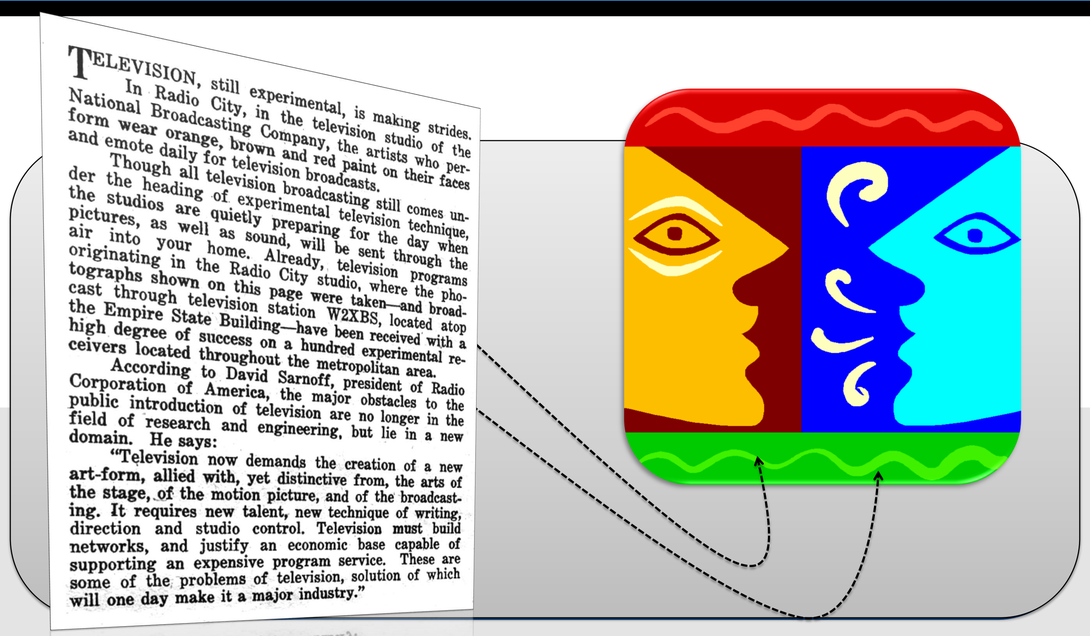
Automated Generation of Multi-Modal Dialogues from Text

Hugo Hernault (The University of Tōkyō)
Helmut Prendinger (National Institute of Informatics, Tōkyō)
Paul Piwek (The Open University, U.K.)
Mitsuru Ishizuka (The University of Tōkyō)



TELEVISION, still experimental, is making strides. In Radio City, in the television studio of the National Broadcasting Company, the artists who perform wear orange, brown and red paint on their faces and emote daily for television broadcasts.

Though all television broadcasting still comes under the heading of experimental television technique, the studios are quietly preparing for the day when pictures, as well as sound, will be sent through the air into your home. Already, television programs originating in the Radio City studio, where the photographs shown on this page were taken—and broadcast through television station W2XBS, located atop the Empire State Building—have been received with a high degree of success on a hundred experimental receivers located throughout the metropolitan area.

According to David Sarnoff, president of Radio Corporation of America, the major obstacles to the public introduction of television are no longer in the field of research and engineering, but lie in a new domain. He says:

"Television now demands the creation of a new art-form, allied with, yet distinctive from, the arts of the stage, of the motion picture, and of the broadcasting. It requires new talent, new technique of writing, direction and studio control. Television must build networks, and justify an economic base capable of supporting an expensive program service. These are some of the problems of television, solution of which will one day make it a major industry."

Goals

- Enable anyone to create multi-modal presentations easily from any text. Use virtual agents in Second Life to "act-out" the dialogues.
- Create an intuitive and engaging way to convey information.
- Employ this system in the domains of e-Healthcare, e-Learning and e-News.

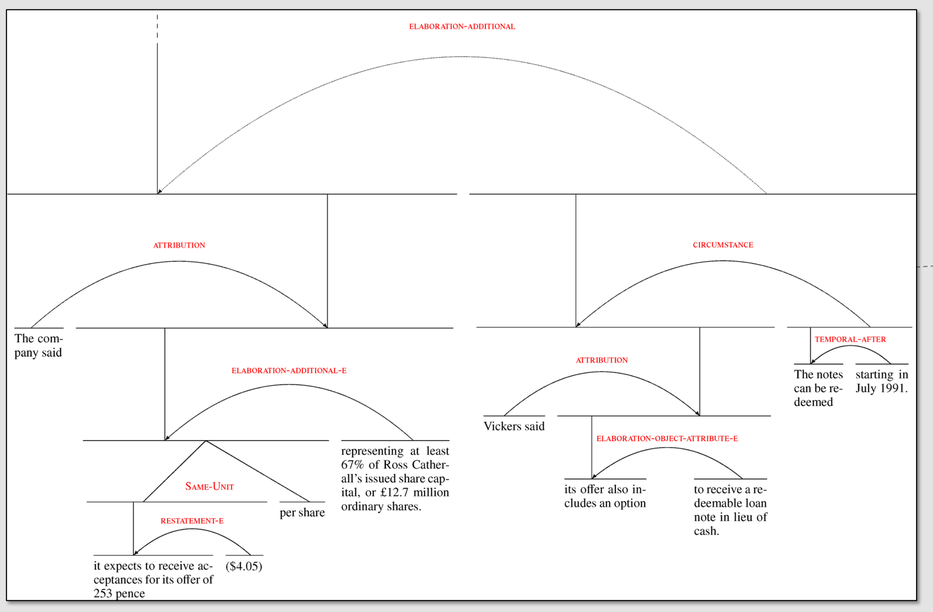
Method

ATTRIBUTION(*P,Q*) ⇒
Layman > *What did + getSubject(P+Q) + getMainVerbLemma(P+Q)?*
Expert > *AddIfNotPresentIn(Q,That) + Q*

Alternate-Speaker Rule:
ATTRIBUTION(*P,Q*) ⇒
Expert > *removeIfPresentIn(Q, That) + Q*
Layman > *Who getMainVerbFromSentence(P+Q) + that?*
Expert > *getSubjectFromSentence(P+Q) + generateWordForm(do, getMainVerbMorphoTagsFromSentence(P+Q)*

Mapping rules for the ATTRIBUTION relation

```
<Data>
    […]
    <Utterance Speaker="Layman" Listener="Expert">
        What did the company say?
    </Utterance>
    <Utterance Speaker="Expert" Listener="Layman">
        That it expects to receive acceptances for its offer of 253 pence  per share representing at least 67%
        of Ross Catherall's issued share capital, or 12.7 million ordinary shares.
    </Utterance>
    <Utterance Speaker="Layman" Listener="Expert">
        Anything else?
    </Utterance>
    <Utterance Speaker="Expert" Listener="Layman">
        Its offer also includes an option to receive a redeemable loan note in lieu of cash.
    </Utterance>
    <Utterance Speaker="Layman" Listener="Expert">
        Who said that?
    </Utterance>
    <Utterance Speaker="Expert" Listener="Layman">
        Vickers did.
    </Utterance>
    <Utterance Speaker="Layman" Listener="Expert">
        Under what circumstances?
    </Utterance>
    <Utterance Speaker="Expert" Listener="Layman">
        The notes can be redeemed starting in July 1991.
    </Utterance>
    […]
</Data>
```



Excerpt from the RST Tree representation of the input text.

"Vickers PLC, a British aerospace, defense and automotive conglomerate, said it reached an agreed cash bid of 108.2 million pounds for Ross Catherall Group PLC, a maker of specialty alloy and ceramics. The company said it expects to receive acceptances for its offer of 253 pence per share representing at least 67% of Ross Catherall's issued share capital, or 12.7 million ordinary shares. Vickers said its offer also includes an option to receive a redeemable loan note in lieu of cash.The notes can be redeemed starting in July 1991. The company said its acquisition of Ross Catherall will be covered largely by cash raised in its July disposal of Howson-Algraphy for 241.7 million pounds."

Input text: Financial newspaper article (wsj_1173 from the RST Discourse Treebank).

1. Analysis of the functional relations of the text using Rhetorical Structure Theory (RST).

2. After filtering-out irrelevant and unimplemented RST relations based on their priority, a *DialogueNet* structure is created by mapping the tree's relations to query-answer pairs.

3. Using MPML3D, the *DialogueNet* is transformed into a multimodal dialogue and "acted-out" by virtual character agents in Second Life.

Percentage of RST relations mapped	31.7%
Minimum number of turns per dialogue	1
Maximum number of turns per dialogue	215
Mean number of turns per dialogue	46.9
Median number or turns per dialogue	32
Standard deviation of the number of turns per dialogue	43.3
Total number of turns for Layman	5649
Shortest turn length for Layman (words)	2
Longest turn length for Layman (words)	38
Mean turn length for Layman (words)	2.7
Median turn length for Layman (words)	2
Standard deviation of the turn length for Layman (words)	1.4
Total number of turns for Expert	10655
Shortest turn length for Expert (words)	1
Longest turn length for Expert (words)	1177
Mean turn length for Expert (words)	18.8
Median turn length for Expert (words)	12
Standard deviation of the turn length for Expert (words)	38.7
Global mean turn length (words)	13.2
Global median turn length (words)	5
Global standard deviation of the turn length (words)	32.2

Characteristics of our dialogues generated from the RST Discourse Treebank, with 5 implemented mapping rules.

NII 連絡先：姓 名（Helmut PRENDINGER）／ 国立情報学研究所 コンテンツ科学研究系 准教授
TEL：03-4212-2650 FAX：03-4212-2650 Email：helmut@nii.ac.jp