Coherence in Child Language Narratives: A Case Study of Annotation and Automatic Prediction of Coherence

Khairun-nisa Hassanali¹, Yang Liu¹ and Thamar Solorio²

¹The University of Texas at Dallas
²University of Alabama at Birmingham
Agenda

- Introduction
- Data
- Annotation Scheme
- Automatic Prediction of Coherence
- Automatic Prediction of Language Impairment
- Conclusions
Introduction

- Child language challenges
  - Presence of disfluencies
  - Limited vocabulary

- Language Impairment
  - Typically Developing (TD) children
  - Children with Language Impairment (LI)

- Coherence
  - A characteristic of proficiency in language
Transcripts for a story telling task based on a 24 page picture storybook “Frog, Where are You?”

frog, where are you? by mercer mayer
Data Contd

- Contains 118 transcripts generated by 14 year old adolescents
- Transcripts were annotated for language impairment
  - 99 belonged to TD group
  - 19 belonged to LI group
Annotators were native English speakers
Six computer science undergraduate students
Asked to annotate the transcripts for
  ◦ Coherence
  ◦ Narrative structures
  ◦ Usage of certain narrative quality constructs
Coherence

- Measure coherence on a general scale
- No block in understanding
- Annotated on 2 scales
  - 2–level scale: 0 (coherent) or 1 (incoherent)
  - 3–level scale: 0 (coherent), 1 (somewhat coherent) or 2 (incoherent)
- Lower inter–annotator agreement for 3–level scale
- Asked to ignore spelling mistakes and disfluencies
Coherence Contd

2-Scale Coherence

<table>
<thead>
<tr>
<th>TD</th>
<th>LI</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>13</td>
</tr>
</tbody>
</table>

Coherent | Incoherent

3-Scale Coherence

<table>
<thead>
<tr>
<th>TD</th>
<th>LI</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>6</td>
</tr>
<tr>
<td>43</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>

Coherent | Somewhat coherent | Incoherent
Last week my friend Billy caught a frog. He brought it home and he put it in a jar. That night he went home and left it on the side of his bed. In the morning the frog was gone. He looked everywhere for it. He looked in his boots. He looked under his bed. He even called out the window for it. His little dog stretched so far that he fell off the windowsill. Billy climbed out of his window and picked him up. They went near a forest and started calling there, looking for him. He looked in a little hole. And his dog was looking up a tree. A gopher came out and banged him on the nose. He started looking in a tree. And his dog knocked a bees' nest that was hanging from the tree. And a swarm of wasps came out and stung him. He was looking everywhere. He looked over a rock, under a rock, everywhere you could find. He even got picked up by a strange deer and got thrown in the then he started to look in the water. He could hear some croaking sounds. He told his little dog to be quiet and looked over the log. Behind it was his frog and another frog. And there were lots of little baby frogs as well. One the frog that he had found earlier and his new family gave and so and him and the dog went home.

The story has an overall theme and very clear flow. The narrator never stutters or changes his mind mid-sentence.
Example of an Incoherent Transcript

, frog in the jar.
when he went to sleep that night the fog um the frog
the next day the f the boy opened the jar, the frog was
he quickly got dressed and climbed out the window.
the dog before he climbed out, the dog fell over.
and he had to jump out and see if the dog were all right.
he was shouting for the frog all day.
he was looking down a rabbit hole while the dog was chasing a pea
then suddenly he found a mole in the ground.
the bees suddenly started chasing the dog.
the dog the boy was looking in a everything even a
an owl came down to see what the boy was doing.
the boy fell over.
and the bees started chasing the dog again.
he was shouting on top of a um tall block for the um frog, the
he um he discovered a deer.
the deer weren't so appy about it.
suddenly the boy ran to the edge of a cliff.
he stumbled and him and the dog fell over down to the bottom of the
the dog was very scared of getting wet.
he hid behind a f um log while the dog was swimming
then suddenly he found two of them.
then he looked again.
and he found some babies.
he said goodbye and left his frog with his with his new

The child stops mid-sentence to begin another thought without finishing his sentence. This leaves several gaps in the story.
Example of a Somewhat Coherent Transcript

I was sitting in my room.
where and I'd found a frog, which so I put it in a bowl.
and me and my dog were looking at it.
then I went to bed later.
and the when I woke up the frog had gone.
so I searched my room and with the dog looking everywhere for
I looked out the window, looked on the floor outside, shouted in
I looked in a hole, looked in a beehive, looked in a tree.
  got chased by some bees from a beehive.
  looked all around a big stone, shouted over a stone and
and then I fell in a stream.
and so me and my dog went and found a big log.
so we looked behind the log.
and we found the a family of frogs.
and so I took my frog back home.

Some things are not clear such as finding the frog in the room, does not flow
well in the episode where the boys land in the pond and looks behind the log
The dominant theme is the search for the frog throughout the story. Reiteration of this theme makes the story more coherent. Search theme was annotated as follows:

- 0 – No mention of the frog missing or boy searching for the frog
- 1 – One of the following is mentioned: The frog is missing or the boy is searching for the frog
- 2 – Both of these are mentioned: The frog is missing and the boy is searching for the frog
- 3 – One additional mention of the boy searching for the frog
- 4 – Two or more additional mentions of the boy searching for the frog
Presence of all the critical components of the narrative

Annotate the story for the presence or absence of:

- Instantiation of the story
- Search episodes – narrative has 5 search episodes
- Resolution of the story
Instantiation of the story: The boy has a frog as a pet and the frog is missing
Search Episode 1: Search in the house
Search Episode 2: Search in the hole in the ground
Search Episode 3: Search in the hole in the tree
Search Episode 4: Search near the rock
Search Episode 5: Search behind the log
Resolution of the story: Find the frog and take a frog home
Narrative Quality

- Narrative quality features essential to the overall understanding of the narratives

- Annotators annotated for:
  - Use of cognitive inferences
  - Use of social engagement devices
  - Reference to affective states
  - Use of intensifiers
Use of Cognitive Inferences

- Inferences of character motivation
  - The boy turned around and pushed the dog off

- Inferences of mental state
  - He thinks the frog may be in the hole

- Inferences of causality
  - The beehive dropped to the ground causing all the bees to come out
Use of Social Engagement Devices

- Sound effects
  - “Woof!”
- Character speech
  - The boy said “Sssshhh!”
- Audience hookers
  - Look at the cute little doggie!
Reference to Affective States

- He was crying
- He was suspicious about the deer
- He was upset at the dog
Use of Intensifiers and Hedges

Use of intensifiers
- Repetition
  - The boy looked and looked for the frog
- Use of modifiers
  - The boy searched very hard for the frog

Use of hedges
- The boy probably thinks the frog is in the hole
- And my dog whimpered back, probably injured
Coh–Metrix Coherence Features

- Tool that provides an implementation of 54 coherence related features
- Known to correlate with the coherence of human written texts
- Some features include:
  - Readability metrics: Flesh–Kincaid grade level and Flesh Reading ease
  - General Word and Text Features: Basic count and frequency features such as number of words, number of utterances
  - Syntactic features: Assess syntactic complexity of text
  - Referential and Semantic features: Look at argument overlap and use LSA
Automatic Prediction of Coherence

- Binary classification: Coherent vs Incoherent
- Tried out various classifiers from WEKA
- Bayesian network classifier performed the best

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Coherent</th>
<th></th>
<th>Incoherent</th>
<th></th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F–1</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Narrative</td>
<td>0.87</td>
<td>0.84</td>
<td>0.85</td>
<td>0.59</td>
<td>0.65</td>
</tr>
<tr>
<td>Coh–Metrix</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.82</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Automatic classification of coherence on 2-scale level
Feature Analysis

- Performed feature selection using narrative structure and narrative quality features
- The following features contributed the most:
  - Presence or absence of instantiation of the story
  - Number of social engagement devices
  - Presence or absence of the resolution of the story
  - Number of cognitive inferences
  - Presence or absence of search episode 1
Automatic Prediction of Language Impairment

- Prior work uses several features such as error patterns, speech fluency features, language productivity features, vocabulary knowledge features and sentence complexity features.
- Used the coherence features for prediction of LI.

<table>
<thead>
<tr>
<th>Feature</th>
<th>P</th>
<th>R</th>
<th>F–1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabani</td>
<td>0.737</td>
<td>0.737</td>
<td>0.737</td>
</tr>
<tr>
<td>Coherence</td>
<td>0.385</td>
<td>0.263</td>
<td>0.313</td>
</tr>
<tr>
<td>Coherence + Gabani</td>
<td>0.889</td>
<td>0.842</td>
<td>0.865</td>
</tr>
</tbody>
</table>

Automatic classification of language impairment
Conclusions

- Described a scheme for annotating coherence, narrative structure and narrative quality features
- TD children produced a higher percentage of coherent narratives
- Used the narrative structure and quality features in the automatic prediction of coherence
- Features generated by the Coh–Metrix toolkit were effective in the prediction of coherence