, NNLA quantified the proportion of verbs produced by type (e.g., obligatory/optional 2-place verbs) and the proportion of each produced with correct argument structure. This level of analysis is also not automated in CLAN programs, and in addition, codes for verb types and verb argument structure variables are not available in CHAT.

Statistical Analysis

For variables coded by both the NNLA and the CLAN, we examined between-groups differences for the two systems, using the Mann–Whitney U test. We adopted a standard significance level of 0.05 in all the statistical tests. For variables coded using only NNLA, between-groups comparisons were performed to examine differences between agrammatic and control participant groups using the nonparametric Mann–Whitney U test given the small number of participants in each group.

Results

Results are shown in Tables 2–6 by participant group (agrammatic, control) and coding system (NNLA, CLAN).

General Measures

With regard to general narrative variables, both systems showed that the group with agrammatic aphasia produced significantly fewer words (NNLA: agrammatic M = 235.13, SD = 168.17, control M = 469.30, SD = 197.60, U = 15, p = .03; CLAN: agrammatic M = 298.75, SD = 191.04, control M = 515.8, SD = 210.25, U = 15, p = .00) and lower MLU (NNLA: agrammatic M = 6.66, SD = 0.79, control M = 12.15, SD = 2.53, U = 0.00, p = .00); CLAN: agrammatic M = 6.63, SD = 1.00, control M = 13.33, SD = 2.49, U = 0.00, p = .00) compared with the control participants. Both systems also found no significant differences in the number of utterances produced (NNLA: U = 31.5, p = .45; CLAN: U = 37.5, p = .79). In addition, as Table 2 shows, there was no difference between the two coding systems in quantification of these three variables.

Lexical-Level Variables

Results of the lexical-level analyses are shown in Table 3. The NNLA coding system revealed marginally significant between-groups differences reflecting lexical impairments in the group with agrammatic aphasia in the proportion of open-class words produced, U = 19; p = .06 (group with agrammatic aphasia: M = 53.96%, SD = 8.55 vs. control group: M = 48.43%, SD = 3.10). In line with these results, a marginally significant open-to-closed word (O:C) ratio, U = 19, p = .06 (group with agrammatic aphasia: M = 1.25, SD = 0.48 vs. control group: M = 0.95, SD = 0.12) was noted between the two groups. Interestingly, the proportion of verbs produced by the group with agrammatic aphasia (M = 19.35%, SD = 2.35) did not differ significantly from the control participants (M = 17.84%, SD = 1.03), U = 23, p = .13; however, the group with agrammatic aphasia produced a significantly higher proportion of nouns (M = 25.18%, SD = 7.61) compared with unimpaired control speakers (M = 19.09%, SD = 1.87), U = 13; p = .02. This resulted in a significant group difference for noun-to-verb (N:V) ratio, U = 18; p = .05, with the group with agrammatic aphasia producing a significantly greater proportion of nouns over verbs (M = 1.29, SD = 0.30) than the control group (M = 1.07, SD = 0.13).

For the same lexical-level variables, the CLAN analysis system found different production patterns where no reliable between-groups difference was found for the proportions of open-class words, U = 38, p = .86 (group with agrammatic aphasia: M = 47.08%, SD = 7.19 vs. control group: M = 46.43%, SD = 10.78), and closed-class words, U = 38, p = .86 (group with agrammatic aphasia: M = 52.92%, SD = 7.19 vs. control group: M = 53.97%, SD = 10.78). As such, no between-groups difference was detected for the O:C ratios, U = 38; p = .86 (group with agrammatic aphasia: M = 0.92, SD = 0.29 vs. control group: M = 0.94, SD = 0.43). In addition, whereas CLAN showed no significant between-groups difference for the proportion of verbs produced (as did the NNLA), U = 38, p = .86 (group with agrammatic aphasia: M = 13.78%, SD = 1.9 vs. control

Table 2. General language measures derived from both systems and statistical results.

General language measures	NNLA		CLAN	
	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)
MLU	6.66 (0.79)	12.15 (2.53)	6.63 (1.00)	13.33 (2.49)
Number of utterances	41.63 (25.96)	45.30 (21.67)	50.13 (28.44)	45.70 (21.39)
Number of words	235.13 (168.17)	469.30 (197.60)	298.75 (191.04)	515.80 (210.25)
Mann–Whitney *U* between-groups tests				
	Group with agrammatic aphasia vs. control groups		NNLA vs. CLAN	
General language measures	NNLA	CLAN	Agrammatic	Control
MLU	$U = 0.00; p = .00^*$	$U = 0.00; p = .00^*$	$U = 31; p = .92$	$U = 33; p = .20$
Number of utterances	$U = 31.5; p = .45$	$U = 37; p = .79$	$U = 21; p = .25$	$U = 48; p = .91$
Number of words	$U = 15; p = .03^*$	$U = 15; p = .03^*$	$U = 21; p = .25$	$U = 38.5; p = .38$

Note. NNLA = Northwestern Narrative Language Analysis; CLAN = Computerized Language Analysis; MLU = mean length of utterance.
$^*p < .05$.

group: M = 14.93%, SD = 3.54), unlike NNLA, CLAN found no reliable between-groups difference for the proportion of nouns produced, U = 25, p = .18 (group with agrammatic aphasia: M = 19.09%, SD = 1.87 vs. control group: M = 20.35%, SD = 4.02). As a result, the N:V ratios did not differ between the group with agrammatic aphasia (M = 1.1, SD = 0.36) and the control groups on the basis of CLAN (M = 1.39, SD = 0.23), U = 20; p = .08.

For the majority of other lexical classes coded, similar numbers were derived from NNLA and CLAN, including adjectives, adverbs, determiners, pronouns, auxiliaries, prepositions, negation markers, infinitival markers, quantifiers, and *wh*-words (see Table 3). Significant differences were found only for conjunctions produced by each participant group (group with agrammatic aphasia: U = 9, p = .02; control group: U = 23, p = .04), modals (group with agrammatic aphasia: U = 6, p = .00; control group: U = 15, p = .01), and particles (group with agrammatic aphasia: U = 4, p = .00; control group: U = 0.00, p = .00). CLAN consistently picked out more conjunctions, whereas NNLA found more modals and particles.

Morphological-Level Variables

Both NNLA and CLAN provided frequency counts of comparative suffixes, superlative suffixes, possessive markers, regular and irregular plural markers, third person present tense, regular and irregular past tense markers, regular and irregular perfect participles, and progressive aspect markers (see Table 4). No reliable difference between the two systems in quantification of these variables was found in either participant group. In terms of morphological error analysis, the NNLA system revealed significantly lower accuracy for speakers with agrammatic aphasia compared with healthy control speakers, for both regular, U = 11, p = .00 (group with agrammatic aphasia: M = 86.74%, SD = 10.94 vs. control group: M = 99.23%, SD = 2.43), and irregular grammatical morphemes, U = 5.5; p = .00 (group with agrammatic aphasia: M = 73.78%, SD = 33.63 vs. control group: M = 86.74%, SD = 10.94), with production of irregularly inflected morphemes more impaired than that of regularly inflected morphemes. Because CLAN does not code for errors automatically, these computations were not undertaken for CLAN-coded samples.

As noted previously, automatic coding of utterance-level, sentence-level, and verb argument structure–level variables is not available in CLAN, although CLAN provides hand-coding guidelines for utterance-level variables like syntactic versus semantic flaws. Other sentence-level variables, like number of sentences, number of embedded clauses per sentence, and sentence complexity ratios, can possibly be analyzed within CLAN without additional hand coding, but guidelines for this are not currently available. CLAN does not address verb argument structure–level coding. However, because of the importance of these variables in determining patterns of agrammatic production, we report the results of NNLA for these levels of analysis.

Utterance-Level Variables

For utterance-level codes shown in Table 5, results of the NNLA coding showed that, although both the group with agrammatic aphasia and the control groups produced a similar number of utterances, the group with agrammatic aphasia produced a significantly reduced proportion of sentences (i.e., utterances with verbs; M = 80.02%; SD = 5.86) compared with control speakers (M = 98.22%; SD = 2.20), U = 0.00 p = .00. Reliable differences between controls and participants with agrammatic aphasia were also noted for the following variables: the proportion of grammatically and semantically correct sentences, U = 0.00, p = .00, the proportion of sentences with syntactic flaws, U = .00, p = .00, and the proportion of sentences with semantic flaws, U = 7, p = .00. The group with agrammatic aphasia produced a mean of 44.61% grammatical sentences (SD = 17.71 vs.

Table 3. Lexical-level values derived from both coding systems and statistical results.

Lexical variables	NNLA		CLAN	
	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)
Total number of open-class words produced	124.63 (85.88)	230.20 (106.18)	139.75 (86.86)	234.20 (93.73)
Proportion of open-class words over all words	53.96% (8.55)	48.43% (3.10)	47.08% (7.19)	46.43% (10.78)
Total number of closed-class words produced	110.50 (84.91)	239.10 (92.37)	132.38 (83.09)	239.70 (78.08)
Proportion of closed-class words over all words	46.04% (8.55)	51.57% (3.10)	52.92% (7.19)	53.57% (10.78)
Open-to-closed word ratio	1.25 (0.48)	0.95 (0.12)	0.92 (0.29)	0.94 (0.43)
Total nouns produced	55.75 (35.54)	88.20 (33.99)	67.13 (39.24)	102.60 (42.95)
Proportion of nouns over all words	25.18% (7.61)	23.24% (6.79)	19.09% (1.87)	20.35% (4.20)
Total verbs produced	43.88 (29.36)	84.00 (36.41)	41.75 (27.77)	74.40 (27.46)
Proportion of verbs over all words	19.35% (2.35)	17.84% (1.03)	13.78% (1.9)	14.93% (3.54)
Noun-to-verb ratios	1.29 (0.30)	1.07 (0.13)	1.10 (0.36)	1.39 (0.23)
Total adjectives	9.75 (9.25)	26.80 (21.38)	8.75 (9.11)	20.80 (17.07)
Total adverbs	14.38 (14.56)	31.20 (18.61)	22.13 (16.00)	36.40 (16.73)
Total determiners	31.38 (22.87)	54.20 (21.96)	40.25 (23.46)	55.10 (23.29)
Total pronouns	24.75 (26.36)	61.20 (25.04)	26.00 (24.32)	66.70 (29.28)
Total auxiliaries	7.88 (7.08)	10.70 (6.62)	5.88 (5.44)	12.80 (8.72)
Total conjunctions	8.25 (7.85)	21.20 (10.58)	27.38 (22.44)	38.50 (18.10)
Total modals	2.63 (3.16)	9.80 (5.07)	0.13 (0.35)	4.00 (3.13)
Total prepositions	16.13 (13.41)	35.80 (13.94)	21.25 (14.83)	44.70 (15.66)
Total negation markers	1.75 (1.49)	5.80 (3.12)	1.75 (1.49)	5.20 (2.66)
Total infinitival markers	2.75 (3.49)	6.60 (3.03)	3.25 (3.24)	10.10 (4.38)
Total quantifiers	5.13 (2.53)	14.30 (6.45)	5.38 (2.33)	9.40 (4.81)
Total *wh*-words	.75 (0.89)	5.30 (4.37)	1.13 (1.36)	2.30 (2.41)
Total particles	1.88 (1.25)	8.20 (3.82)	0.00 (0.00)	0.00 (0.00)

Mann–Whitney *U* between-groups tests

Lexical variables	Group with agrammatic aphasia vs. control group		NNLA vs. CLAN	
	NNLA	CLAN	Agrammatic	Control
Total number of open-class words produced	$U = 15$; $p = .03$*	$U = 16.5$; $p = .04$*	$U = 27$; $p = .60$	$U = 46$; $p = .76$
Proportion of open-class words over all words	$U = 19$; $p = .06$+	$U = 38$; $p = .86$	$U = 17$; $p = .12$	$U = 36$; $p = .29$
Total number of closed-class words produced	$U = 12$; $p = .01$*	$U = 14$; $p = .02$*	$U = 25$; $p = .46$	$U = 46$; $p = .76$
Proportion of closed-class words over all words	$U = 19$; $p = .06$+	$U = 38$; $p = .86$	$U = 17$; $p = .12$	$U = 36$; $p = .29$
Open-to-closed word ratio	$U = 19$; $p = .06$+	$U = 38$; $p = .86$	$U = 17$; $p = .12$	$U = 36$; $p = .29$
Total nouns produced	$U = 20$; $p = .08$	$U = 24$; $p = .15$	$U = 25$; $p = .46$	$U = 35$; $p = .26$
Proportion of nouns over all words	$U = 13$; $p = .02$*	$U = 25$; $p = .18$	$U = 25$; $p = .46$	$U = 34$; $p = .23$
Total verbs produced	$U = 15$; $p = .03$*	$U = 15$; $p = .03$*	$U = 29$; $p = .75$	$U = 41$; $p = .50$
Proportion of verbs over all words	$U = 23$; $p = .13$	$U = 38$; $p = .86$	$U = 0.00$; $p = .00$*	$U = 26$; $p = .07$
Noun-to-verb ratios	$U = 18$; $p = .05$*	$U = 20$; $p = .08$	$U = 14.5$; $p = .07$	$U = 12$; $p = .00$*
Total adjectives	$U = 15$; $p = .03$*	$U = 15$; $p = .03$*	$U = 27$; $p = .60$	$U = 34$; $p = .23$
Total adverbs	$U = 13$; $p = .02$*	$U = 18$; $p = .05$*	$U = 20$; $p = .21$	$U = -0.91$; $p = .36$
Total determiners	$U = 18$; $p = .05$*	$U = 26.5$; $p = .23$	$U = 19$; $p = .17$	$U = 48$; $p = .88$
Total pronouns	$U = 12$; $p = .01$*	$U = 12$; $p = .01$*	$U = 27.5$; $p = .64$	$U = 38.5$; $p = .38$
Total auxiliaries	$U = 24.5$; $p = .17$	$U = 19.5$; $p = .07$	$U = 26$; $p = .53$	$U = 38.5$; $p = .38$
Total conjunctions	$U = 9.5$; $p = .01$*	$U = 24.5$; $p = .17$	$U = 9$; $p = .02$*	$U = 23$; $p = .04$*
Total modals	$U = 8$; $p = .00$*	$U = 9.5$; $p = .00$*	$U = 6$; $p = .00$*	$U = 15$; $p = .01$*
Total prepositions	$U = 12$; $p = .01$*	$U = 11.5$; $p = .01$*	$U = 24.5$; $p = .43$	$U = 32$; $p = .17$
Total negation markers	$U = 7$; $p = .00$*	$U = 11$; $p = .01$*	$U = 31.5$; $p = .96$	$U = 44.52$; $p = .67$
Total infinitival markers	$U = 15$; $p = .03$*	$U = 8$; $p = .00$*	$U = 26$; $p = .52$	$U = 28$; $p = .10$
Total quantifiers	$U = 4.5$; $p = .00$*	$U = 16.5$; $p = .03$*	$U = 29$; $p = .75$	$U = 28.5$; $p = .10$
Total *wh*-words	$U = 5$; $p = .00$*	$U = 27$; $p = .23$	$U = 28$; $p = .66$	$U = 25.5$; $p = .06$
Total particles	$U = 4$; $p = .00$*	—	$U = 4$; $p = .00$*	$U = 0.00$; $p = .00$*

Note. Em dash indicates data not [obtained]/[reported]/[available]. NNLA = Northwestern Narrative Language Analysis; CLAN = Computerized Language Analysis.

*$p < .05$. +$p < .06$.

Table 4. Morphological-level values derived from both coding systems and statistical results.

Morphological variables	NNLA		CLAN	
	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)
Total comparative suffixes	0.13 (.35)	0.30 (0.67)	0.13 (.35)	0.50 (0.85)
Total superlative suffixes	0.00 (0.00)	0.10 (0.32)	0.00 (0.00)	0.10 (0.32)
Total possessive markers	0.75 (1.49)	0.80 (1.32)	1.13 (1.36)	0.90 (1.29)
Total regular plural markers	9.88 (10.23)	15.20 (5.79)	13.13 (10.16)	13.20 (7.36)
Total irregular plural forms	1.13 (1.13)	2.60 (1.71)	1.75 (1.91)	3.00 (1.63)
Total third person present tense markers	3.75 (3.73)	12.50 (15.95)	4.88 (7.20)	11.10 (13.89)
Total regular past tense markers	5.13 (4.76)	8.80 (5.57)	5.00 (5.10)	9.50 (5.84)
Total irregular past tense markers	14.13 (15.57)	24.50 (13.80)	16.75 (18.65)	26.90 (13.76)
Total regular perfect aspect markers	6.25 (5.06)	3.60 (3.10)	7.13 (5.28)	3.70 (2.71)
Total irregular perfect participles	1.00 (1.41)	1.50 (0.97)	2.63 (2.33)	1.90 (1.45)
Total progressive aspect markers	0.25 (0.46)	6.60 (3.50)	1.00 (1.07)	6.80 (4.57)
Proportion of correct regular inflection	86.74 (10.94)	99.23 (2.43)	—	—
Proportion of correct irregular inflection	73.78 (33.63)	86.74 (10.94)	—	—

Mann–Whitney *U* between-groups tests

Morphological variables	Group with agrammatic aphasia vs. control groups		NNLA vs. CLAN	
	NNLA	CLAN	Agrammatic	Control
Total comparative suffixes	*U* = 36.5; *p* = .63	*U* = 32; *p* = .33	*U* = 32; *p* = 1.00	*U* = 44.5; *p* = .58
Total superlative suffixes	*U* = 36; *p* = .37	*U* = 36; *p* = .37	*U* = 32; *p* = 1.00	*U* = 50; *p* = 1.00
Total possessive markers	*U* = 36; *p* = .67	*U* = 35; *p* = .64	*U* = 23; *p* = .30	*U* = 46; *p* = .74
Total regular plural markers	*U* = 19.5; *p* = .07	*U* = 38; *p* = .86	*U* = 24.5; *p* = .43	*U* = 33; *p* = .20
Total irregular plural forms	*U* = 20; *p* = .07	*U* = 24; *p* = .15	*U* = 27.5; *p* = .62	*U* = 44; *p* = .64
Total third person present tense markers	*U* = 35; *p* = .65	*U* = 33.5; *p* = .56	*U* = 30; *p* = .83	*U* = 47; *p* = .82
Total regular past tense markers	*U* = 26.5; *p* = .23	*U* = 20; *p* = .07	*U* = 31; *p* = .92	*U* = 45.5; *p* = .73
Total irregular past tense markers	*U* = 24; *p* = .16	*U* = 26.5; *p* = .23	*U* = 30.5; *p* = .87	*U* = 40.5; *p* = .47
Total regular perfect aspect markers	*U* = 14; *p* = .02*	*U* = 30; *p* = .39	*U* = 16; *p* = .09	*U* = 48; *p* = .88
Total irregular perfect participles	*U* = 12; *p* = .01*	*U* = 25.5; *p* = .18	*U* = 18; *p* = .10	*U* = 39; *p* = .38
Total progressive aspect markers	*U* = 33.5; *p* = .56	*U* = 37.5; *p* = .83	*U* = 28; *p* = .67	*U* = 47.5; *p* = .85
Proportion of correct regular inflection	*U* = 11; *p* = .00*	—	—	—
Proportion of correct irregular inflection	*U* = 5.5; *p* = .00*	—	—	—

Note. Em dashes indicate data not [obtained]/[reported]/[available]. NNLA = Northwestern Narrative Language Analysis; CLAN = Computerized Language Analysis.

*$p < .05$.

Table 5. Utterance- and sentence-level values derived from NNLA and statistical results.

Variables	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)	Stats
Utterance level			
Proportion of sentences produced	80.02% (5.86)	98.22% (2.20)	*U* = 0.00; *p* = .00*
Proportion of sentences with correct syntax and semantics	44.61% (17.72)	97.21% (3.25)	*U* = 0.00; *p* = .00*
Proportion of sentences with flawed syntax	49.73% (17.76)	2.18% (3.29)	*U* = 0.00; *p* = .00*
Proportion of sentences with flawed semantics	11.90% (8.47)	.38% (0.89)	*U* = 7; *p* = .00*
Sentence level			
Sentence complexity ratio	0.24 (0.11)	0.82 (0.26)	*U* = 0.00; *p* = .00*
Number of embedded clauses per sentence	0.19 (0.09)	0.71 (0.20)	*U* = 0.00; *p* = .00*

Note. NNLA = Northwestern Narrative Language Analysis.

*$p < .05$.

Table 6. Verb argument structure–level: proportion data derived from NNLA and statistical results.

Argument structure variables	Group with agrammatic aphasia, *M* (*SD*)	Control group, *M* (*SD*)	Stats
1-place verbs over all verbs	29.39% (14.01)	31.16% (5.39)	$U = 34$; $p = .60$
2-place verbs over all verbs	54.78% (16.70)	62.12% (7.12)	$U = 31$; $p = .42$
3-place verbs over all verbs	4.57% (6.12)	5.97% (4.06)	$U = 28$; $p = .28$
1-place verbs with correct argument structure	89.31 (11.27)	100.00 (0.00)	$U = 15$; $p = .00$*
2-place verbs with correct argument structure	89.78 (5.19)	100.00 (0.00)	$U = 5$; $p = .00$*
3-place verbs with correct argument structure	88.28 (7.42)	99.00 (3.16)	$U = 5$; $p = .00$*

Note. NNLA = Northwestern Narrative Language Analysis.
*$p < .05$.

control: $M = 97.21\%$; $SD = 3.25$), with the rest of their narrative production flawed by syntactically ill-formed sentence structures (agrammatic: $M = 49.73\%$; $SD = 17.76$ vs. control: $M = 2.18\%$; $SD = 3.29$), more so than incorrect word meanings (agrammatic: $M = 11.90\%$; $SD = 8.47$ vs. control: $M = 0.38\%$; $SD = 0.89$).

Sentence-Level Variables

Table 5 also presents sentence-level performance patterns derived from NNLA, which show that speakers with agrammatic aphasia produced decreased proportions of complex sentences compared with controls, with lower sentence complexity ratios, $U = 0.00$, $p = .00$ (group with agrammatic aphasia: $M = 0.24$, $SD = 0.11$ vs. control group: $M = 0.82$, $SD = 0.26$). The NNLA also showed that the participants with agrammatic aphasia compared with controls produced fewer embedded clauses per sentence, $U = 0.00$, $p = .00$ (group with agrammatic aphasia: $M = 0.19$, $SD = 0.09$ vs. control group: $M = 0.71$, $SD = 0.20$).

Verb Argument Structure–Level Variables

Finally, the results of verb argument structure coding, provided by the NNLA, are shown in Table 6. NNLA results showed that both groups produced more 2-place and 1-place verbs than 3-place verbs, and the proportion of each verb type produced did not differ significantly between the group with agrammatic aphasia and the control group. In terms of accuracy, the group with agrammatic aphasia produced a significantly lower proportion of verbs with correct argument structure across verb types. Whereas the control groups produced verbs with correct verb arguments almost 100% of the time, the group with agrammatic aphasia produced between 88% and 89% of verbs with correct arguments. These between-groups differences in accuracy across verb types were statistically reliable (1-place verbs: $U = 15$, $p = .00$; 2-place verbs: $U = 5$, $p = .00$; 3-place verbs: $U = 5$, $p = .00$). This pattern of results is in line with agrammatic verb argument structure impairments, as well as is in keeping with pretesting language patterns derived from the NAVS subtest, Argument Structure Production Test (see Table 1, Cho-Reyes & Thompson, 2012).

Discussion and Conclusions

This study compared the results from a manual narrative language coding system, NNLA, with those from the automated CLAN analysis programs. To our knowledge, no studies have previously used an automated coding system to analyze agrammatic speech production nor compared the results of such system to a manual coding system. In order to do this, we compared variables automatically coded by CLAN programs to variables manually coded using NNLA (a) to ascertain whether they yield similar results for reliability purposes and (b) to evaluate CLAN for its ability to automatically identify language variables important for detailing agrammatic production patterns. Importantly, because our purpose was to evaluate the accuracy of automated coding, comparisons between the two systems were made only for fully automated CLAN codes that require no hand-coding (i.e., lexical- and morphological-level variables). We used NNLA coding to detail additional aspects of language that are important for characterizing agrammatic speech but rely on hand-coding, particularly sentence and verb argument structure–level variables and error analyses, and point out that, although there are methods for analysis of some of these variables within CLAN, like the NNLA, hand-coding is required.

Results showed that both systems code general language measures similarly, including MLU, as well as the number of utterances and words produced in language samples, with results showing no significant differences between these measures when the two systems were compared. Both NNLA and CLAN also revealed similar patterns of lexical and morphological production, with codes for the same word classes, including codes for both open-class and closed-class words, and grammatical morphemes. When directly comparing overlapping lexical measures from NNLA and CLAN, we found high agreement between the two systems. Importantly, however, NNLA, but not CLAN, found deficits in N:V and O:C ratios. This likely resulted from differences in transcription conventions used across systems. Both systems exclude repetitions, revisions, interjections, false starts, comments, and perseverations because these reflect word retrieval difficulty rather than propositional utterance production. However, NNLA has more strict exclusionary criteria. For example, NNLA excludes comments such as *I think, I guess,* which are frequently used by speakers with agrammatic aphasia, but CHAT does not (although this can be altered manually). Hence, these are included for CLAN analysis (notably, *I* is a closed-class pronoun, and *think* and *guess* are verbs). In addition, NNLA also excludes all sentence-initial conjunctions that do not serve the function of connecting two clauses (e.g., *and, so* at the beginning of the sentence, which are both closed-class items). For this reason, the total numbers of conjunctions counted in NNLA-coded and CLAN-coded samples differed significantly. Consider, for example, the following utterance produced by one of the study participants:

"and uh slipper I mean glass slipper. I think so."

CHAT transcription results in inclusion of all words except "*uh slipper,*" whereas NNLA transcription retains only "*glass slipper*" for coding.

NNLA: ~~and uh slipper I mean~~ glass slipper. ~~I think so~~.
CHAT: and ~~uh slipper~~ I mean glass slipper I think so.

Other cross-system differences in the counts of modals and particles reflect differences in data treatment. Phrases like "be going to" and "had to" are coded as single modals in NNLA, reflecting their grammatical function in the sentence, whereas in CLAN, such phrases are coded by word, as separate lexical items [i.e., *be going to* is coded as an auxiliary (closed class), a verb (open class), and an infinitival marker (closed class)]. Similarly, NNLA codes adverbs or prepositions following a single verb (e.g., *up* in *pick up*) as verb particles (closed class), whereas, the CLAN codes these items as adverbs (open class) and/or prepositions (closed class).

At the morphological level, both systems code overlapping, bound grammatical morphemes, providing frequency counts by morpheme. Results showed no significant differences between NNLA and CLAN counts for these variables. NNLA revealed that irregular inflected morphemes are more impaired than regular morphemes in agrammatic aphasia. This may result from the fact that the Cinderella story elicits more irregular verbs than regular verbs (MacWhinney et al., 2010). Examining differences between regular and irregular production, particularly verb inflection, is important clinically, as well as for understanding the nature of production impairments in agrammatism. CLAN does not automatically detect morphemic errors, although it offers guidelines (similar to NNLA) for hand-coding errors at this level.

Several variables, which are particularly important for detailing agrammatic production patterns, are not automatically coded in existing CLAN programs. These are primarily associated with utterance, sentence, and verb argument coding levels. At the utterance level, the proportion of sentences produced across all utterances is not quantified by CLAN. NNLA showed that even though the participants with agrammatic aphasia produced as many utterances as the unimpaired control participants, only about 80% qualified as sentences (vs. 98.22% in the control group). Since the NNLA definition of a sentence is any utterance that contains at least a lexical verb and verbs are required for grammatical sentences, this measure is particularly sensitive to sentence-level impairments. Although it is possible to use CLAN commands like KWAL or FREQ to derive frequency counts of utterances with verbs, CLAN guidelines do not provide an operational definition of a sentence; hence, it is left to users of the program to determine parameters that constitute a sentence versus, for example, a sentence fragment, which could potentially introduce inconsistencies in results.

At the sentence level, NNLA, but not CLAN programs, codes sentence complexity, yielding a sentence complexity ratio, and revealed lower sentence complexity ratios in speakers with agrammatic aphasia, compared with unimpaired control speakers. NNLA also codes the number or type of embedded clauses produced and the nature of sentence errors, that is, whether they are syntactically or semantically flawed. Notably, NNLA analysis of sentence types indicated that over 55% of the sentences of speakers with agrammatic aphasia were syntactically and/or semantically ill-formed, with a greater proportion of syntactic errors compared with semantic errors. NNLA also revealed impoverished production of embedded clauses in the participants with agrammatic aphasia. Indeed, these are important variables to detail in agrammatic speech because many agrammatic sentences are syntactically, rather than semantically, flawed (frequently observed in people with aphasia of other types, e.g., anomic aphasia), and because few embedded sentences are produced. Once again, however, even though CLAN does not automatically code these variables, CHAT provides guidelines for manual sentence-level error coding.

Finally, while NNLA includes a verb argument structure level of coding, focused on detailing verbs by type, CLAN does not support automated verb argument structure production analysis, nor does it provide guidelines for hand coding such variables in CHAT. Verb argument structure production has been extensively studied in agrammatic aphasia across languages, with results indicating that this is an important diagnostic indicator of agrammatic production (e.g., Dragoy & Bastiaanse, 2010; Jonkers & Bastiaanse, 1996; Kim & Thompson, 2000, 2004; Thompson, 2003; Thompson, Shapiro, Li et al., 1995; Thompson, Shapiro, Tait et al., 1995; Webster et al., 2007). Although it can be difficult to assign thematic roles in narrative discourse, an automated annotation corpus, the Proposition Bank (Palmer, Gildea, & Kingsbury, 2005) has been developed under the PennTree Bank system to serve this purpose (Marcus, Santorini, & Marcinkiewicz, 1993). The Proposition Bank automatically codes the argument structure of common verbs in sentence contexts, although it has not been perfected to code for all verb types (i.e., some verbs still require manual annotation/disambiguation, e.g., alternating unaccusatives), and the accuracy of automated annotation is reportedly 80.9% (Palmer et al., 2005). Until automatic verb argument structure coding is reliably developed, CHAT would benefit from incorporating guidelines for hand coding verb argument structures (as well as errors).

In summary, findings from this study indicate that CLAN's automated coding at general, lexical, and morphological levels yields results largely consistent with those derived from NNLA's manual coding, except for a few measures noted above. This suggests that CLAN programs can reliably facilitate users with or without extensive linguistic knowledge to accomplish most lexical and morphological analyses efficiently. However, the current version of CLAN does not offer programs automatically detailing other measures important for identifying and documenting agrammatic production patterns, namely, utterance-level, sentence-level, and verb argument structure–level variables, as well as error analysis. Deficits on these levels are particularly difficult to track with standardized tests alone. When treatment goals involve utterance, sentence, and verb production, narrative tasks are important to ascertain individuals' ability to use language. Presently, however, hand coding is required to quantify these important variables. While CHAT provides guidelines for error coding on the sentence, lexical, and morphological levels, there are currently no guidelines to facilitate manual coding on other variables associated with the sentence and verb argument structure levels.

In conclusion, a fully automated system for analysis of agrammatic speech is needed for clinicians and researchers alike. The AphasiaBank analysis tools, including the CHAT format and the automated CLAN programs, have potential to fill this void. As such, it could be used to quantify agrammatic production deficit patterns in people with aphasia and evaluate changes in production over time. However, in its present form, manual coding cannot be completely replaced, especially when evaluating utterance, sentence, and verb production patterns and changes. Therefore, it is recommended that clinicians and researchers combine the benefit of automated coding on lexical and morphological levels with manual coding on other important variables to fully capture different aspects of agrammatic production ability.

Acknowledgments

The authors would like to thank Jennifer E. Mack and Jiyeon Lee for reliability coding and the researchers from the AphasiaBank project at Carnegie Mellon University for their training and for preparing the agrammatic samples.

References

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

Bastiaanse, R., Bamyaci, E., Hsu, C. J., Lee, J., Duman, T. Y., & Thompson, C. K. (2011). Time reference in agrammatic aphasia: A cross-linguistic study. *Journal of Neurolinguistics, 24*(6), 652–673.

Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology, 12,* 951.

Bird, H., & Franklin, S. (1996). Cinderella revisited: A comparison of fluent and non-fluent aphasia. *Journal of Neurolinguistics, 9,* 187–206.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org

Boyle, M. (2015). Stability of word-retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology, 24,* S953–S960.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277.

Dragoy, O., & Bastiaanse, R. (2010). Verb production and word order in Russian agrammatic speakers. *Aphasiology, 24,* 28–55.

Edwards, S. (1995). Profiling fluent aphasic spontaneous speech: A comparison of two methodologies. *International Journal of Language & Communication Disorders, 30*(3), 333–345.

Faroqi-Shah, Y., & Thompson, C. K. (2007). Verb inflections in agrammatic aphasia: Encoding of tense features. *Journal of Memory and Language, 56*(1), 129–151.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25,* 1414–1430.

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology, 30,* 364–380.

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25,* 1261–1278.

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language 33*(3), 217–222.

Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain, 135*(12), 3815–3829.

Fromm, D., Holland, A., Armstrong, E., Forbes, M., MacWhinney, B., Risko, A., & Mattison, N. (2011). "Better but no cigar": Persons with aphasia speak about their speech. *Aphasiology, 25*(11), 1431–1447.

Jacobs, B., & Thompson, C. K. (2000). Cross-modal generalisation effects of training non-canonical sentence comprehension and production in agrammatic aphasia. *Journal of Speech, Language, and Hearing Research, 43,* 5–20.

Jonkers, R., & Bastiaanse, R. (1996). The influence of instrumentality and transitivity on action naming in Broca's and anomic aphasia. *Brain and Language, 55*(1), 37–39.

Kertesz, A. (2006). *Western Aphasia Battery–Revised*. San Antonio, TX: Harcourt Assessment Inc.

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language, 74,* 1–25.

Kim, M., & Thompson, C. K. (2004). Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and Language, 88,* 1–20.

Kong, A. P. H., Law, S. P., Kwan, C. C. Y., Lai, C., & Lam, V. (2015). A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a Database of Speech and GEsture (DoSaGE). *Journal of Nonverbal Behavior, 39*(1), 93–111.

Kong, A. P. H., Law, S. P., Wat, W. K. C., & Lai, C. (2015). Co-verbal gestures among speakers with aphasia: Influence of aphasia severity, linguistic and semantic skills, and hemiplegia on gesture employment in oral discourse. *Journal of Communication Disorders, 56,* 88–102.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology, 24*(6), 856–868.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn TreeBank. *Computational Linguistics, 19*(2), 313–330.

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex, 20,* 207–220.

Miller, J. F., & Chapman, R. (2000). Systematic analysis of language transcripts [Computer software]. Madison, WI: Language Analysis Laboratory, Waisman Center.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics, 31*(1), 71–106.

Parisse, C., & Le Normand, M. T. (2000). How children build their morphosyntax: The case of French. *Journal of Child Language, 27*(2), 267–292.

Prins, R., & Bastiaanse, R. (2004). Analyzing the spontaneous speech of aphasic speakers. *Aphasiology, 18*(12), 1075–1091.

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology, 30*(1), 45–73.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language, 72*(3), 193–218.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2007, June). High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition* (pp. 25–32). Stroudsburg, PA: Association for Computational Linguistics.

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics, 16*(2–3), 151–167.

Thompson, C. K. (2017). Northwestern Assessment of Verb Inflection (NAVI) [Measurement instrument]. Evanston, IL: Northwestern University.

Thompson, C. K., Ballard, K. J., Tait, M. E., Weintraub, S., & Mesulam, M. (1997). Patterns of language decline in non-fluent primary progressive aphasia. *Aphasiology, 11,* 297–321.

Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B. B., . . . Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology, 26*(1), 20–43.

Thompson, C. K., Shapiro, L. P., Li, L., & Schendel, L. (1995). Analysis of verb and verb-argument structure: A method for quantification of Aphasic language production. *Clinical Aphasiology, 23,* 121–140.

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.

Webster, J., Franklin, S., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of Neurolinguistics, 20,* 363–394.

Wright, H. H., & Capilouto, G. J. (2012). Considering a multilevel approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology, 26*(5), 656–672.

Yarbay Duman, T., & Bastiaanse, R. (2009). Time reference through verb inflection in Turkish agrammatic aphasia. *Brain and Language, 108,* 30–39.