Automated Parsing of Dialogue Games in Open Ended Dialogue

By: Matthew Marchesani

Adviser: Prof. Nick Webb

Background

In Computer Science, there is a field called Natural Language Processing (NLP). Dialogue systems are a research area in NLP. Dialogue systems can be used to the benefit people in many ways.

My Research Question

I wanted to find out if patterns in dialogues could be represented abstractly. This similar to the way sentences can be parsed into parts of speech.

Definition of Terminology

- Dialogue
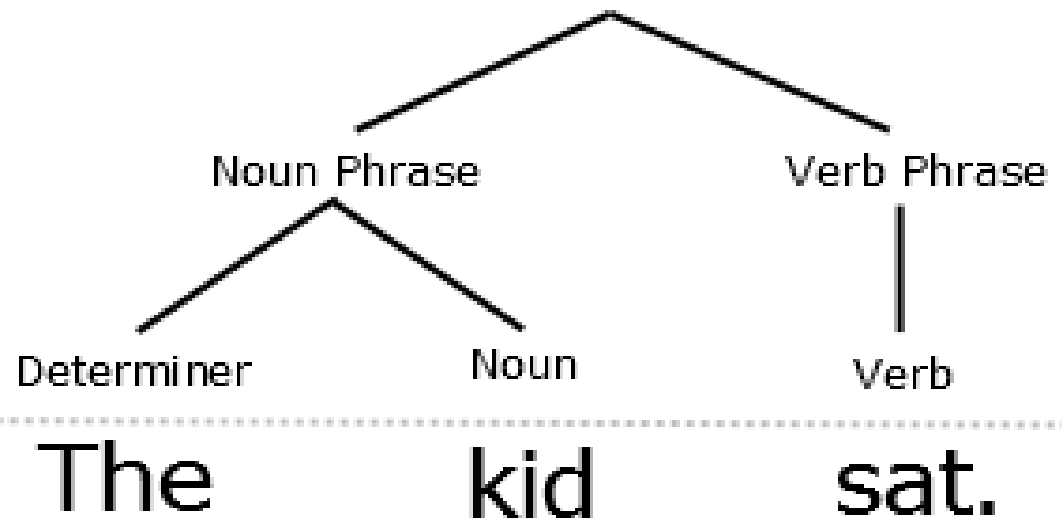- Turn
- Utterance
- Dialogue act

Dialogue Act Example

Dialogue Act: Statement (non-opinion)

....

Grammatical Unit: Sentence

....

Part of Speech:

```
                                        Sentence
                              ╱                    ╲
                      Noun Phrase              Verb Phrase
                     ╱          ╲                   │
              Determiner       Noun               Verb
```

The   kid   sat.

Dialogue Games

Greeting Dialogue Game
- 1: Greeting   "Hello."
- 2: Greeting   "Hi."

End

Question Dialogue Game
- 1: Question   "How are you?"
- 2: Answer   "Awesome."

End

Question Dialogue Game
- 1: Question   "What is your address?"

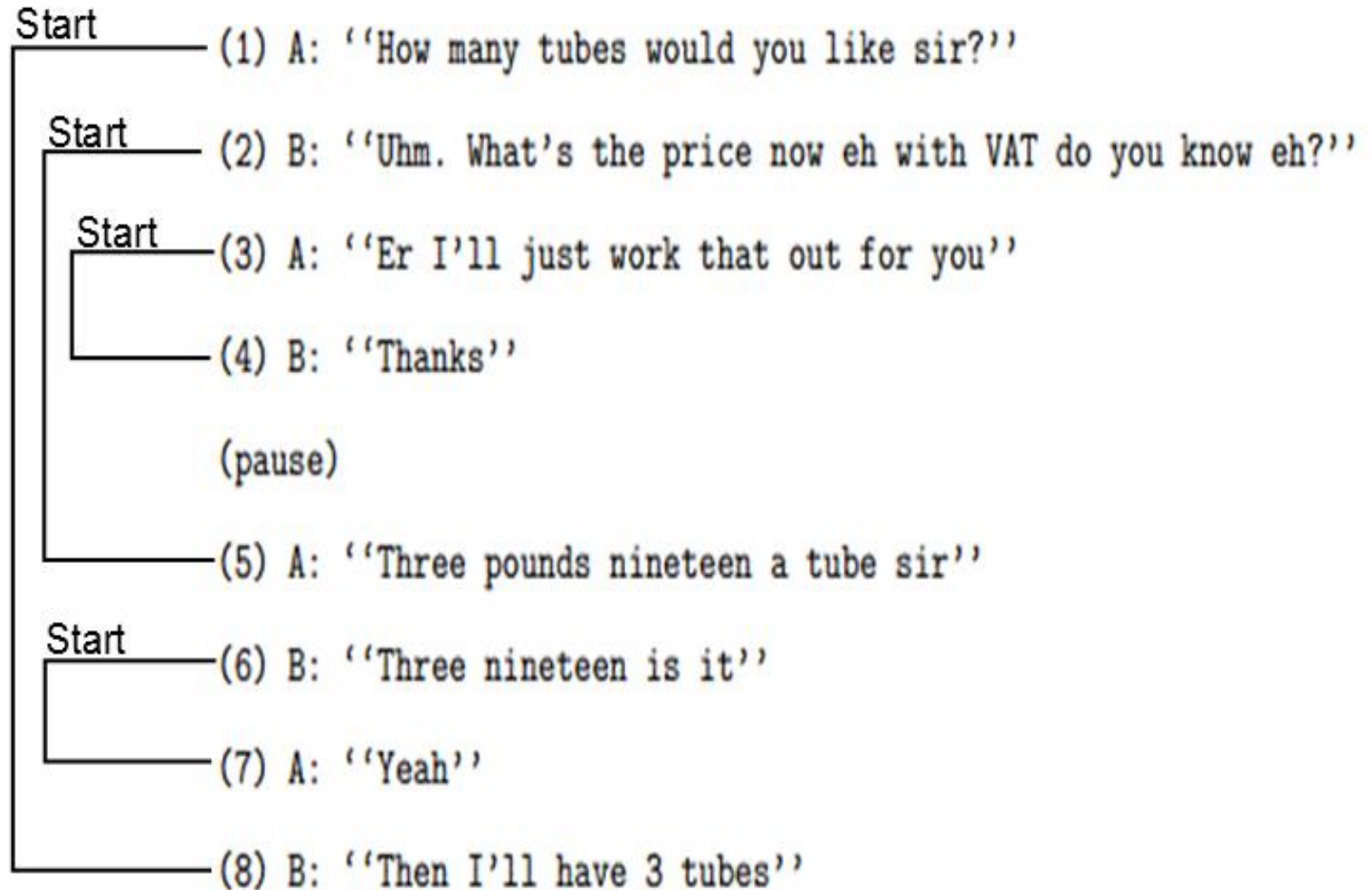   **Question Dialogue Game**
- 2: Question   "Why do you want to know."
- 1: Answer   "I have a package for you."

    End

- 2: Answer   "807 Union Street."

End

DG Example

Start ——— (1) A: ''How many tubes would you like sir?''

Start ——— (2) B: ''Uhm. What's the price now eh with VAT do you know eh?''

Start ——— (3) A: ''Er I'll just work that out for you''

——— (4) B: ''Thanks''

(pause)

——— (5) A: ''Three pounds nineteen a tube sir''

Start ——— (6) B: ''Three nineteen is it''

——— (7) A: ''Yeah''

——— (8) B: ''Then I'll have 3 tubes''

DG Example

Q-begin
*f Do you want it to go below the carpenter?
QUERY-YN

*g < No, I want you to go up the left hand side of it towards /
REPLY-N

*f Ok.
ACKNOWLEDGE

*g Green Bay and make it a slightly diagonal line, towards,
em sloping to the right.>
Q-end REPLY-N–continued

C-begin ***f** So you want me to go above the carpenter?
CHECK

*g Uh-huh.
C-end REPLY-Y

Switch Board Corpus
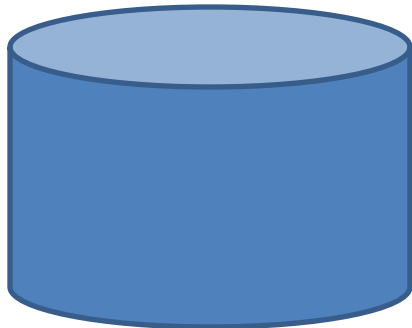
Dialogues: 1155

Turns: 105879

Utterances: 204915

Average turns per dialogue: 91

Highest # turns in a dialogue: 263

Least # turns in a dialogue: 14

Mode: 67 (occurred 23 times)
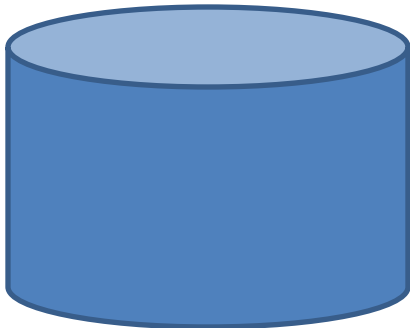
Switch Board Corpus

Highest # Utterances in a turn: 30

Lowest # Utterances in a turn: 1

Mode: 1 (occurred 56929 times)

DA Acts

Statement-non-opinion	"Me, I'm in the legal department."	72,824	36%
Backchannel	"Uh-huh."	37,096	19%
Statement-opinion	"I think it's great"	25,197	13%
Agree/Accept	"That's exactly it"	10,820	5%
Abandoned or Turn-Exit	"So, -"	10,569	5%
Appreciation	"I can imagine."	4,633	2%
Yes-No-Question	"Do you have to have any special training"	4,624	2%
Non-verbal	"[Laughter]"	3,548	2%
Yes answers	"Yes."	2,934	1%
Conventional-closing	"Well, it's been nice talking to you."	2,486	1%
Uninterpretable	"But, uh, yeah"	2,158	1%
Wh-Question	"Well, how old are you?"	1,911	1%
No answers	"No."	1,340	1%
Acknowledgement	"Oh, okay."	1,277	1%
Hedge	"I don't know if I'm making sense or not."	1,182	1%
Declarative Question	"So you can afford to get a house?"	1,174	1%

Strategy for Processing the Corpus

- Simplify the corpus
  - Remove DAs that do not change the flow of the dialogue.
  - Combine DAs by the same speaker if they are adjacent and have the same type.

Strategy for Processing the Corpus

- Simple parsing of Question Answer DGs
  - A question at a turn change is answered by all the following utterances of the speaker answering the question.

- Sophisticated parsing of Question Answer DGs
  - A question at a turn change is answered by utterances with the same topic.

Utterances Removed

- Acknowledge-(backchannel)
- Backchannel-in-question-form
- Non-verbal
- Other
- Quotation
- Uniterpretable

Utterances Conflated

1: yes-no-question, or-clause

2: no-answers[3], statement-non-opinion

1: statement-opinion, yes-no-question[2]

2: yes-answers, statement-non-opinion[2]

1: statement-non-opinion, statement-non-opinion

2: agree/accept[2], summarize/reformulate

1: appreciation, statement-opinion

2: agree/accept[2], statement-opinion

Removing Utterances

Speaker A:

DA="statement-non-opinion": You know, in some of the European countries, they have numbers of parties that run.

DA="statement-non-opinion": And then lot of more voters come out to vote than vote in this country.

Speaker B:

DA="acknowledge-(backchannel)": Uh-huh.

DA="acknowledge-(backchannel)": Right.

DA="uniterpretable": And sometimes it's, uh.

DA="statement-opinion": Well, not sometimes, but many times it's who has the most money to spend on the largest campaign.

DA="statement-opinion": And, and issues get, uh, bogged down.

Conflating Utterances

Speaker A:

DA="statement-non-opinion": You know, in some of the European countries, they have numbers of parties that run.

DA="statement-non-opinion": And then lot of more voters come out to vote than vote in this country.

Speaker B:

DA="statement-opinion": Well, not sometimes, but many times it's who has the most money to spend on the largest campaign.

DA="statement-opinion": And, and issues get, uh, bogged down.

Result

Speaker A:

DA="statement-non-opinion": You know, in some of the European countries, they have numbers of parties that run. And then lot of more voters come out to vote than vote in this country.

Speaker B:

DA="statement-opinion": Well, not sometimes, but many times it's who has the most money to spend on the largest campaign. And, and issues get, uh, bogged down.

Change of Topic

Speaker A:

DA = wh-question, nouns = night:

    And what night is it on?

Speaker B:

DA = hold-before-answer/agreement, nouns=[]

    Well, we watched it,

DA = statement, nouns = times excuse friday night

    The times we've seen it, it's been like on Satur-, or, excuse me, Friday night.

DA = hedge, nouns = couple times

    We've only seen it a couple of times.

Parsing Statistics

	No Topic	Topic
DGs parsed:	7262	6479
Questions parsed:	80.79%	72.01%
Total utterances parsed:	26.94%	20.24%
DGs immediately followed by DGs:	26.94%	15.41%
Non-parsed question followed by DG:	25.06%	19.62%

Next Step

- Divide dialogues into sections based on topic.

- Conflate based on topic.

- Write a parser for nested DGs.

- Once 90% of a dialogue is parsed, make a DG representation in XML using DGs.

- Continue gathering statistical information about DGs.

Questions