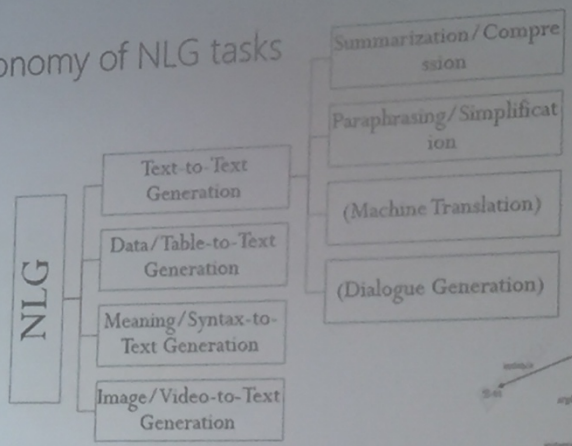
Automatic Text Generation: Recent Advances and Challenges

Xiaojun Wan

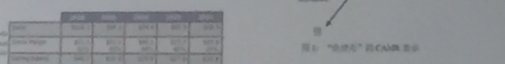
Peking University
http://www.icst.pku.edu.cn/lcwm/wanxj/

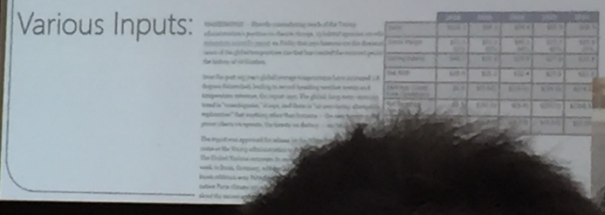
Aug. 30, 2018

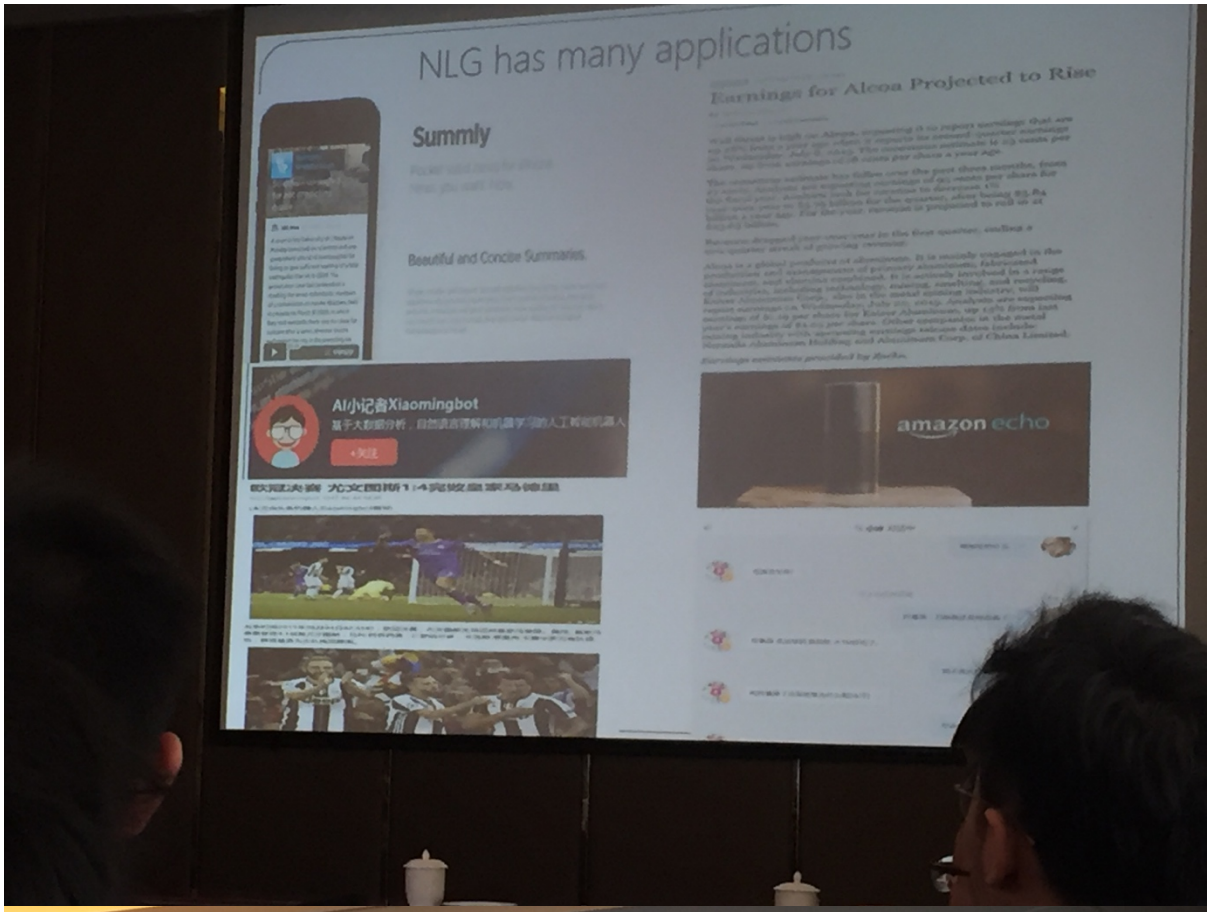
Taxonomy of NLG tasks

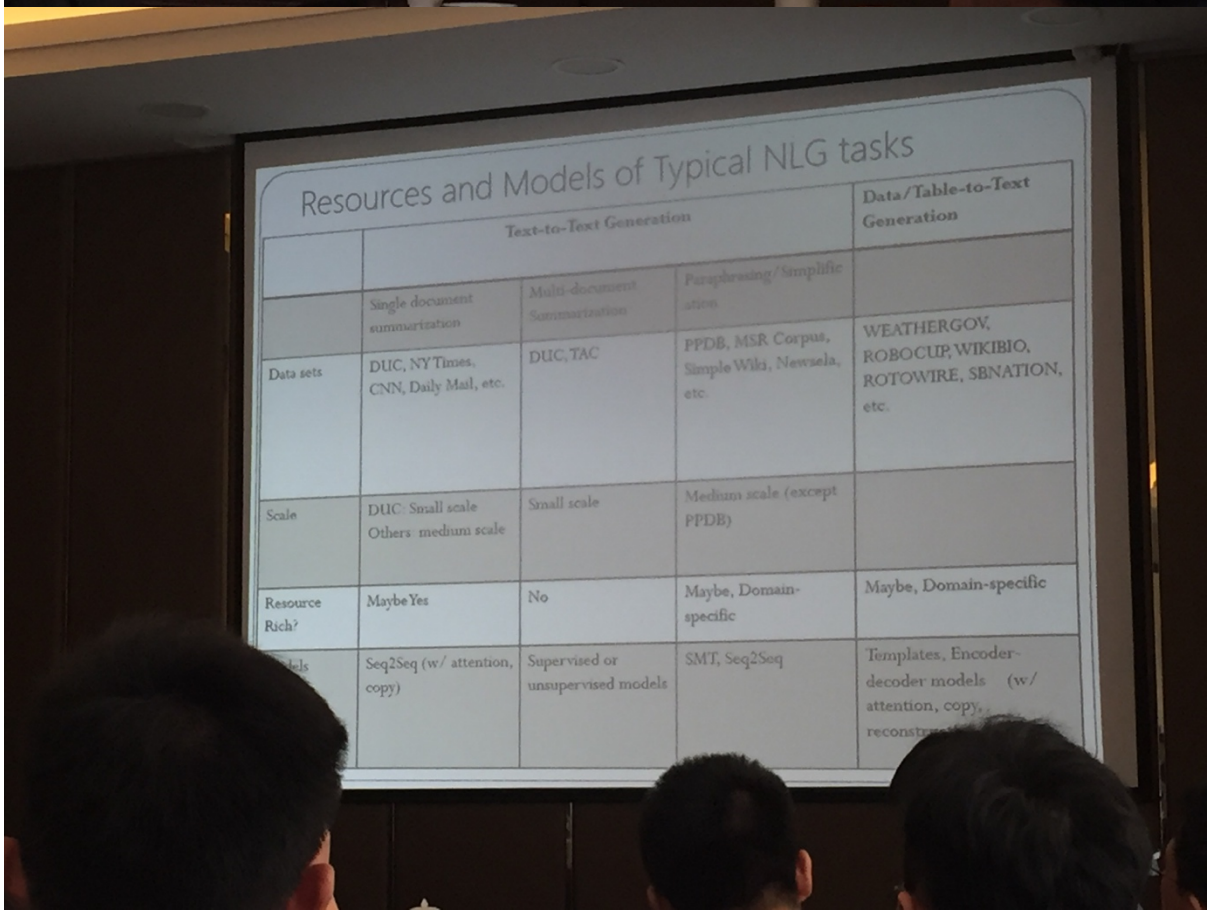


Various Inputs:

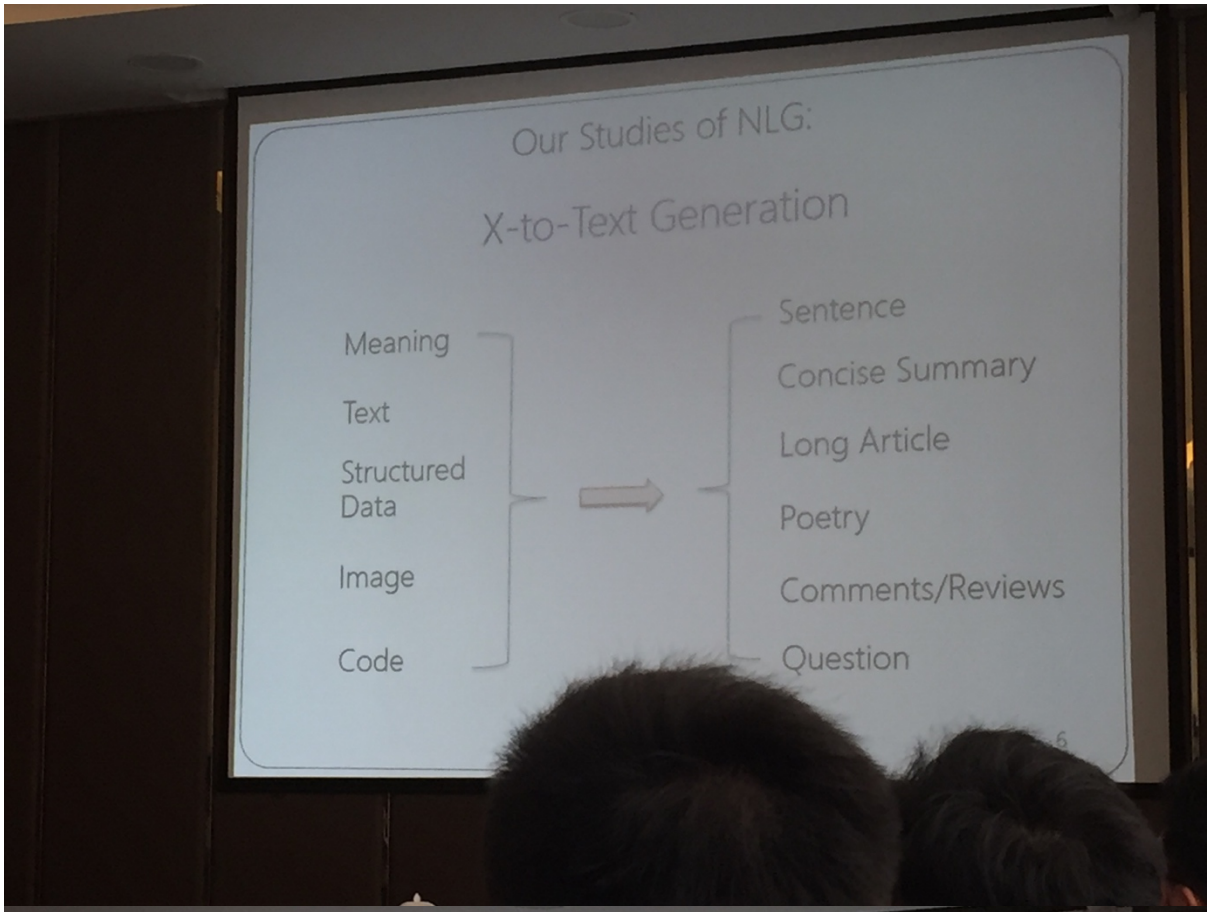


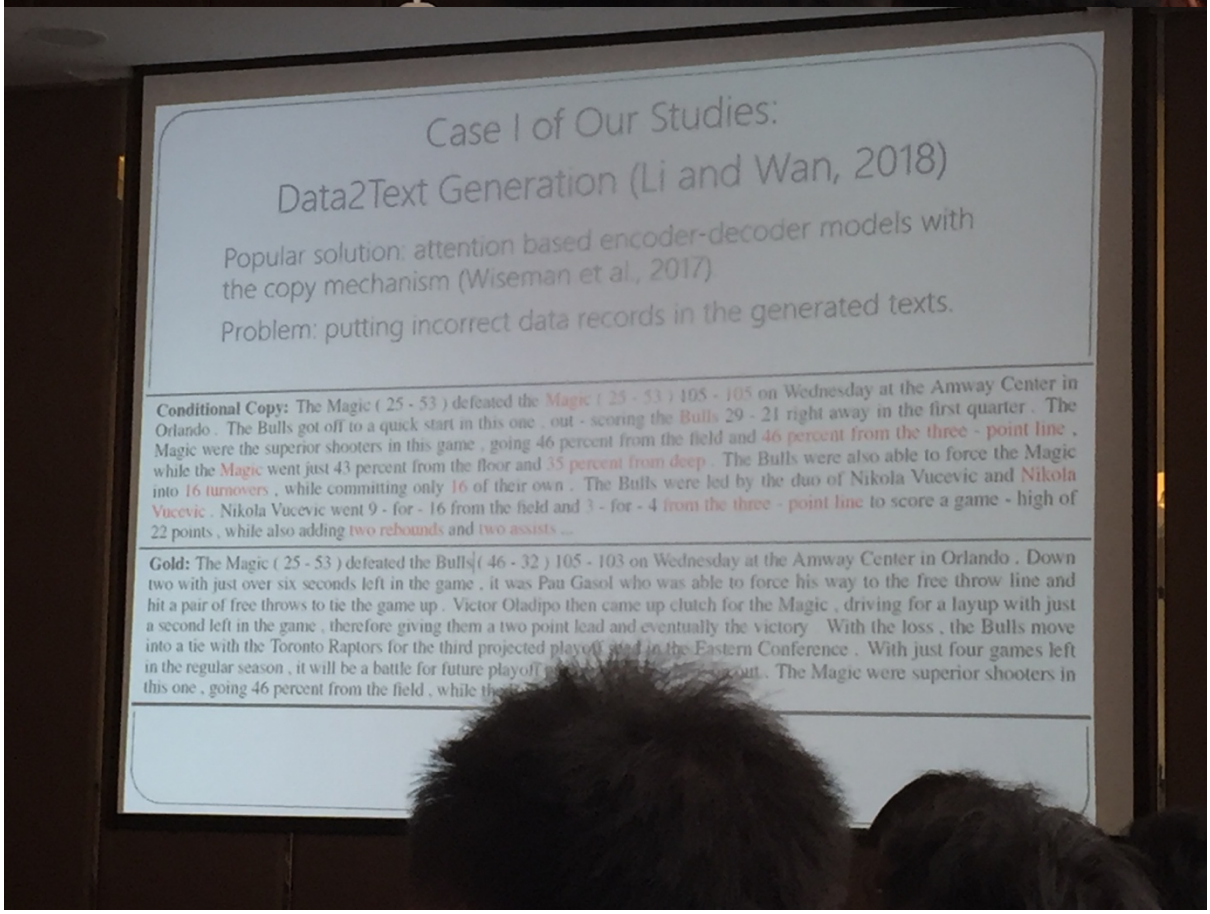


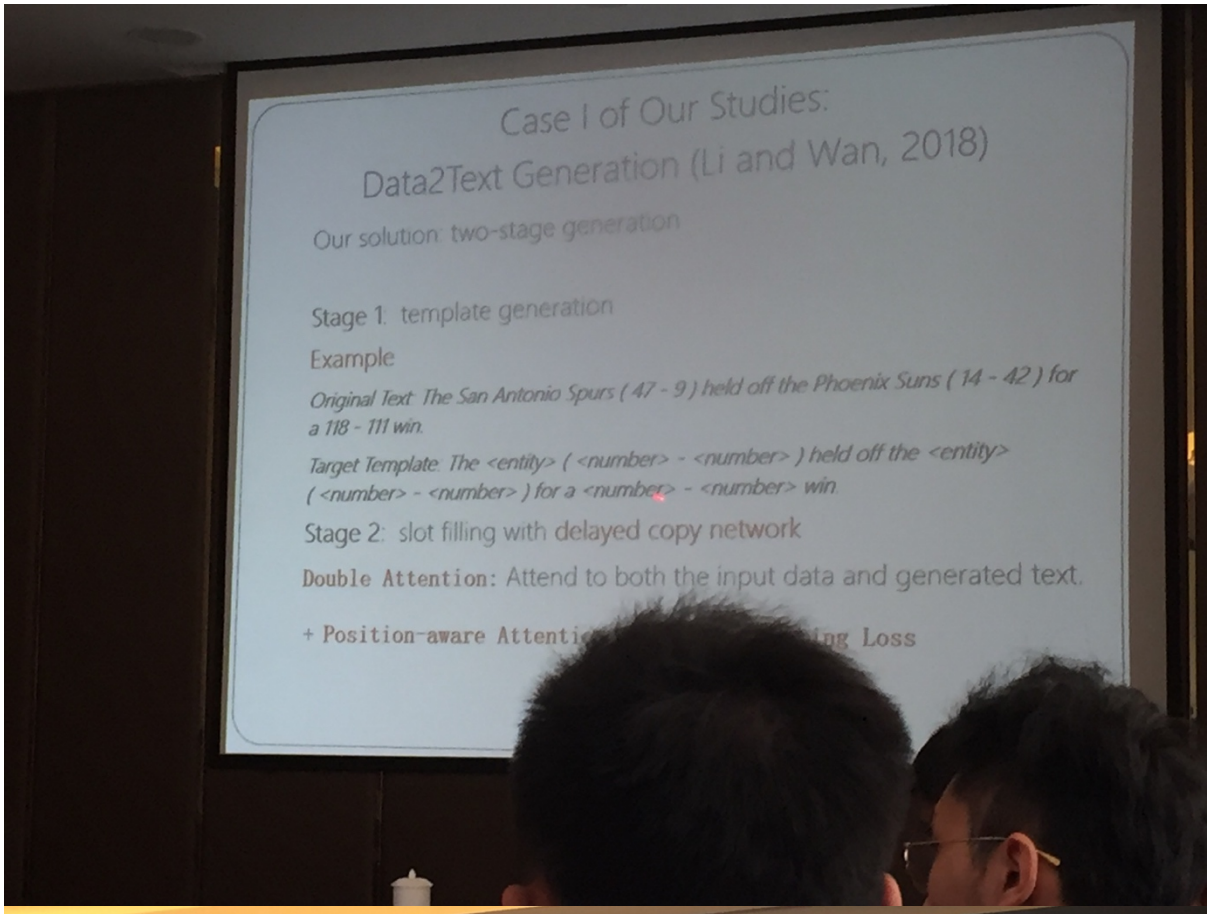


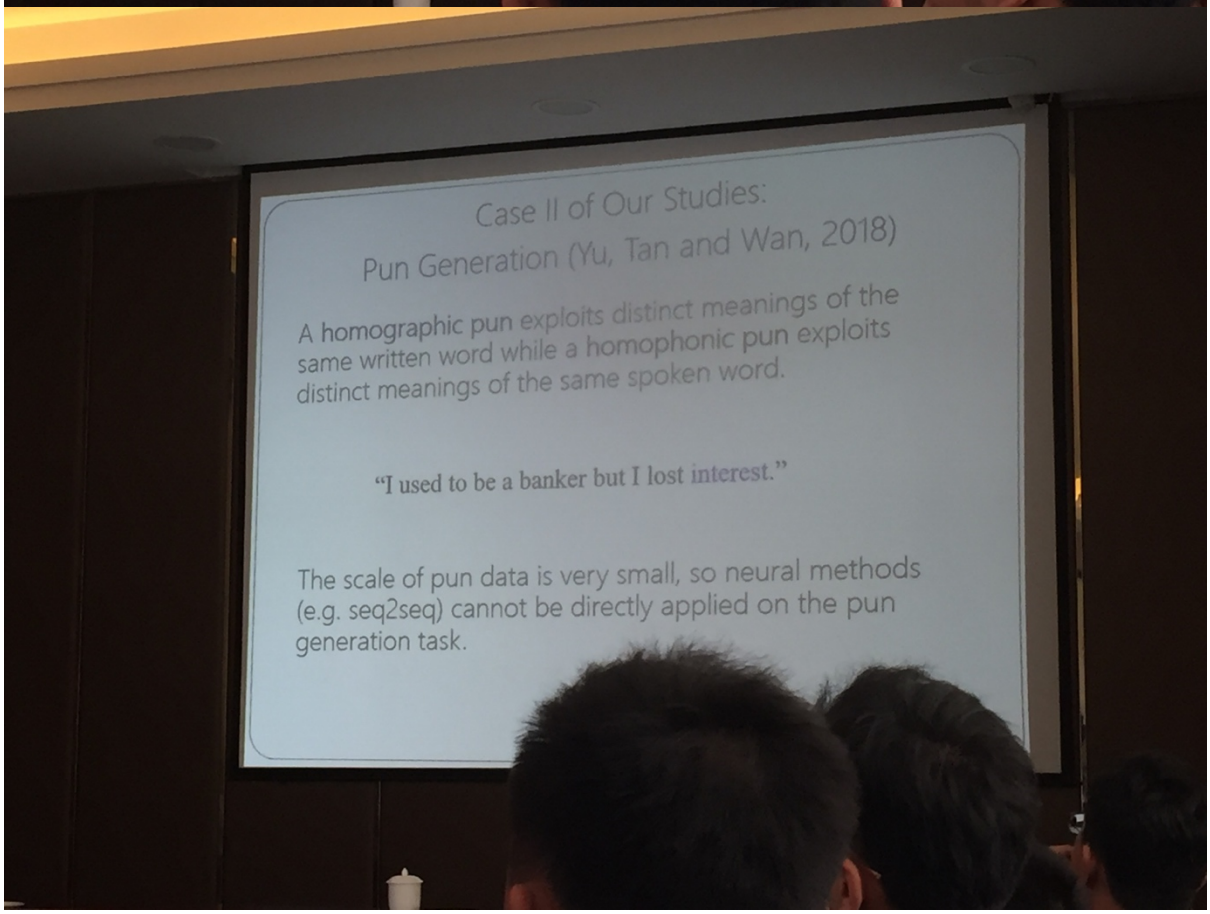


	Text-to-Text Generation			Data/Table-to-Text Generation
	Single document summarization	Multi-document Summarization	Paraphrasing/Simplification	
Data sets	DUC, NY Times, CNN, Daily Mail, etc.	DUC, TAC	PPDB, MSR Corpus, Simple Wiki, Newsela, etc.	WEATHERGOV, ROBOCUP, WIKIBIO, ROTOWIRE, SBNATION, etc.
Scale	DUC: Small scale Others: medium scale	Small scale	Medium scale (except PPDB)	
Resource Rich?	Maybe Yes	No	Maybe, Domain-specific	Maybe, Domain-specific
els	Seq2Seq (w/ attention, copy)	Supervised or unsupervised models	SMT, Seq2Seq	Templates, Encoder-decoder models (w/ attention, copy, reconstr


Our Studies of NLG:
X-to-Text Generation
Meaning
Text
Structured Data
Image
Code
Sentence
Concise Summary
Long Article
Poetry
Comments/Reviews
Question

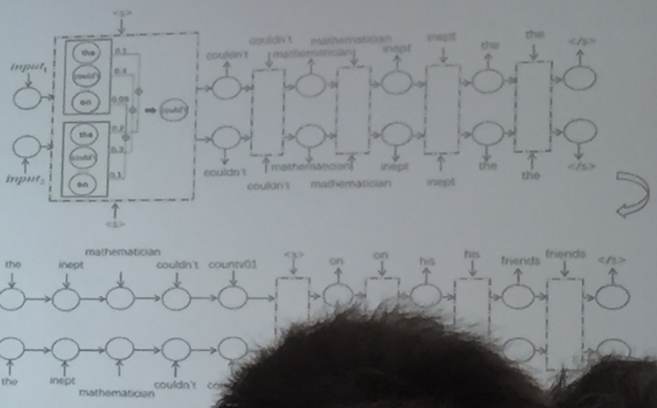

Case I of Our Studies:
Data2Text Generation (Li and Wan, 2018)
Popular solution: attention based encoder-decoder models with the copy mechanism (Wiseman et al., 2017)
Problem: putting incorrect data records in the generated texts.
Conditional Copy: The Magic (25 - 53) defeated the Magic (25 - 53) 105 - 105 on Wednesday at the Amway Center in Orlando . The Bulls got off to a quick start in this one , out - scoring the Bulls 29 - 21 right away in the first quarter . The Magic were the superior shooters in this game , going 46 percent from the field and 46 percent from the three - point line , while the Magic went just 43 percent from the floor and 35 percent from deep . The Bulls were also able to force the Magic into 16 turnovers , while committing only 16 of their own . The Bulls were led by the duo of Nikola Vucevic and Nikola Vucevic . Nikola Vucevic went 9 - for - 16 from the field and 3 - for - 4 from the three - point line to score a game - high of 22 points , while also adding two rebounds and two assists ...
Gold: The Magic (25 - 53) defeated the Bulls (46 - 32) 105 - 103 on Wednesday at the Amway Center in Orlando . Down two with just over six seconds left in the game , it was Pau Gasol who was able to force his way to the free throw line and hit a pair of free throws to tie the game up . Victor Oladipo then came up clutch for the Magic , driving for a layup with just a second left in the game , therefore giving them a two point lead and eventually the victory . With the loss , the Bulls move into a tie with the Toronto Raptors for the third projected playoff spot in the Eastern Conference . With just four games left in the regular season , it will be a battle for future playoff
out . The Magic were superior shooters in this one , going 46 percent from the field , while the


Case I of Our Studies:
Data2Text Generation (Li and Wan, 2018)
Our solution: two-stage generation
Stage 1: template generation
Example
Original Text: The San Antonio Spurs (47 - 9) held off the Phoenix Suns (14 - 42) for a 118 - 111 win.
Target Template: The <entity> (<number> - <number>) held off the <entity> (<number> - <number>) for a <number> - <number> win.
Stage 2: slot filling with delayed copy network
Double Attention: Attend to both the input data and generated text.


Case II of Our Studies:
Pun Generation (Yu, Tan and Wan, 2018)
A homographic pun exploits distinct meanings of the same written word while a homophonic pun exploits distinct meanings of the same spoken word.
"I used to be a banker but I lost interest."
The scale of pun data is very small, so neural methods (e.g. seq2seq) cannot be directly applied on the pun generation task.

Case II of Our Studies:
Pun Generation (Yu, Tan and Wan, 2018)

Our proposed neural method:
Conditional language model + Joint Beam Search + Associated words



Case III of Our Studies:
Images2Poem Generation (Liu, Wan&Guo, 2018)

Poetry Generation: A challenging text generation task

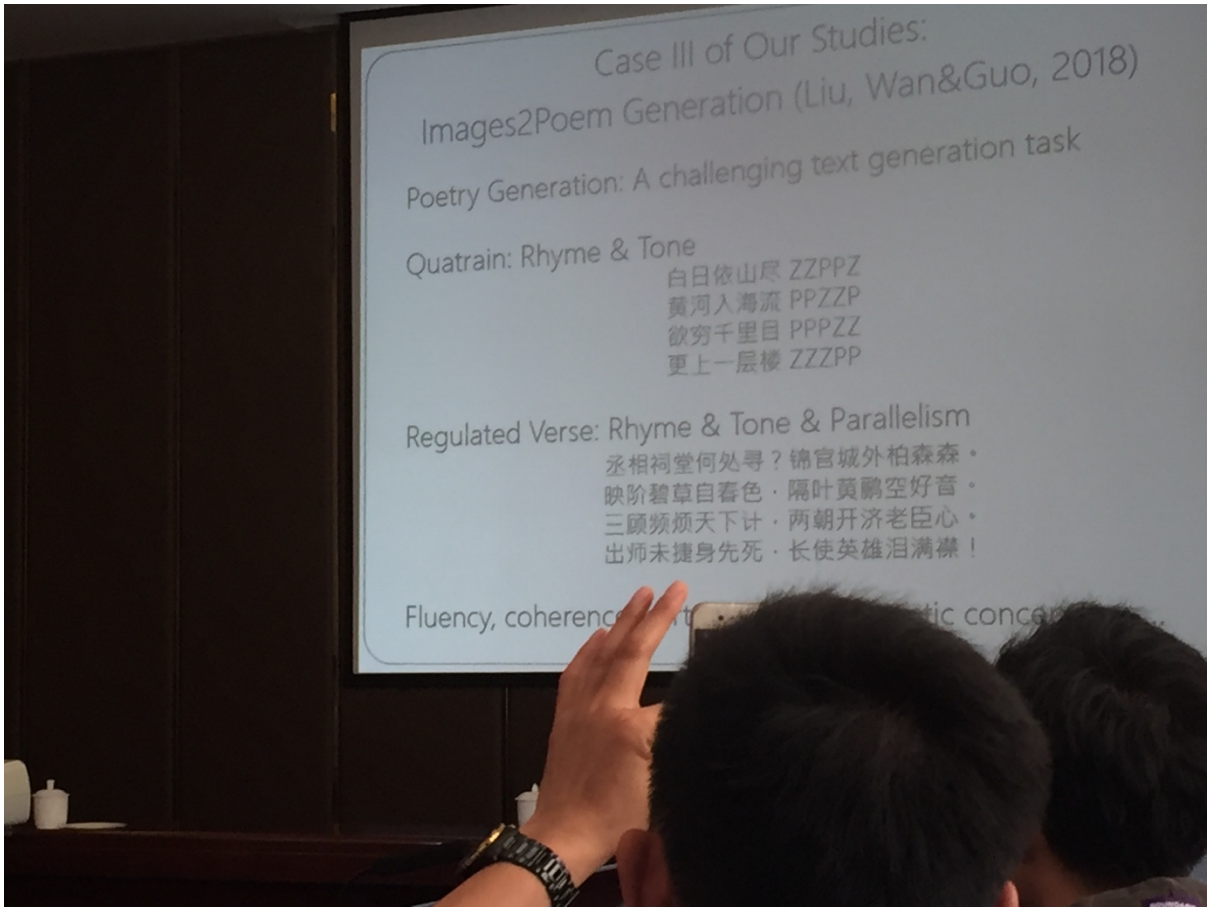
Quatrain: Rhyme & Tone

白日依山尽 ZZPPZ
黄河入海流 PPZZP
欲穷千里目 PPPZZ
更上一层楼 ZZZPP

Regulated Verse: Rhyme & Tone & Parallelism

丞相祠堂何处寻？锦官城外柏森森。
映阶碧草自春色，隔叶黄鹂空好音。
三顾频烦天下计，两朝开济老臣心。
出师未捷身先死，长使英雄泪满襟！

Fluency, coherenc[illegible] ic conce[illegible]


Case III of Our Studies:
Images2Poem Generation (Liu, Wan&Guo, 2018)
Poetry Generation: A challenging text generation task
Quatrain: Rhyme & Tone
白日依山尽 ZZPPZ
黄河入海流 PPZZP
欲穷千里目 PPPZZ
更上一层楼 ZZZPP
Regulated Verse: Rhyme & Tone & Parallelism
丞相祠堂何处寻？锦官城外柏森森·
映阶碧草自春色·隔叶黄鹂空好音·
三顾频烦天下计·两朝开济老臣心·
出师未捷身先死·长使英雄泪满襟！