

Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders

Johanna K. Lake · Karin R. Humphreys · Shannon Cardy

Published online: 10 November 2010
© Psychonomic Society, Inc. 2010

Abstract This study investigates the role of disfluencies such as “um” or “uh” in conversation to discern whether these features of speech serve listener- or speaker-oriented functions by looking at their occurrence (or lack of occurrence) in the speech of participants with autism. Since the characteristic egocentricity of individuals with autism means they should engage in minimal listener-oriented behavior, they are a useful group to differentiate these functions. Transcription, analysis and categorization of 26 spontaneous language samples were derived from age-matched native English-speaking controls and high-functioning individuals with Autism Spectrum Disorders (ASDs). Results showed that individuals with ASD produced fewer filled-pause words (ums and uhs) and revisions than controls, but more silent pauses and disfluent repetitions. Filled-pause words therefore appear to be listener-oriented features of speech.

Keywords Autism · Pragmatics · Disfluencies · Language production · Conversation

The language mechanisms underlying conversation are highly complex and intricately coordinated, yet they are largely effortless processes. One question that arises is to what extent this ease of communication is due to the work being done by the speaker or by the listener, the latter often in spite of a speaker’s failure to be helpful. The goal here is to identify what aspects of a speaker’s behavior can be

identified as cooperative, or otherwise. In the past, some approaches to this problem have involved referential communication tasks to better understand the roles of speakers and listeners in communicative exchanges (see Clark & Fox Tree, 2002; Fox Tree, 2001; Fox Tree & Schrock, 1999; Girbau, 2001; Horton & Keysar, 1996; Mangold & Pobel, 1988). This research has had success in determining which aspects of speech are helpful for a listener, but it remains unclear to what extent this is being done by the speaker for the benefit of the listener (i.e., is listener-oriented), or whether it is merely a regularity in the speaker’s behavior that a listener may be able to exploit, and is not performed by the speaker with the listener’s needs in mind (speaker-oriented) (see Bock, 1996; Brennan & Clark, 1996; Fox Tree & Schrock, 1999). For example, recent research has suggested that disfluencies such as filled-pause words serve a useful discourse function, indicating that the speaker is not finished speaking, and is trying to compose their next thought or find the correct word, and that listeners are able to utilize this information (Fox Tree, 2001). However, it is still unclear whether this is being done for the benefit of the listener.

In this work, we make the assumption that individuals with autism spectrum disorders will not tend to display elements of speech that are listener-oriented, thereby providing an alternate form of evidence as to whether certain discourse behaviors within typical populations are listener- or speaker-oriented. Individuals with autism display a variety of impairments in many areas including social skills and language development. One of the most profound characteristics of individuals with autism is egocentricity, which manifests as a lack of interest in interacting with other people, a failure to develop social relationships and difficulty with social interactions. It has been suggested that underlying these problems may be a

J. K. Lake · K. R. Humphreys (✉) · S. Cardy
Department of Psychology, Neuroscience & Behaviour,
McMaster University,
1280 Main St West,
Hamilton, Ontario L8S 4K1, Canada
e-mail: krh@mcmaster.ca

problem of “theory of mind,” which describes how individuals with autism are unable to form representation of another’s mental state (American Psychiatric Association, 2000; Baron-Cohen, 1995). A critical pragmatic aspect of conversational language use is the ability and willingness to recognize the listener’s perspective and knowledge, a task often referred to as establishing ‘common ground’ (Clark, 1996). Common ground involves understanding the speaker’s intention and beliefs such that a shared understanding of mental state is developed (Clark, 1996). Typical speakers are generally very good at using this knowledge and carry on conversations with little effort (e.g., Brennan & Clark, 1996). In contrast, individuals with autism tend to take egocentric approaches to conversation and usually have poor pragmatic skills; this is true even of very high-functioning individuals with autism, who are not judged to be language impaired by usual measures (Baltaxe, 1977; Bishop, 1998; de Villiers, Fine, Ginsberg, Vaccarella, & Szatmari, 2007; Fine, Bartolucci, Ginsberg, & Szatmari 1991; Wetherby & Prutting, 1984; Young, Diehl, Morris, Hyman, & Bennetto, 2005; Ziatas, Durkin, & Pratt, 2003).

This group, therefore, presents an opportunity to explore which functions of language are produced by a speaker for the benefit of the listener and which are independent of the perceived needs of the listener. Specifically, we predict that if high-functioning individuals with autism are seen to produce specific pragmatic aspects of speech at a normal rate, that feature is likely *not* being done for the benefit of the listener. Conversely, for those pragmatic aspects of speech that are relatively absent in an individual with autism, this constitutes some evidence that this feature may be listener-oriented in normal speech.

In this study we specifically examine the role of disfluencies in speech. It has been suggested that the use of filled-pause words or disfluencies in normal speech, such as *um* and *uh*, may represent an important role in conversation. Fox Tree (2001) examined the effect of *ums* and *uhs* during on-line processing of speech, and showed that *um* and *uh* may be utilized by a listener to facilitate conversations. *Uh* appeared to signal an upcoming short delay, while *um* a longer delay. The use of *uh* increased the speed at which listeners were able to recognize words; however, *um* had no effect on listeners’ speech recognition (Fox Tree, 2001, 2002). Fox Tree suggests that *ums* and *uhs* help listeners by alerting them that the speaker is still speaking (that it is not the listener’s turn yet) and indicating the length of the upcoming delay in speech. However, we do not know if speakers are intentionally using this function of speech to aid the listener or if this is merely a regularity in the speaker’s behavior that listeners are able to take advantage of. By studying the speech of individuals with autism, who by definition are unlikely to engage in listener-oriented functions of speech, this can provide evidence as

to whether these types of disfluencies appear to be a speaker- or listener-oriented function.

Method

Participants

Participants with autism were recruited from a facility in Hamilton, Ontario, Canada, providing services to high-functioning individuals with autism spectrum disorders. Fourteen native English-speaking individuals with ASD (13 male) took part in the experiment, all of whom had been diagnosed by an outside agency. One male participant was subsequently excluded for a verbal IQ score that fell below the normal range. According to the Autism Diagnostic Observation Schedule (ADOS: WPS Version; Lord, Rutter, DiLavore, & Risi, 1999), as administered by the facility, six of the remaining participants had a prior diagnosis of autism spectrum disorder, four Asperger’s syndrome, and three autism. The ADOS was repeated by a psychiatrist at McMaster University for four of the participants (others were not retested due to time constraints), and in all cases, the original diagnosis was confirmed. The mean age of participants with autism was 27 years, with a range of 19–35. Wechsler Adult Intelligence (WAIS, Wechsler, 1939) scores were obtained for participants with ASD with an average verbal IQ of 99 and a range of 83 to 117, all within normal ranges. Thirteen age- and gender-matched control participants also took part in the experiment. Control participants were native English-speaking students of McMaster University and members of the community who volunteered to participate. IQ information was not obtained for controls nor were the participants matched in terms of education level.

Materials

A spontaneous language sample was obtained from a 5–10-min recorded conversation. Participants were asked a variety of general questions related to their interests and hobbies. Following each question, participants were given roughly 5 s (as estimated by the trained experimenter) to respond before the experimenter used further prompting to achieve a reply. The same set of questions was used for both groups, and all conversations were digitally recorded. Two experimenters listened to these recordings and transcribed the conversations using SALT software (Systematic Analysis of Language Transcripts, Miller, & Chapman, 1983). Transcriptions were completed independently and then compared, with discrepancies resolved by one of the original transcribers. As per SALT conventions,

the first 49 utterances produced by each participant were analyzed using SALT guidelines with regard to syntactic, phonological, semantic, and pragmatic properties. Transcripts were further categorized according to the rate per hundred words of revisions, filled pauses (ums and uhs), silent pauses (greater than 2 s), and disfluent repetitions.

Results

Conversation sample

The mean length of utterance (MLU) for control participants was 9.1 words, and for ASD participants, 5.7 words; the latter tended to answer questions with shorter responses, particularly one-word replies. However, even when one-word utterances were excluded from the analysis, the MLU for control participants was still larger, ranging from 7.7 to 11.5 compared to 4.6 to 8.8 for ASD participants. Informal analysis of the conversation samples of ASD participants revealed no obvious deficits, and semantic and syntactic aspects of speech were comparable between the groups.

Throughout the conversation samples, ASD participants responded to 84.5% of questions, compared to 99% for control participants. The experimenter frequently had to pose and re-phrase the questions several times before obtaining a response from participants with autism. Disfluencies were coded into four categories: revisions, repetitions, silent pauses, and filled pauses. Table 1 gives examples of each of these types of disfluencies. Figure 1 shows boxplots of the disfluency rates per hundred words by group, demonstrating a striking lack of overlap between the groups' distributions. A series of Mann-Whitney tests for independent samples revealed significant differences between the control group and individuals with ASD with respect to disfluencies per hundred words. Participants with ASD were found to produce fewer filled-pause words (ums and uhs) than control participants, $U(25) = 150$, $Z = 3.36$, $p = .001$ (means of 1.7 vs. 5.0). Conversely, participants with ASD produced more silent pauses than control participants $U(25) = 169$, $Z = 4.33$, $p < .001$ (with means of 4.0 and zero, respectively). Of the silent pauses produced by individuals with ASD, 68% occurred at the beginning of an utterance and 32% within

utterances. Participants with ASD produced significantly fewer revisions than controls $U(25) = 129$, $Z = 2.28$, $p = .02$ (2.7 vs. 3.8), but significantly more disfluent repetitions than $U(25) = 169$, $Z = 4.33$, $p < .001$ (4.71 vs. .49).

Another important way in which the groups differed was MLU, which could potentially account for the effects seen here (i.e., a greater MLU may lead to more opportunities for disfluencies). To investigate this possibility we plot MLU against the various disfluency measures, organized by group in Fig. 2. In all four disfluency measures, while we see a clear relationship with MLU, it should be noted that the two groups show separate regression lines, with different intercepts, even when MLUs overlap.

Discussion

There are several aspects of conversational speech that could be listener- or speaker-oriented. The use of filled pauses during disfluencies, including ums and uhs, appear to help listeners (Fox Tree, 2001), but it is unclear whether these are produced for their benefit.

Their characteristic egocentricity, perhaps due to challenges in understanding the perspective of another – theory of mind deficits – makes individuals with autism by definition unlikely to engage in listener-oriented behavior. Therefore, if individuals with autism employ certain aspects of speech in the same way as typical speakers, we argue that this feature must not be listener-oriented and if they do not employ it, this is some evidence that it may be listener-oriented.

Results of the present experiment demonstrated that participants with ASD produced far fewer filled-pause words than controls. Interestingly, ASD participants appeared to use silent pauses in the place of filled pauses. ASD participants used far more silent pauses than controls and engaged in these silent pauses at virtually the same rate as control participants used ums and uhs. However, unlike filled pauses, silent pauses made it difficult for the speaker to know when the listener was finished speaking. In this sense, silent pauses may reflect the same speaker-originating disfluencies in production, but do not attempt to remediate the potential confusion they cause to an interlocutor. One issue that needs to be taken into account is that our coding scheme, by counting only quite lengthy

Table 1 Examples of disfluencies

Disfluency category	Example utterance
Filled pause	“My favorite...um...animal...uh...is an um...cat”
Silent pause >2 s	“My favorite...(>2 s silence) animal is a ...(>2 s silence) cat”
Revision	“My fat...favorite...best animal is a dog...my favorite animal is a cat”
Repetition	“My favorite..my favorite animal is a cat.

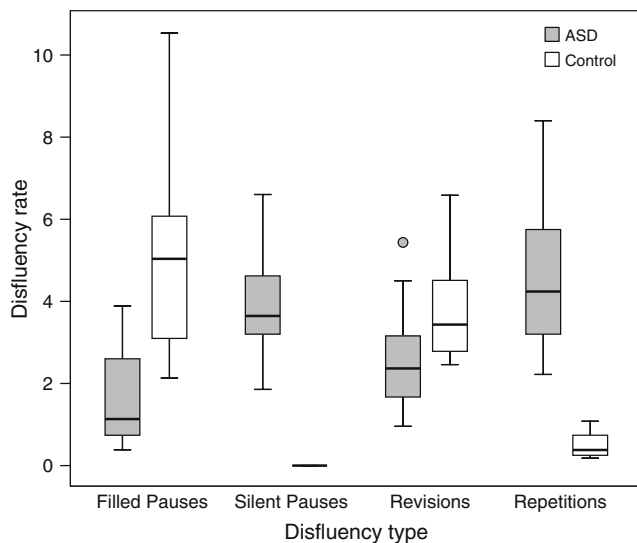


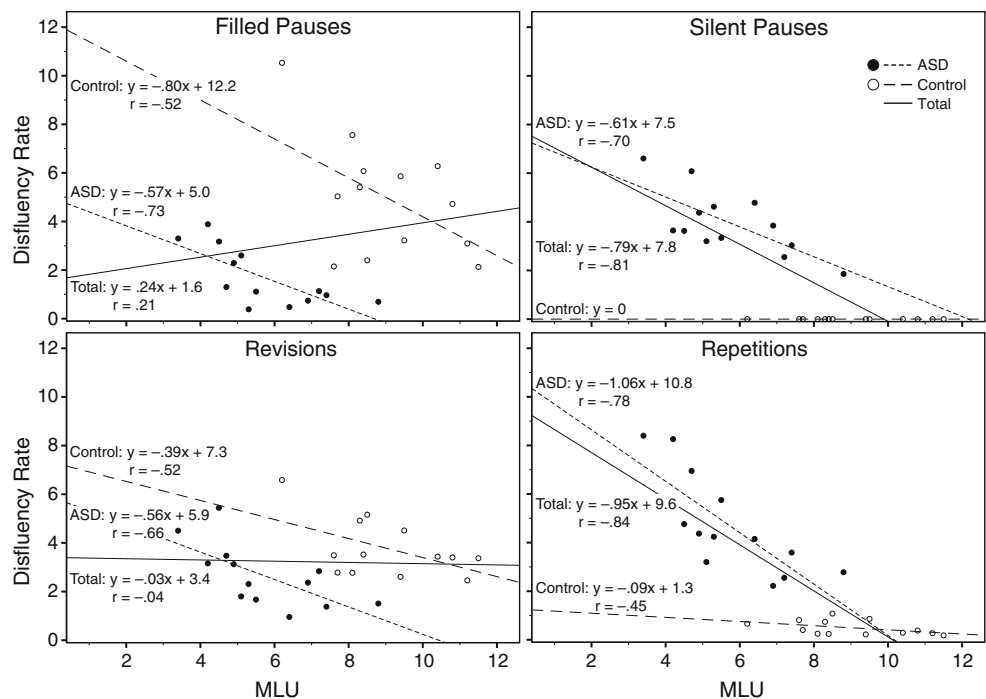
Fig. 1 Note: This figure demonstrates the general lack of overlap between the disfluency measures for ASD and control groups. For filled pauses, the median score for the ASD group was less than the minimum for the control group; this was true even for the median score of only the six ASD participants with VIQs above 100 (1.7). For silent pauses, there was no overlap between the distributions; median for ASD participants with VIQs above 100 was 3.7. Revisions showed somewhat more overlap, although the median number of revisions for the ASD participants was still below the minimum score for the control participants (2.4 vs. 2.5), and the median for the speakers with VIQs over 100 was slightly above the first quartile for the control speakers (2.9 vs. 2.8). For repetitions, there was also no overlap at all between the distributions, and the median for speakers with VIQs over 100 was 3.0

silent pauses, likely missed some shorter disfluent silent pauses from both groups. This was done in order to ensure that we excluded normal prosodic pauses. However, with a less strict criterion, the two groups would probably look more similar on this measure. It is also important to note that all speakers with ASD did use at least some filled pauses, and the current data cannot distinguish between the potential explanations that either some filled pauses are not listener-oriented, or that these individuals with ASD did engage in some listener-oriented behavior.

Participants with ASD also revised their speech significantly less often than controls. Belser and Sudhalter (2001) also found low levels of revisions in lower functioning young adults with ASD, as well as in the speech of individuals with mental retardation, although this was not in comparison to typical controls. Revising speech involves self-repair, whereby a speaker detects a problem and formulates a revision or replacement to correct it (Levelt, 1983). Given this information, one could conclude that participants with autism either make fewer mistakes, or don't detect problems in their own speech the same way as controls do. One other alternative we suggest is that they may be able to detect their own formulation problems adequately, but may be less aware of the problems this may have caused a listener, and are therefore less likely to attempt to clarify and revise their utterance to aid a listener. The data here cannot distinguish between these possibilities, but it remains an intriguing possibility for follow-up work.

We also find that individuals with ASD in the study showed far more disfluent repetitions of elements of their

Fig. 2 Note: For each disfluency type, regression lines are plotted for ASD and control groups individually, as well as for the ungrouped total. In all cases, while there is a strong linear relationship with MLU and disfluency rate, the two groups show notably different regression lines, with different intercepts



speech than did our controls. This helps to distinguish between alternative accounts of repetitions, as described in Clark and Wasow (1998). One explanation is that repetitions are akin to filled pauses, in that they are an attempt to hold the floor, and maintain fluency, and as such are listener-oriented (part of Clark & Wasow's "continuity" hypothesis). However, our results here, showing that adults with ASD are more likely to use disfluent repetitions, suggest that repetitions are not listener-oriented. They also show that repetitions tend to pattern very differently in these groups than do revisions and filled pauses. However, Clark and Wasow (1998) point out that speaker's preference for continuity could be either due to speaker-oriented or listener-oriented processes. If they are not listener-oriented, and designed to help an addressee cope with disrupted speech, it still might be easier for a speaker to repeat elements after a disruption in order to produce a full constituent rather than a fragment. Alternatively, our results are also consistent with the hypothesis that repeated words are not an attempt to restore fluency, but are the more automatic result of immanent errors detected in one's speech plan (e.g., Levelt, 1983).

While we argue that these results demonstrate the use of filled pauses as a listener-oriented behavior, the question remains, to what extent this is a volitional choice. The experience of at least many speakers seems to be a limited ability to inhibit ums and uhs, even if speaking only to themselves, making a purely volitional account problematic (although see Fox Tree, 2007, for a report that people are able to suppress these sometimes). We suggest instead that ums and uhs may become a habitual part of speech in typically developing children resulting from a responsiveness to interlocutors' states of mind. When one is interrupted when pausing before finished speaking, it seems likely that theory of mind reasoning would be required to understand that the interlocutor mistook the silence for the end of the utterance and that filling the pause with verbal material would be required to hold one's turn. Similarly, lengthy silent pauses typically make one's interlocutors quite uncomfortable. Anecdotal experience from this study was that experimenters found it very awkward to simply wait for participants with ASD to resume speaking, and that at times it was difficult to follow experimental protocol and not fill the silence themselves. On the other hand, participants with ASD appeared to either not perceive, or at least not be concerned by, any potential discomfort on the part of their conversational partners.

One limitation of this study is the fact that while the participants with ASD were high functioning and had good verbal skills, the control participants were not matched to them on IQ or education, only on age and gender. While we do not have IQ information on our control participants, it seems likely that their scores would have been higher than

those of the ASD participants, although there should be at least some overlap. Furthermore, there was a sizable difference in MLU, which could allow different levels of disfluency opportunities. However, inspection of the distributions suggests that while both VIQ and MLU do show some associations with measures of disfluency, they alone do not appear to account for the large differences we see between the groups.

We do note, however, that these findings are not entirely consistent with some other reports in the literature. Shriberg et al. (2001) found that in a sample of high-functioning males with ASD aged 10-50, the ASD participants showed an increased rate of disfluencies, which they describe as an increased rate of one word repetitions and revisions, as compared to controls. However, their data actually show that while the ASD participants did indeed show significantly higher rates of one word repetitions, it was in fact the controls who showed significantly higher rates of revisions than the ASD participants (Fig. 1, panel C, p 1105). Although the discussion in Shriberg et al. (2001) glosses over the distinction between the revisions and repetitions, and states that the ASD participants showed greater rates of disfluency, their data are very much in line with the findings that we present here, in that our ASD participants also showed decreased rates of revisions as compared to controls, but increased rates of repetitions. Thurber and Tager-Flusberg (1993) showed that a sample of 12-year-old children with autism showed *lower* rates of silent pauses within phrases than did typical children matched for verbal mental age (approximately 8 year olds), which they attribute to lower levels of communicative and cognitive demand from the stories told by the children with autism. However, they did not report filled pauses. This discrepancy between findings has several possible sources – these were from children with autism, who were also relatively lower functioning. Furthermore, our silent pauses were defined to be significantly longer than the brief hesitations described in this study. However, a more complete understanding will require an investigation of the developmental trajectory of disfluencies in children with ASD.

These results add further support to the findings of Fox Tree and colleagues (Clark & Fox Tree, 2002; Fox Tree, 2001; Fox Tree & Schrock, 1999) in showing the useful nature of ums and uhs for both speakers and listeners in conversation. We also provide convergent evidence for the idea that ums and uhs are not simply meaningless fillers that listeners have opportunistically discovered how to make use of. Instead, we find that speakers with ASD who have normal verbal IQs, who are by definition egocentric (and therefore not likely to be listener-oriented), do not use ums and uhs, but instead appear to use silent pauses. We therefore argue that ums and uhs have likely become part of normal speaking as a response to listeners' needs, even if we eventually lose some volitional control over their usage.

Acknowledgements We thank Mel Rutherford and the Human Development Lab for help with recruiting participants for this study. We also thank Megan Keilty for assistance with data collection, and Elizabeth Hall, Amy Beth Warriner, Scott Watter, and Judith Shedden, as well as members of the Cognitive Science Lab, and two anonymous reviewers for helpful comments and suggestions. This work was supported by Natural Sciences and Engineering Research Council of Canada grant no. 293145 to KRH. Correspondence regarding this article should be addressed to either Johanna Lake<lakejk@mcmaster.ca>or Karin Humphreys<krh@mcmaster.ca>, Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main St West, Hamilton, Ontario, Canada, L8S 4 K1.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Baltaxe, C. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology*, *2*, 176–180. doi:10.1093/jpepsy/2.4.176
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Belser, R. C., & Sudhalter, V. (2001). Conversational characteristics of children with fragile X syndrome: Repetitive speech. *American Journal of Mental Retardation*, *106*, 28–38.
- Bishop, D. V. M. (1998). Development of the children's communicative checklist (ccc): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry*, *39*, 879–891. doi:10.1017/S0021963098002832
- Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, *3*, 395–421.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.
- Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111. doi:10.1016/S0010-0277(02)00017-3
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242. doi:10.1006/cogp.1998.0693
- De Villiers, J., Fine, J., Ginsberg, G., Vaccarella, L., & Szatmari, P. (2007). A scale for rating conversational impairment in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*, 1375–1380. doi:10.1007/s10803-006-0264-1
- Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *32*, 771–782. doi:10.1111/j.1469-7610.1991.tb01901.x
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, *29*, 320–326.
- Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes*, *34*, 37–55. doi:10.1207/S15326950DP3401_2
- Fox Tree, J. E. (2007). Folk notions of um and uh, you know, and like. *Text & Talk*, *27*, 297–314. doi:10.1515/TEXT.2007.012
- Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*, 280–295. doi:10.1006/jmla.1998.2613
- Girbau, D. (2001). Children's referential communication failure: The ambiguity and abbreviation message. *Journal of Language and Social Psychology*, *20*, 81–89. doi:1177/0261927X01020001004
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117. doi:10.1016/0010-0277(96)81418-1
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104. doi:10.1016/0010-0277(83)90026-4
- Lord, C., Rutter, M., DiLavore, D. C., & Risi, S. (1999). *Autism diagnostic observation schedule*. Los Angeles, CA: Western Psychological Services.
- Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language Social Psychology*, *7*, 181–191. doi:10.1177/0261927X8800700403
- Miller, J. F., & Chapman, R. S. (1983). *Systematic analysis of language transcripts (SALT)*. San Diego: College Hill Press.
- Shriberg, L. D., Paul, R., McSweeney, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*, 1097–1115. doi:10.1044/1092-4388(2001/087)
- Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders*, *23*, 309–322. doi:10.1007/BF01046222
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wetherby, A., & Prutting, C. (1984). Profiles of communicative and cognitive social abilities in autistic children. *Journal of Speech and Hearing Research*, *27*, 364–377.
- Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, Speech, and Hearing Services in Schools*, *36*, 62–72. doi:10.1044/0161-1461(2005/006)
- Ziatas, K., Durkin, K., & Pratt, C. (2003). Differences in assertive speech acts produced by children with autism, Asperger syndrome, specific language impairment, and normal development. *Development and Psychopathology*, *15*, 73–94. doi:10.1017/S0954579403000051