Natural language generation

Tomer Meshorer
What is NLG?

NLG is the process of constructing natural language outputs from non-linguistic inputs. [VanLinden]

NLG is mapping some communication goal to some surface utterance that satisfies the goal. [Reiter & Dale]
This presentation describes the use of natural language generation for:

- Generating stories (“How was school today”) and adult conversation (“soccer match”) [Ehud Reiter, et al.]
- Generate written output from BLISS [Yael Netzer]
- Generate Jokes [ANNALU WALLER]

- Using NLG to help language-impaired users tell stories and participate in social dialogues
- Using Semantic Authoring for Blissymbols Communication Boards.
- Evaluating the STANDUP pun generating software with children with cerebral palsy
USING NLG TO HELP LANGUAGE-IMPAIRED USERS TELL STORIES AND PARTICIPATE IN SOCIAL DIALOGUES

E. REITER, R. TURNER, N. ALM, R. BLACK, M. DEMPSTER AND A. WALLER.

IN 12TH EUROPEAN WORKSHOP ON NATURAL LANGUAGE GENERATION (2009).
MOTIVATION

- Communication rates are extremely low, averaging 2-10 words per minute. Can be slower, based on the disability.
- Due to slow communication rates (for example, due to head mouse), the current systems are only adequate for communicating basic needs.
- Word prediction does not help.
- Want to give AAC users more control on the conversation and get closer to real human to human interaction.
- According to Melanie, the goals of communications:
  - Basic needs and wants
  - Sharing new information
  - Social closeness. 78% of human conversations (twitter)
MOTIVATION – Existing system

- Number of research system strived to support more human to human conversation.
- System provide fixed conversational moves:
  - openings and closings, such as Hello and How are you.
  - backchannel communication such as Uh-huh, Great!, and Sorry, can you repeat that
- But, fixed conversation shape, does not take into account context and goals.
- Other system worked with conversation scripts. For example Make An Appointment script. But to work in general would need a lot of scripts
- Systems were also created to support story telling vital to maintaining a full impression of one’s personality in dealing with others
- But again, they support only fixed conversations.
NLG

- Generate texts in English and other human languages from non-linguistic input
- Less rigid, NLG based MP3 player, BMW SAMMIE
  http://www.coli.uni-saarland.de/~korbay/Publications/pais06-sammie.pdf
- The companion project: http://www.companions-project.org/demonstrators/english/
- NLG from data sources.
- NLG from Bliss (Next paper)
Data to text NLG Architecture

- Data analysis – read data from sensors/web. Analyze the data and identify messages
- Editing – add annotation (“BAD” to soccer game). Rank messages
- Narration – Select messages and conversational moves
- NLG – Generate the actual utterance.
- Speech synthesis – talk the actual messages
“How was school today” system

- Goal: let the kid tell a story about what they did in school today
- Story: a series of linked real-world events which are unusual or otherwise interesting, possibly annotated with information about the child’s feelings, which can be narrated orally.
- Motivation:
  - Increase child vocabulary and language competence
  - Let the child take control of the conversation
  - Increase the child narrative (written and spoken) skills.
- Video:
  http://news.bbc.co.uk/2/hi/health/8084422.stm
Data analysis combine child schedule / domain activities to create events. It then rank events by interest.

Editing interface allow children to manage events and add opinions.
How was school today (example)

M1: I had Arts and crafts this afternoon with Mrs. Smith
M2: Rolf and ross were there

M1: I went to the Hall instead of Physiotherapy
M2: I did not like it

It was great
“How was school today” results

- Two children for four days (11 and 13)
- Qualitative evaluation: Talk with teachers/SLP/mother)
- The system worked very well
- Need more phrases. Wider range of annotations
- Stuff very supportive.
- General rollout.
- Data acquisition seems to be a problem (e.g. lunch menu)
Social conversation with adults

- Adults with cerebral palsy
- Much more general (soccer, movie, weather)
- Help to develop interpersonal relationships
- Based on the TALK system: [http://www.dundee.ac.uk/psychology/people/academics/jtodman/video23.htm](http://www.dundee.ac.uk/psychology/people/academics/jtodman/video23.htm)
- Conversation rate 10X, but need a lot of manual preparation
- Commercial version available by Dynavox.
Adult conversations

- **Data sources:**
  - Ontologies of events (watching movies, listening to music track)
  - For each event collect: when, where, who
  - Use IMDB database

- User annotate with their thought / feeling

- NLG component is based on Natural OWL (an open source java NLG engine)
Adult conversation example

- User went to concert at 8:00 PM
- NLG engine generate number of messages
Unsupervised Modeling of Twitter Conversations

- **Goal**: discover the sequential dialogue structure of conversation.
- **HMM**: Each state represents a speech act.
- **5-40 acts**: Train the HMM using EM.

NLG challenges

- NLG should adopt to the user language:
  - Instead of: I had gone swimming at 11
  - Generate: I want swimming at 11

- Keep lexical choice simple.

- Main micro planning challenges relate to discourse coherence: referring expressions and temporal descriptions.

- Want narrative that are:
  - Not just facts and events but with beginning and end
  - Want links between narrative components.
  - Interesting.

- Want to monitor what the conversation partner is doing (ASR)

- Want to let the user control the NLG system (content and expression).

- My belief: To increase conversation rate 10X, Use NLG and ASR from AAC patient aid to control an avatar. NLG / ASR / NLP are key to next generation AAC systems. (Summer project)
USING SEMANTIC AUTHORING FOR BLISSYMBOLS COMMUNICATION BOARDS

YAEL NETZER, MICHAEL ELHADAD

NAACL 2006
MOTIVATION

- **Goal:** We use NLG techniques to produce utterances automatically from the sequence of symbols, while the content determination is done by the AAC user.

- Previous techniques perform parsing a telegraphic sequence, then re-generating a full sentence in natural language. But this method lacks the main cues of morphological and syntactic structure that exist in natural language. Hence the vocabulary is small.

- Rate enhancement technique. Allow the user to select basic content and have the system expand it to full sentence.

- Support Bliss – Hebrew generation.

- Example (McCoy 1997):
  - Original: “Mary think 3 watch give John Andrew.”
  - Expanded: “Mary thinks that the 3 watches were given to John by Andrew.”

  [http://acl.ldc.upenn.edu/W/W97/W97-0503.pdf](http://acl.ldc.upenn.edu/W/W97/W97-0503.pdf)
Bliss

Pictographs
- House
- Animal
- Flag

Arbitrary symbols
- Indefinite article
- Definite Article
- That
- Action

Ideographs
- Before
- After
- In
- Out
- Up
- Down

Compound symbols
- Happy
- Sad
- School
- University
Real therapists and AAC users' conversations [McCoy 1994]

- original word order was changed by the therapist:
  - S: <boy> <table> <dusting> <grand mom> <floor> <sweep>
  - T: Boy is dusting the table and the grand mom is sweeping the floor

- missing agent
  - S: <wash> <clothes>
  - T: They are washing clothes.

- Missing verb
  - S: <toys>
  - T: They have toys.

- Add Future Tense, plural number
  - S: <girl> <make> <in> <pan> <egg> <breakfast>
  - T: Girl will make the eggs in the pan for breakfast
Semantic Authoring overview

- Each step of input insertion is controlled by a set of constraints and rules, which are drawn from an ontology.
- Generation is an incremental process and the full utterance's input for the syntactic realizer is revised with each step taken.
- If the final input is not complete, missing constituents are given default values.
- Support sentence starters.
Bliss lexicon

- Was developed from several sources:
  - The WordNet lexical database
  - English Verb classes
  - COMLEX syntax dictionary

- Database of the list of symbols accessible to the user, along with their graphic representation, semantic information

- Can be searched by keyword (learn), or by semantic/graphic component:
  - For example: searching all the words that contain both “person” and “medical” will return: aiding tool, dentist, doctor, nurse.
General NLG architecture

Legend:
- Data Structure
- Knowledge Source
- Module

Example:

[Boy: #I]-(Name) --> [Word: "Felipe"]
-(Age) --> [12]

[Boy: #Pablo]-(Brother-of) --> [Boy: #I]
- (Name) --> [Word: "Pablo"]
- (Age) --> [10]

WordNet Senses play(5 34)
Thematic Roles
Acton1[+animate]
Acton2[+animate]

Frames:
- Intransitive (+ with-PP) "Brenda met with Molly."
  Actor1 V Prep(with) Actor2
- Intransitive (plural subject) "The committee met."
  Actor1[+plural] V
- Simple Reciprocal Alternation Intransitive () "Brenda and Molly met."
  Actor1 and Actor2 V
- With Preposition Drop Alternation () "Anne met Cathy."
  Actor1 V Actor2
Verbs in same subclass
[consult, meet, play, visit]
Dynamics: Pablo and I are playing

1. The user first select participants. This create semantic description (see prev slides). The system create concept graph.
2. System show verbs the require Boy in one of their argument. Verbs are ordered by their freq.
3. The user chose “to_play” symbol. The system automatically generated: “Pablo and I are playing”
4. The word “play” belong to “meet” class of “verb” (See prev slides). This drives the generation of:
   - I played with Pablo
   - Pablo and I played
Evaluation

- Evaluated on regular users. 10 non-disabled users.
- Collected set of 19 sentences written in BLISS and their full English generation. Total number of choices is 133 while total number of words is 122. No real saving.
- Saving with dynamic communication board use on dynamic board instead of 50 boards.
- Saving in selection due to constraints.
- Generation in other languages (Hebrew) without the need for translation.
System To Augment Non-speaker’s Dialogue Using Puns

EVALUATING THE STANDUP PUN GENERATING SOFTWARE WITH CHILDREN WITH CEREBRAL PALSY

ANNA L. WALLER, ROLF BLACK, AND DAVID A. O’MARA

IN 12TH EUROPEAN WORKSHOP ON NATURAL LANGUAGE GENERATION (2009),
Motivation

- Word play is a critical part of language development in children.
- Jokes are a type of word play which offer opportunity for conversation and interaction skills.
- Enable experimentation with rhyming and ambiguity as well as learning pragmatic skills such as turn taking.
- Humor as tool
  - Used as a mean to acquire good improving their syntactic and semantic skills.
  - Pun riddles shown to improve reading comprehension.
  - Language-impaired children and adolescents have been found to have significantly poorer comprehension of humor than their peers with normal language development.
  - Important social function.
JAPE [Binsted et al, 1997]

- Designed to generate a variety of types of novel punning riddles. Q and A form.
  - A: A cereal killer.”

- Issues with JAPE:
  - Did not filter jokes according to their complexity
  - Did not support comparing words for similarity of sound
  - No interactive support (command line interface)
  - No support for “black lists”
STANDUP

- Based on JAPE but with:
  - Extensive lexicon with information about phonetic similarity of words.
  - Provision of speech output
  - Picture-word matching using two widely used AAC graphic symbol sets

System creates new jokes (not pre-stored)
Lexical requirements

- POS tags
- Phonetic spelling, for computing:
  - Time thyme
  - Pub tub
  - Bare/spank spare/bank
- Compound nouns and their components. e.g. long time, traffic jam
- Distinct senses of word phrases
- Semantic relations:
  - Strange / bizarre (synonyms)
  - Thyme herb (hypernyms)
  - Traffic car (meronyms)
STANDUP Lexicon

- **Sources:**
  - Wordnet (200K word senses, synonyms, hypernyms)
  - Unisyn (115K word forms, pronunciation dictionary)
  - SemCor (subset of brown corpus), 230K words
  - Problem: Unsuitable, Unfamiliar, American

- **Additional resources:**
  - MRC Psycholinguistic Database
  - BNC spoken Corpus (frequency ratings for compounds nouns)
  - Widgit conceptcodes: >11K concepts linked to >6K widgit rebus symbols, > 4K Mayer-Johnson PCS symbols
  - Schonell spelling list: spelling list of 3K for children 7-12. Used as preferred source of “familiar” words
STANDUP Lexicon (Data preparation)

- Disambiguation
  - Automatic using Widgit & Schonell, MRC database
  - Manual using Widgit conceptcodes and Schonell spelling list
- Define Phonetic relations (Similarity, rhymes)
- Familiarity scoring.
STANDUP database (final)

- Core Lexicon
  - 130K lexmes
    - Senses
    - POS
    - F score
  - 79K word forms
    - Orthography
    - Phonetic spelling
  - 32K compound nouns
    - Head
    - Modifier
  - 85K concepts
  - 65K hypernym pairs.
  - 10K Widgit to wordnet matches
  - >500K phonetic similarity ratings
Tell joke screen

Travel according to the progress in joke generation
Are you ready?

Travel according to the progress in joke generation

What do you call a washing machine with a September? An autumn-atic washer.
Tell question screen

What do you call a washing machine with a September?
An autumn-atic washer.
STANDUP Evaluation

1. 9 children. At least 6 years old. Normal hearing and vision. Use keyboard and scanning switch.
2. Evaluation over a period of 10 weeks with five phases. CELF and PIPA tests were administered before and after the study.
3. First two weeks consist of introductory training sessions.
4. Four weeks of intervention phase where participants were encouraged to explore the software.
5. Two weeks of evaluation phase, where participants performed various tasks, without help from the therapist.
6. After each session, participant choose their favorite joke of the session, and place it in a special joke folder.
7. Performance were estimated by analysis of video data by a researcher.
## STANDUP tasks and results

<table>
<thead>
<tr>
<th>Group</th>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A1</td>
<td>Find name (log onto the system)</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>End program (log off from the system)</td>
</tr>
<tr>
<td>B</td>
<td>B1</td>
<td>Generate any joke from new jokes</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>Speak a joke using speech synthesis</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>Save a joke to favorites</td>
</tr>
<tr>
<td></td>
<td>B4</td>
<td>Choose a joke from favorites</td>
</tr>
<tr>
<td>C</td>
<td>C1</td>
<td>Generate a joke on specified topic (e.g., about an animal)</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>Generate a joke on a specified sub topic (e.g., about a wild animal)</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>Choose a joke from old joke collection not saved to favorites.</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>Generate a joke of a particular Joke Class</td>
</tr>
<tr>
<td></td>
<td>C5</td>
<td>Generate a joke by keyword, from topics</td>
</tr>
<tr>
<td>D</td>
<td>D1</td>
<td>Generate a joke by keyword, browsing a dictionary</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>Generate a joke by keyword, typing in word</td>
</tr>
<tr>
<td>E</td>
<td>E1</td>
<td>Generate a joke appropriate to a current conversation topic.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>AB2</td>
<td>AB4</td>
<td>AB2</td>
<td>AB4</td>
<td>AB3</td>
<td>AB4</td>
<td>AB2</td>
<td>AB2</td>
<td>AB2</td>
</tr>
<tr>
<td>Inter</td>
<td>ABC3</td>
<td>ABC3</td>
<td>ABC1</td>
<td>ABC3</td>
<td>ABC3</td>
<td>ABC5</td>
<td>AB3</td>
<td>AB4</td>
<td>ABC3</td>
</tr>
<tr>
<td>Eval</td>
<td>ABCE1</td>
<td>ABC3</td>
<td>ABCE1</td>
<td>ABCD2</td>
<td>AB2</td>
<td>ABCDE1</td>
<td>ABC3</td>
<td>AB4</td>
<td>ABC1</td>
</tr>
</tbody>
</table>
## CELF and PIPA

### Table III. Pre/Poststudy Test Results

<table>
<thead>
<tr>
<th>Participant</th>
<th>CELF Word Classes Out of 27</th>
<th>PIPA Out of 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-study</td>
<td>Post-study</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>4.1</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>
## Participant feedback

<table>
<thead>
<tr>
<th>Participant Issue</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jester Character</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Ok/Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Telling jokes</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Ok</td>
<td>Ok/Good</td>
<td>Good</td>
<td>Good</td>
<td>Bad/ok</td>
</tr>
<tr>
<td>Joke quality</td>
<td>Ok</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Ok</td>
<td>Ok+</td>
<td>Good</td>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td>Interface screens</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Bad</td>
<td>Bad/ok</td>
</tr>
<tr>
<td>Scanning</td>
<td>Ok</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Touch screen</td>
<td>Bad</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Ok+</td>
<td></td>
<td>Bad/ok</td>
<td>Ok</td>
<td>Ok</td>
</tr>
<tr>
<td>Voice quality</td>
<td>Bad</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Bad</td>
<td>Bad</td>
<td>Bad</td>
<td>Ok+</td>
<td>Bad/ok</td>
</tr>
<tr>
<td>Would participant like to use STANDUP again?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Discussion

- Quality and ease of use was good. The GUI interface help the children to enjoy the software and increase the complexity of the task involved.
- The generative capabilities permitted novel explorative learning.
- Improvement in CELF score suggesting that joke generation and telling activities may impact on underlying linguistic abilities.
- Children confidence and social involvement as evidence by joke telling at home was markedly increased.
- Generation of jokes was random, which led to:
  - Older participants found the quality of jokes as bad
  - Random joke generation encourage language play and exploration.