Interactions Between Turn-taking Gaps, Disfluencies and Social Obligation

Rebecca Lunsford, Peter A. Heeman, and Jan P. H. van Santen
Center for Spoken Language Understanding, OHSU, Beaverton, OR, USA
rebecca@luns4d.com

Abstract
Speakers strive to minimize inter-turn gaps when engaged in a dialogue. However, little work has addressed what impact this might have on the fluency of the following speech. In this paper we explore whether there are interactions between turn-taking gaps and turn-initial disfluencies and if the social pressure to respond to questions plays a role in that interaction. Our results indicate that child speakers are more likely to become disfluent both after a question and as the gap length increases, and that the two interact to further increase the likelihood. We also compared the speech of children with Typical Development (TD) to those with Autism Spectrum Disorder (ASD) or Developmental Language Disorder (DLD), where we found that those with ASD were less likely to become disfluent after a question. This finding suggests that the trade-off between timing and disfluencies is driven by social obligation, and that speakers are willing to tolerate disfluencies so as to maintain a short delay.

**Index Terms:** turn-taking gaps, disfluencies, social pressure, language impairments

1. Introduction

Although human listeners are particularly apt at interpreting speech that includes disfluencies [1, 2, 3], disfluent speech provides challenges for Spoken Dialogue Systems (SDS). From an interface design perspective, knowing what factors increase a user’s likelihood of initiating their utterance with less easily recognized speech could help designers build systems that manage the interaction so as to avoid these types of disfluencies.

In previous work, we found that children with ASD had significantly longer gaps after questions than did children with TD, but that the two groups did not differ in the length of their gaps after non-questions (e.g., statements) [4]. This finding suggests that gaps lengths after questions are impacted by social or language processing (dis)abilities. This paper builds on that work by exploring whether gaps interact with disfluencies, and whether speakers are more likely to produce turn-initial disfluencies after being asked a question. We also explore whether social or language processing deficits impact the likelihood of turn-initial disfluencies.

2. Background and Related Work

SDS will typically predefine some “time-out” value, after which, if the user has not responded, it is assumed that a user needs re-prompting or assistance [5]. This appears a reasonable assumption, in that people strive to minimize gaps during turn transitions [6]. In addition, inter-turn gap length is consistent across languages and cultures, with responses generally forthcoming within 0.5 seconds [7].

Yet, the length of gaps can vary. For example, gap lengths tend to be shorter when no visual cues of attention are present (i.e., during a telephone conversation) [8], a common context for SDS. In addition, gaps tend to be longer when the responding utterance is longer [9], when the responder needs clarification [10], and when asked a difficult question [11].

Responses to questions are of particular interest, in that, by asking a question a speaker places an obligation to respond on the listener [12], and this impacts how the listener responds. In work examining the way speakers respond to factual questions, Smith and Clark [11] found that when speakers are uncertain whether they know the answer, they will respond more slowly, give a non-answer more quickly, and add the fillers “um” and “uh”. They suggest that a desire to preserve self-presentation prompts responders to avoid excessive delays, which might cause the questioner to view them as uncooperative or slow-witted. Instead the responder will provide signals to account for delays and indicate their level of confidence in their answer. Thus, it appears that the obligation to respond places both social and cognitive pressure on the listener.

Lickley [13] found that the highest rate of disfluent words are found in replies to wh-questions (e.g., “Who”, “What”), negative replies, and instructions and clarifications. This work did not separate turn-initial from utterance-medial disfluencies, thus it is unclear whether responding to a question results in generally disfluent speech, or if the effect is localized to the beginning of the utterance. However, it seems likely that these disfluencies would be more prevalent in the turn-initial position because, due to the social pressure to respond in a timely manner, a listener could start to respond before they have fully planned their response. However, to the best of our knowledge, no previous work has explored whether there is an interaction between disfluencies, gaps, and social pressure.

2.1. Use of Children with ASD and DLD

ASD and DLD are both conditions that impact a child’s ability to engage in spoken communication. Children with ASD are characterized by impaired reciprocal social interaction and communication, repetitive behaviors, and restricted interests according to the APA’s DSM-IV-TR [14]. In terms of communication, even high-functioning children with ASD, who have semantically and syntactically correct spoken language, will have faulty pragmatics due to difficulties in understanding and using social cues during conversations. Children with DLD are characterized by an inability to communicate in a manner appropriate to the child’s age, in which the inability is not attributable to physical or intellectual impairments, or ASD [14]. These inequalities can include shortcomings in expressive language, resulting in difficulties organizing and formulating an utterance, or in receptive language, leading to difficulty in comprehending others’ speech.

In children with TD, the ability to recognize, and respond, to turn-release cues is learned early in childhood, starting in
infancy [15]. As such, we can expect that children with TD have a good sense of the rules, and will respond appropriately to the cues that indicate a speaker intends to release the turn. For this research, we include children with TD, DLD and ASD to provide insight into the extent turn-taking is impacted by social pressure. If it is, we would anticipate that the children with DLD would manage turn-taking in a manner similar to those with TD. In contrast, if the DLD group resembles the ASD group, it is likely that turn-taking is more strongly impacted by processing (dis)abilities.

3. Data and Coding

The data used in this paper was collected during administration of the Autism Diagnostic Observation Schedule [16], in which 91 children, 27 with TD, 21 with DLD, and 43 with ASD, engaged in different activities with a clinician. The children ranged in age from 3.9 to 8.9.

The activities consisted of having a conversation, describing a wordless picture or book, and playing with toys. Each of these activities is designed to allow the clinician to observe different communication skills. The Conversation activity allows the clinician to observe how aware the child is of his own emotions. In this activity, the clinician asks the child about their personal experiences with different emotions and the physical manifestations of emotion. During the Description activity, the clinician and child peruse and discuss a wordless picture/book. The clinician then tells a story from her past, relating it to the picture/book. In this activity, the clinician observes whether the child engages the clinician, interacting with the clinician to gain a better understanding of the story. During the Play activity, toys are made available for the child and clinician to engage in joint play. This activity allows the clinician to observe how well the child engages in interactive play and maintains a joint activity.

The clinician and child’s speech was transcribed using Praat [17] and segmented into communication-units (C-units) [18]. Semantically and syntactically complete C-units were transcribed with ‘.’, ‘!’, and ‘?’; and ‘<’ was used to mark incomplete ones. Overlaps between the child’s and clinician’s speech was annotated by placing the overlapped regions of the transcript within angle brackets (i.e., ‘<’). Finally, the activity was annotated in a separate tier.

Here, we use data only for those turn-exchanges from clinician to child in which there is no overlap annotated at the beginning of the child’s C-unit, or at the end of the clinician’s C-unit. Data were also excluded if the clinician’s speech was annotated as incomplete.

3.1. Gaps

For gap lengths, we used the CSLU speech recognizer to refine the silence durations, via a forced alignment between the text transcriptions for the clinician’s speech proceeding the gap, and the child’s speech following the gap. The resulting end and start times were then used to compute an unbiased measurement of the gap lengths. Gap lengths, in milliseconds, were then log-transformed. Each subject’s far outliers (i.e., those outside 1.5 * inter-quartile range) were removed. Far outliers accounted for 2.4% of the data.

3.2. Mazes

As defined by the Systematic Analysis of Language Transcripts (SALT) guidelines [18], mazes include false starts, repetitions of a word or phrase, revisions, fillers, and stutters. Speech identified as mazes was placed within parenthesis (i.e., ‘( )’). Each C-unit was annotated as starting with, or without, a maze.

3.3. Length

As utterance length (in terms of the number of words) has been shown to impact a subject’s rate of disfluencies [19], we include the length of each C-unit as a potential predictor of turn-initial mazes. Length was log-transformed prior to analyses.

3.4. Questions

For each turn-initial C-unit produced by a child, it was determined whether the preceding clinician C-unit was a non-question or a question. In addition, we determined whether each clinician question was a consecutive question, i.e., was immediately preceded by a clinician question with no intervening child speech.

3.5. Analysis

To determine what factors influence the likelihood of a child producing an utterance-initial maze, we created mixed-effect logistic regression models for each group of children, using lmer [20]. These models predict the log-odds of the C-unit starting with a maze. Prior to analyses, gap length was centered about the mean to avoid potential co-correlation effects. The models were iteratively refined, removing factors that did not contribute either alone or by interaction with other factors.

To determine the goodness-of-fit for each model, the model’s predictions are correlated to the actual values. However, as a correlation between the binary responses and the continuous predicted probabilities would not produce usable results, the reported r values were computed thusly: (1) the predicted probabilities were sectioned into 10 bins of size 0.1 (e.g., 0-0.1, 0.1-0.2, etc.); (2) the mean of the predicted probabilities was computed for each bin; (3) the mean for the actual responses (i.e., 0= no maze, 1= maze) was computed for each bin; and (4) the mean probabilities are correlated to the mean actuals.

4. Results

Data for 19,624 turns were available for analyses; 5529 for the TD group, 4767 for the DLD group, and 9328 for the ASD group.

Table 1 shows the models created to ascertain if the likelihood of a turn-initial maze is related 1) to the child’s inter-turn gap length (Gap), and 2) to whether the clinician produced a non-question, a question (Question), or asked two or more consecutive questions (Question+) prior to the child’s response. As it has been shown that the rate of disfluencies is correlated to the length of the utterance [19], we also include 3) C-unit length (UtterLength) in the models. In previous work, we found that the rates of utterance-initial fillers were lower during the Description activity [21], thus we also include 4) Activity. Factors that did not contribute either alone or by interaction with other factors (e.g., sex and age) are not included in these final models.

For these models, INTERCEPT represents the predictions using the reference levels for each factor, which are ACTIVITY=Conversation, Non-question, log(UtterLength)=0, and mean log(Gap). The remaining coefficients indicate offsets from the Intercept for other factor levels. So, for example, the
negative values in the second row indicate that the children are less likely to produce a maze during the Description activity than during Conversation. The second-to-last row shows the standard deviation of the subjects' individual offsets, and the last row shows the goodness-of-fit measure $r$.

### 4.1. By-group Analysis

First, looking at the coefficients for the TD, DLD and ASD groups' models in Table 1 we see, as anticipated, that both Activity and Utterance Length significantly contribute to the likelihood of a maze, with utterance length increasing the likelihood (i.e., TD=1.07, DLD=1.20, ASD=1.25) and the activities Description (i.e., TD=0.82, DLD=0.45, ASD=-0.25) and Play (i.e., TD=-0.20, DLD=-0.38, ASD=-0.16) decreasing the likelihood.

Next looking at gaps, we see that the likelihood of a maze increases with the length of the gap.\(^1\) In addition, we see that Gap and Utterance Length interact, with a negative coefficient. Put simply, this interaction means that a short gap followed by a long utterance will be more likely to have a maze than a long gap followed by a short utterance.

Looking next at questions we see that having been asked a question, or a series of consecutive questions, increases the likelihood that the child will produce a maze. In addition, the interaction between Question and Gap shows that gap length has a greater effect on the likelihood of a maze following a question.

Finally, we see that the included factors produce well-fitted models for the children with TD and ASD (both $r$’s >0.90), but a less well-fit model for the DLD group.\(^2\)

### 4.2. Between-group Comparisons

We look next at the combined model in the last column of Table 1. As we are interested in how the children with language disorders (i.e., DLD and ASD) differ from those with TD, here we chose DX=TD as the reference level. Additionally, for this model, we collapsed Question and Question+ and did not include interactions between DX and Gap or interactions with Gap as these exclusions resulted in a better fitting model. The resulting model had an $r=0.98$.

Here we see that the children with DLD did not differ significantly from the children with TD (i.e., DX:DLD and the interactions with DLD are not significant). Comparing the children with TD to those with ASD, we see that the ASD group is more likely to produce a maze during the Description activity, and is more likely than the TD group to produce a maze as their utterance length increases. Of particular interest, we see that the ASD group is less likely to produce a maze after a question.

### 5. Discussion

In this paper, we examined whether mazes (i.e., disfluencies), inter-turn gaps and social pressure are related. For the children with TD, we found that gaps and utterance-initial mazes do interact, with the likelihood of a maze increasing as the length of the gap increases. Furthermore, we found that when the social pressure to respond was increased (i.e., following a question), children with TD were even more likely to produce an utterance-initial maze and that this likelihood increased with the length of the preceding gap. These findings suggests that child-

---

\(^1\)This coefficient is not significant for the DLD group.

\(^2\)The $r$ computed here is particularly sensitive to bins with sparse data. For example, the low $r$ value for the DLD group was due to a bin containing a single prediction of .72, in which no maze was present in the actual data. An alternate treatment, which divides the data into slightly overlapping bins, each containing an equal number of data points, resulted in $r$ values of over 0.98 for all models.
children react to the social pressure to respond in a timely manner by either speaking before fully prepared or by producing a filler. For the children with DLD, gap length alone did not appear to influence the likelihood of a maze. However, for this group we did find an effect of questions and an interaction between gaps and questions. Once again, this result suggests that these children have an increasing likelihood of producing mazes with increased social pressure to respond.

Comparing the children with ASD to those with TD, we found that the children with ASD were less likely to produce a maze after a question. Viewed in concert with our previous finding that children with ASD have longer gaps [4], this result suggests that the children with ASD may wait until they are ready to speak, rather than using delaying mechanisms or producing incompletely planned, disfluent speech. This further suggests that the interaction between questions and mazes is, at least partially, driven by an awareness of, or a desire to meet, the social obligation to respond in a timely manner following a question.

For all the children, we found that the likelihood of an utterance-initial maze was diminished during the Description activity, but the effect was attenuated for the children with ASD. The Play activity also decreased the likelihood, but was not a significant factor in the combined model. This suggests that the nature of the interaction also plays a role in the production of mazes, perhaps decreasing the pressure to respond.

5.1. Lessons for HCI

This research presents opportunities for improving SDS. First, we found that children without social impairments (i.e., those with TD and DLD) are more likely to produce a maze when responding to a question. With this in mind, SDS could be designed that are less question-response centric, thus alleviating the social pressure to respond quickly. Alternatively, SDS could anticipate an increased likelihood of disfluent speech as the gap length increases, thus providing a more natural dialogue flow and improved speech recognition results.

5.2. Implications for ASD and DLD

Although this paper focuses on how gap lengths and mazes interact and are impacted by social pressure, this work also has implications for the differential diagnosis of ASD and DLD. A primary differentiator between children with ASD and DLD is social impairment. Because gap lengths and mazing appear to be impacted by social impairments, the differences shown here could be used as additional tools to help separate the two disorders.

6. Acknowledgments

Funding gratefully received from the National Institute of Heath under grants IR21DC010239 and 5R01DC007129, and the National Science Foundation under IIS-0713698. The views herein are those of the authors and do not reflect the views of the funding agencies.

7. References


