Exploring Pragmatic Language Dysfunction in High Functioning Speakers with Autism Spectrum Disorder: A Reliability Study for Transcription using Systematic Analysis of Language Transcripts (SALT)

Megan Conrad, Brandis Duby, Katryna Lysay, Kalista Smith

Supervisor: Joanne Volden Reader: Karen Pollock

A Reliability Study for Transcription

ABSTRACT

Children with Autism Spectrum Disorder (ASD) commonly experience difficulties with pragmatic language use. To determine how cognitive load affects the pragmatic language use of high-functioning children with ASD, 11 children between the ages of 7 and 12 and diagnosed with ASD performed talk-aloud number line tasks of increasing difficulty. Their performance will be compared to 11 typically developing children matched on chronological age and non-verbal mental age. All children involved in the study spoke English as their first language, and had no known neurological disorder. A future study will focus on the pragmatic performance of the participants and the investigators expect to find that children with ASD make more pragmatic errors than matched controls and that all participants make more errors on longer number lines.. This phase of the study focused on transcription of the data.. With appropriate training, graduate research students transcribed the samples using Systematic Analysis of Language Transcripts (SALT) software. Inter-rate reliability of bettr than 80%was achieved.

According to the Diagnostic and Statistical Manual 5 (DSM-5, American Psychiatric Association, 2013), Autism Spectrum Disorder (ASD) is characterized by deficits in social communication and interaction, as well as restricted and repetitive patterns of behaviour, interests, and activities (American Psychiatric Association, 2013). These difficulties are present from early childhood, but may not be fully exhibited until social demands become too great (American Psychiatric Association, 2013). ASD is described as a spectrum disorder because the severity is not uniform across all of those who are diagnosed. The United States Centers for Disease Control and Prevention (CDC) estimate that the prevalence of ASD in children aged 8 is 1 in 88 (2012).

While those with ASD may also have speech delays or a language disorder and the severity varies across the spectrum, pragmatic deficits and difficulties with social language are universal within the ASD population (Tager-Flusberg, Paul & Lord, 2005). The overarching goal of this project is to investigate whether cognitive load has an effect on the social language use of those with ASD. Social language use is an aspect of pragmatics, which is an important domain of language (Paul& Norbury,., 2012). Characteristics of pragmatic language can be captured using the Pragmatic Rating Scale (PRS) originally developed by Landa et al. (1992) to rate the social language use of parents with children who have ASD. Landa (1992) found that the scale reliably distinguished between a control group of parents and parents of children with ASD. The PRS was later refined by Paul, Orlovski, Marcinko and Volkmar (2009) for direct use with those who have ASD. Items on the PRS are organized into three major groups. They are: pragmatic behaviours, speech and prosodic behaviours, and paralinguistic features. Each item

within a group can be scored as inappropriate, mildly inappropriate, or appropriate.

In order to explore the relationship between changes in cognitive load and pragmatics, language samples were taken from videotaped examiner-child interactions, collected from a project focused on strategic thinking in estimating targets on a variety of number lines. This project provided an opportunity to observe spontaneous pragmatic language in a controlled situation across children with ASD and matched controls. Participants with ASD were compared to matched controls on their ability to place a mark on a number line where they estimated that a target number would be found. The children were given number lines displayed on iPads with only the numbers at each end displayed. They were then asked to tap where certain numbers fell between the two end points. Number lines of varying difficulty were presented to each child. The easiest and most conventional number lines were from 0 to 100. Others were more difficult, such as 0 to 1000, or had more atypical ranges, such as from 0 to 220. Two different types of trials were completed across both the simple and more complex lines. During some trials the children were only required to indicate the location of the number. On other trials the children were required to "think out loud" and explain how they were deciding where the number should go on the line. After they did this, the examiner would then probe their response to gain more insight into why they decided the number belonged in that exact location.

Although the ultimate goal is to examine these language samples for evidence of pragmatic difficulties, the samples needed to be transcribed before they could be scored on the PRS. The current study involved transcription of the videotaped language samples using the Systematic Analysis of Language Transcripts (SALT) software and establishing reliability in that

transcription. The SALT website describes the product as "software that manages the process of eliciting, transcribing, and analyzing language samples. The software includes a transcription editor, standard reports, and reference databases for age-matched comparisons." ("What is SALT?," n.d., para. 1) In this study the videos were transcribed by four Western Canadian graduate research students in the Speech-Language Pathology program at the University of Alberta and reliability was calculated to ensure accuracy and consistency among the four transcribers. In the absence of any information about current standards for transcription reliability,, a standard of 80% or better among four transcribers was targeted to ensure the substantial consistency necessary to proceed with the second phase of this study.

METHODS

Training

Four native-English-speaking graduate research students were trained together on how to transcribe using the SALT software. The students were trained using an older version of the SALT manual and online training courses available on the SALT software website. The newest conventions in the online training courses were used when there were conflicting conventions. SALT allows for unique coding to be defined by the user for purposes of coding aspects of utterances that are unconventional. For this study it was relevant to code utterances that were pragmatically awkward. The unique code decided for capturing this aspect of utterances was [AWK].

The students practiced using the SALT conventions by transcribing two interviews collected for a study of adult conversation. These interviews were transcribed by all the

students and the transcripts compared with one completed by an experienced clinician. The students and clinician discussed differences and discrepancies among their transcripts using both the SALT manual and the website as a reference. These decisions about coding were recorded to ensure consistency throughout future transcripts.

The next step in training was to practice transcribing videos from the number line project. The students transcribed seven videos from the number line project that were not being used for the current study. All four transcripts from the first video transcribed were compared, each discrepancy was discussed, and a consensus was achieved as to the proper transcription. Subsequent videos were compared in pairs – each member of the pair of students compared her own transcript to that of the other transcriber. Comparisons were done word by word, code by code and utterance boundary by utterance boundary in order to identify any discrepancies in transcription. Then inter-rater reliability was calculated on 100 utterances from the middle of the transcript, using the following formula:

$$\text{Reliability (\%)} = \frac{\textit{total agreements for words, codes and utterance boundaries}}{\textit{total agreements + total disagreements for words, codes, utterance boundaries}} \times 100$$

100 utterances from each transcript was deemed to be an appropriate sample to represent the entire transcript. Reliability was calculated by comparing the number of agreements on words, codes and utterance boundaries to the total number of agreements plus disagreements on words, codes and utterance boundaries. Words that were part of mazes (i.e., repetitions and reformulations) or filled pauses were not counted as part of the calculation. Membership in the pair of students coding each particular transcript was balanced so that each

student compared her transcription to each other student at least once. Inter-rater reliability on these practice transcripts was above the target reliability of 80%.

Participants

The experimental group was composed of 11 children with ASD, aged 7-12. The control group consisted of 11 typically developing children in grades 2, 4, and 6. The typically developing children were matched to the children with ASD on non-verbal intelligence quotient (as measured by the Brief Intellectual Ability subtest on the Woodcock-Johnson III; Woodcock, McGrew, & Mather, 2007) and on chronological age.

Transcription

Forty-four videos of twenty-two participants were transcribed. There were two videos from each participant taken during the third and fourth sessions conducted in the number line project. An attempt was made to blind the transcribers to the identity of the children and their group by assigning number codes to each video.

Initially, each video was assigned to two students to transcribe. The pairs were alternated so that each student was paired with each other student equally. Videos were transcribed in ascending numerical order as provided by the principal investigator in the number-line study.

Throughout the transcription process, reliability was periodically assessed by examining the transcripts from both videos from every fifth participant. These transcripts were checked for reliability by comparing 100 utterances of each transcript to the other for word by word, code by code and utterance boundary by utterance boundary agreement, following the same methods used during the training period. In total, 10 videos were checked for reliability

throughout the course of the study. Reliability checks were conducted throughout the transcription process so that any major discrepancies could be discussed and a consensus could be made before further videos were transcribed. Very few discrepancies were encountered throughout the transcription process. The discrepancies that were discussed were minimal (e.g., whether to spell alright as “alright” or “all right”). It was never deemed necessary to go back to previous transcripts and make adjustments.

After the first 13 videos were transcribed by two students each, reliability was found to be above 90% for all reliability checks (i.e., reliability checks done on both videos from participant 100 and participant 105), and it was deemed unnecessary to continue with double transcriptions for every video. Subsequent videos were first transcribed by one student and then another student watched the video while following along on the transcript, fixing any mistakes and making any changes as needed. When a consistent disagreement arose between two transcribers, the discrepancy was discussed and a consensus was achieved. Videos that were checked for reliability (i.e., both videos from participants 110, 115 and 120) were still transcribed by two students individually and compared, in the same manner as the previous reliability checks had been done.

RESULTS

The reliability achieved exceeded the original goal of 80% reliability between raters. Inter-rater reliability was calculated from ten videos that were transcribed by two students. These videos were of every fifth participant after initial reliability was found to be high. Reliability between transcripts remained high, ranging from 91% to 95%, and the average

reliability calculated across the ten transcripts was 93%. Table 1 shows the ten transcripts that were compared, along with the calculations used to obtain the percentage of inter-rater reliability for each. As transcribers were blind to which group participants belonged to, it is unclear whether there was a difference in reliability between the experimental group and the control group. However, as reliability among 10 transcripts remained above 90%, it is unlikely that there was a large discrepancy between the two groups.

Transcript Number	Total Agreements/Total Agreements+Disagreements	Percentage Reliability
100_S3	806/851	95%
100_S4	833/907	92%
105_S3	601/649	93%
105_S4	589/627	94%
110_S3	720/765	94%
110_S4	683/751	91%
115_S3	749/820	91%
115_S4	647/690	94%
120_S3	481/530	91%
120_S4	731/766	95%

Table 1: Inter-rater reliability between two transcribers every five participants.

DISCUSSION

The SALT transcription system provides a coding protocol designed to capture detailed

information about spoken language. These rules were closely followed and high reliability among the four raters was established during the training phase of this study. The procedure then called for monitoring reliability by double-coding every fifth subject, which ensured that consistency was maintained throughout the months of the transcription period. Checking the transcripts of every fifth subject was important to enable discussions around any discrepancies encountered in the study. As a result, high levels of transcription agreement were consistently reached for all examined transcripts across the period of data analysis. While each transcript was verified by only two students, each student obtained high reliability with each of the other three students. This means that collectively, reliability was achieved among all members of the research team.

While the only a portion of the utterances were selected for inter-transcriber reliability purposes, varying which utterances were selected ensured the best representation of the transcript. A total of one hundred utterances were selected with fifty consecutive utterances selected randomly from the first half of the transcript, and fifty consecutive utterances selected randomly from the second half. This encompassed many different contexts that may have arisen throughout the session.

Furthermore, transcription reliability remained relatively stable across transcripts. No significant increase or decrease was found in the percentage of inter-transcriber reliability across the time period of transcription.

The average level of 93% inter-transcriber agreement surpassed the original target level of 80%. These consistently high levels of reliability indicate that the transcriptions were as close as possible to the actual spoken utterances.

Despite the achievement of a high level of reliability, there were some limitations in data collection. Because pragmatic analysis was not the primary purpose of the study, data were not collected with language analysis in mind. The audio component of the video was recorded by the camera as opposed to a table or clip-on microphone. As a result, the speech was sometimes quiet or muffled. On occasion participants walked away from the camera and microphone making the speech indiscernible.

To view the data collected during the number line study, the focus of the camera was the iPad rather than the participant. This left the face of the participant out of the frame of the video and the body was focused towards the iPad rather than the camera or the interviewer. The study of pragmatics is larger than the study of the language that the participants used. Facial expression and body language factor into how we perceive the speaker. As gestures and pragmatic interpretations can be transcribed as comments, these observations should be included in the transcription as they are relevant to the purpose of this study. Without the view of the face and an obscured view of the body, interpretations of body language and other non-verbal cues were unable to be completed. This impeded not only transcription, but will likely influence the pragmatic judgments in the second phase of this study. Emphasis was placed on verbalization of internal mathematical thought for the purposes of the number line experiment rather than pragmatic subtleties of the body language (e.g., sighing in frustration, hesitancy of gestures). While interviewers may have commented on some pragmatic cues of interest (e.g., by offering a break due to signs of fatigue) other possible pragmatic cues could not be observed without video of the face and body of the participants.

The administration of this task may have influenced the transcriptions of the students

by revealing group status. During administration of the number line tasks, examiners offered breaks between the number line tasks and external rewards such as stickers to the participants with ASD, but not to the controls. In addition, the presence of extraneous adults signalled a participant with ASD and on occasion the participants themselves sometimes mentioned ASD. Thus, transcribers were not always blind to group status. Preconceived ideas about group status may influence pragmatic judgments of utterances for the next phase of this study.

The design of the task may also be questioned when determining if this is a suitable way to measure pragmatic skill. This task is very structured. The predetermined topic, and level of detail required, mean that there is less opportunity for a participant to initiate conversation of their choice. Tasks of a more conversational nature, because they are more flexible, often reveal more pragmatic difficulties in the ASD population (Paul et al., 2009). However, it is difficult to control for level of cognitive demand during conversational tasks, which is something that this number line task allowed us to do.

The final consideration for transcription is the degree of effort required to complete the task itself. Transcription of these videos proved a labour intensive task. We estimate that transcriptions of these videos took about one hour per ten minutes of video. Transcribing long videos or large number of participants would quickly become very time consuming. In circumstances with a time constraint, transcription, however beneficial, might not be a realistic option for researchers.

FUTURE DIRECTIONS

The subsequent phase of this study will involve other students using the PRS to rate

each participant in the study based on his or her pragmatic language skills throughout number line activities. Thorough training in using the PRS will be necessary for students to effectively rate each participant. Furthermore, performing regular inter-rater reliability checks between students using the PRS will again ensure consistent and accurate results. The overarching goal of this study is to determine the effect of cognitive load on pragmatic language use in children with ASD. Using the PRS to rate participants during number line tasks with varying degrees of difficulty may reveal a relationship between the complexity of a task and the ability of children with ASD to communicate effectively. The rating scale may also be effective in finding differences in pragmatic language use between children with ASD and matched controls during cognitively demanding activities.

We also recommend that additional procedures during data collection be put in place to minimize complications to the transcription process. Improving the audio quality of the videos would greatly decrease the amount of time spent deciphering what is being said during transcription. For future studies we suggest that clip-on microphones or table microphones be used in order to improve the quality of the audio portion of the videos. Multiple camera angles in which the child's face is shown would also be recommended for future projects. This would further improve accuracy of transcription, as transcribers could read participants' lips as they verbalize. Having a better view of the participants would also allow the transcribers to adequately gauge facial expressions and attitudes of participants. Additionally, transcribers would be able to describe actions and gestures that participants use during the videos more accurately. However, if a full view of participants were to be included, a more systematic way of commenting on participant actions and facial expressions would be necessary in order to

ensure accurate and reliable transcripts. If transcribers agreed on a system to accurately code and comment on non-verbal, pragmatic elements, transcripts could be checked and compared more thoroughly. A more detailed and accurate transcription would be desired when rating each participant with the PRS. Descriptions of facial expressions, attitudes and actions of participants may give raters a better sense of non-verbal pragmatic difficulties.

CONCLUSIONS

The results of this phase of the study show that reliable transcription can be achieved by graduate student clinicians with a relatively short training period. While much can be gained by watching video in terms of affect, tempo, and inflection, the transcript plays an integral role in clarifying ambiguous utterances. Language measures can be easily obtained through the use of language sample analysis software (e.g., SALT). The completed transcript lends itself easily to secondary uses such as coding with the PRS.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th

ed.). Arlington, VA; American Psychiatric Publishing

Centers for Disease Control and Prevention (2012). Prevalence of Autism Spectrum Disorders – Autism and Developmental Disabilities Monitoring Network. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6103a1.htm?s_cid=ss6103a1_w

Paul, R., Orlovski, S. M., Marcinko, H.C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Development Disorders, 39*(1):115-125

Landa R, Piven J, Wzorek M, Gayle J, Chase G, Folstein S. (1992). Social language use in parents of autistic individuals. *Psychological Medicine*, *22*, 245–254.

Paul, R., Norbury, C.F., (2012). Language disorders from infancy through adolescence: Listening, speaking, reading, writing and communicating (4th ed.) St. Louis, MO: Elsevier Mosby.

Self-paced Online Training. (n.d.). In *SALT Software.* Retrieved from http://www.saltsoftware.com

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, D. Cohen (Eds.) *Handbook of autism and pervasive developmental disorders (pp. 335-365).* Hoboken, New Jersey: John Wiley & Sons, Inc.

What is SALT?. (n.d.). In *SALT Software.* Retrieved from http://www.saltsoftware.com

Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.