



THE OHIO STATE UNIVERSITY

Preliminary Study of Automated Analysis of Event Reports from Nuclear Power Plants Based on Natural Language Processing Techniques

Yunfei Zhao (Co-authors: Xiaoxu Diao, Carol Smidts)

Department of Mechanical and Aerospace Engineering

The Ohio State University

Sep. 19, 2018

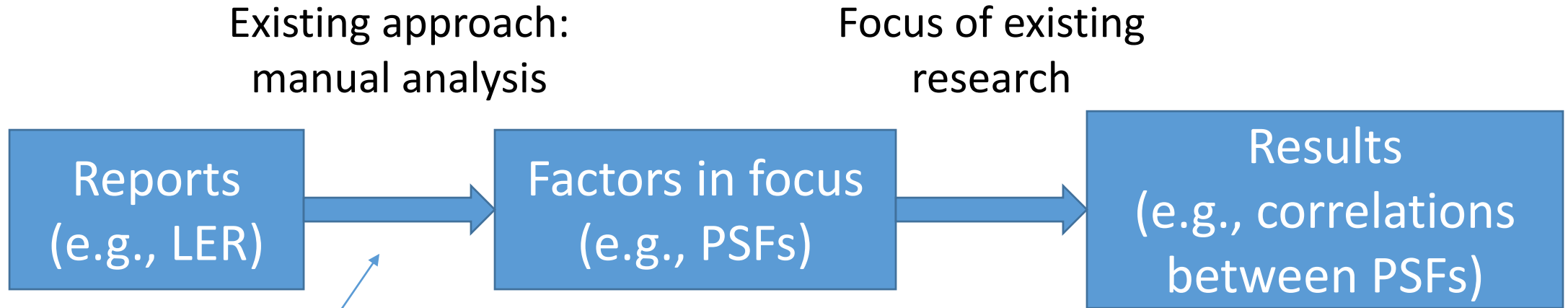


Background

- A large number of text-based event reports available:
 - US NRC Licensee Event Reports
 - INPO Operating Experience database, etc.
- What is in the reports:
 - Root causes
 - The effects of these causes and the following effects of these effects ...
 - Basically, how the event was initiated and progressed
- This information is particularly useful to us:
 - We can analyze the important factors contributing to the events
 - Can also analyze how these factors are related



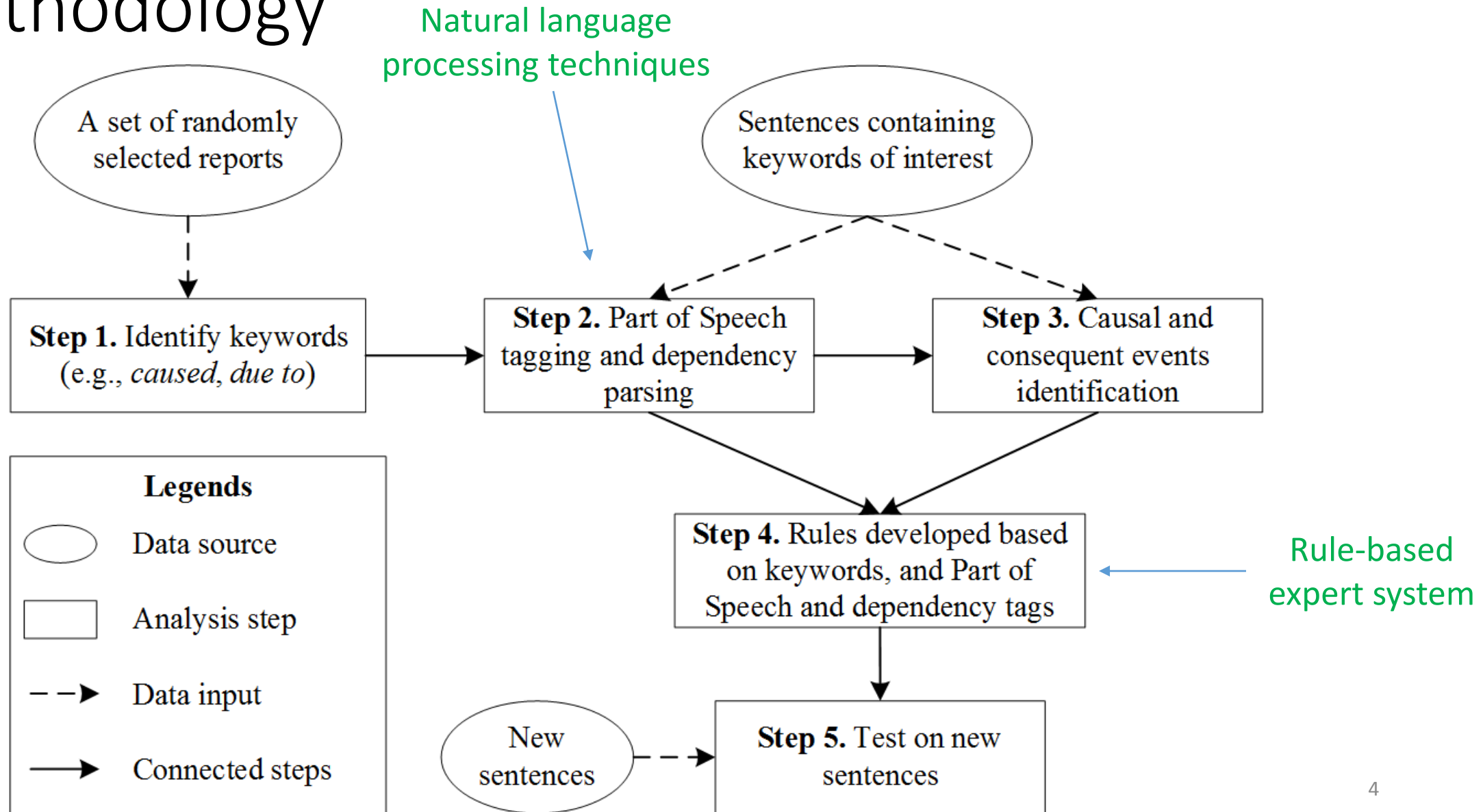
Background (Continued)



- **Our goal:**
 - **Automate** the analysis process.
 - Specifically, identify **causal relationships** in text-based LER reports **automatically**.



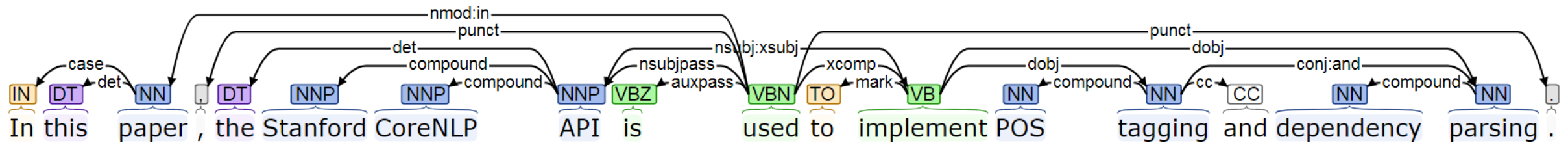
Methodology





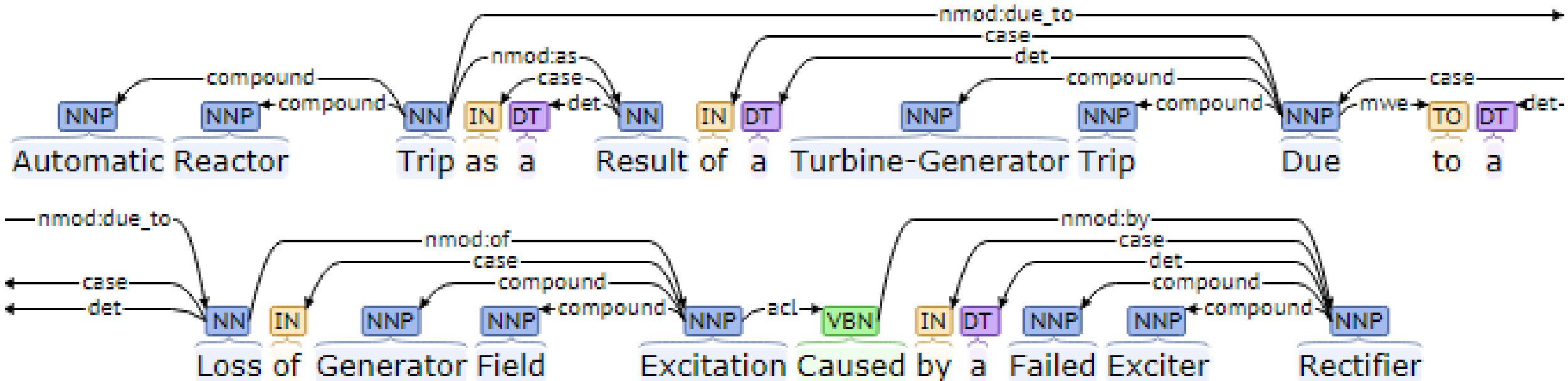
Methodology (Continued)

- Part of speech: word classes or categories, e.g., noun or verb
- Dependency relation: the semantic relations between words in a sentence
 - example: *She looks very beautiful. She is the nominal subject (nsubj) of looks.*
- Two commonly used features in NLP for more complicated tasks.
- Implemented by Stanford CoreNLP API.



Methodology (Continued)

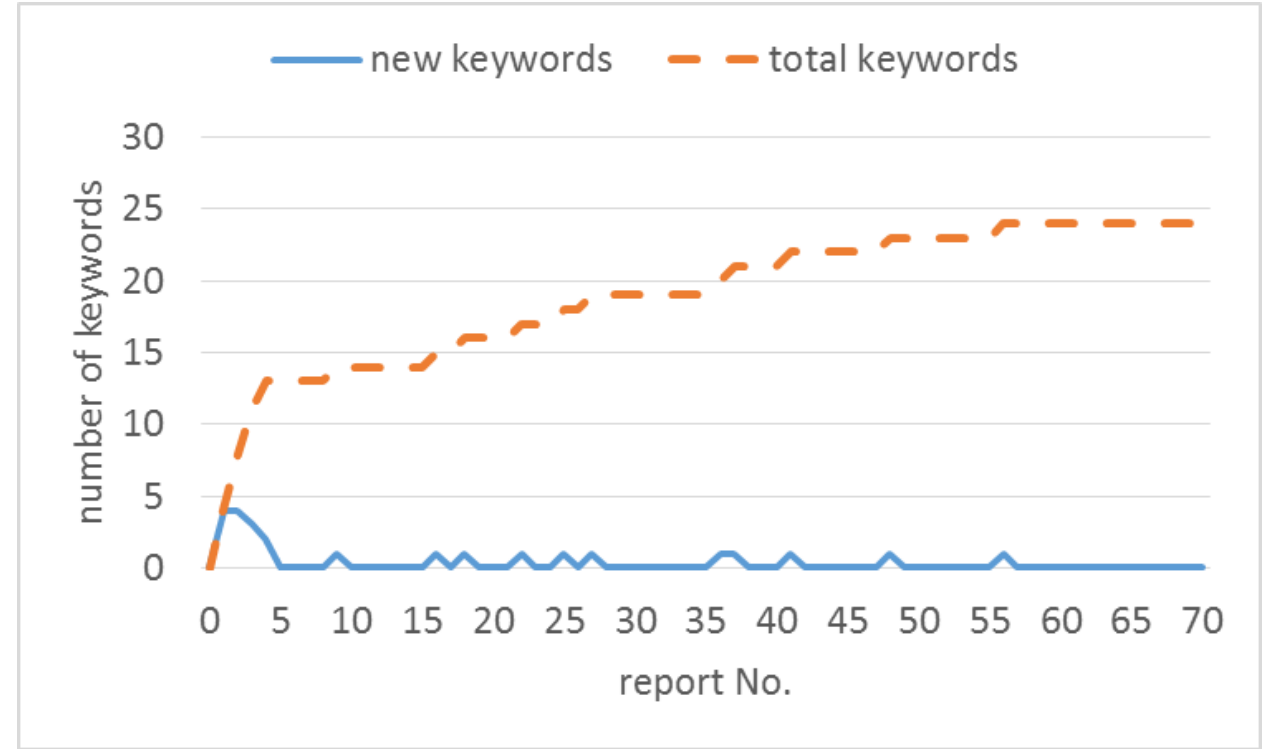
- How do we develop the rules?
 - Manually analyze the sentence and identify the cause and effect
 - Identify corresponding part of speech and dependency for cause and effect
 - In this example, the rule is (*acl, NNP*) for effect and (*nmod:by, NNP*) for cause





Results - Keywords

- Randomly selected 70 event reports;
- Identified 24 keywords in total;
- Figure on right shows the relation between # of keywords and reports analyzed;
- Table on right shows the keywords.



NO.	Keyword	NO.	Keyword
1	Result in	13	As a result
2	Caused	14	Because
3	Due to	15	Impact
4	Be caused by	16	Be attributed to
5	Result from	17	Initiate
6	Follow	18	Produce
7	As a result of	19	Lead to
8	[the/a ...] cause(s) of	20	Actuate
9	Be followed by	21	Because of
10	Be due to	22	Be required to
11	[the/a ...] cause(s) be	23	Contributing factor(s) be
12	In response to	24	Give rise to



Results - Rules

- Focused on two keywords
 - *Caused* and *Caused by*
- Fifty-six sentences manually analyzed
- Nine rules developed for each keyword
- Table on right shows the rules for *Caused*

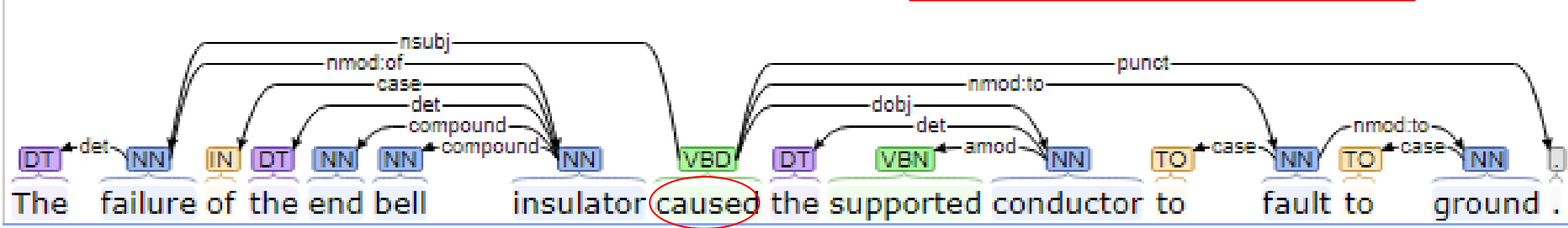
NO.	Causal Event		Consequent Event	
	Dependency	Part of Speech	Dependency	Part of Speech
1	nsubj	NN	dobj	NNS
2	nsubj	NNS	dobj	JJ
3	nsubj	NN	dobj	NN
4	nsubj	DT	dobj	NN
5	nsubj	VB	dobj	NN
6	nsubj	DT	ccomp	NN
7	nsubj	NN	dobj	NNP
8	acl:relcl	WDT	dobj	NN
9	nsubj:xsubj	NN	dobj	NN

Results – Test

- How to use the rules?
- Step 1:
 - Identify the keyword
- Step 2:
 - Implement part of speech tagging and dependency parsing
- Step 3:
 - Match rule

NO.	Causal Event		Consequent Event	
	Dependency	Part of Speech	Dependency	Part of Speech
1	nsubj	NN	dobj	NNS
2	nsubj	NNS	dobj	JJ
3	nsubj	NN	dobj	NN
4	nsubj	DT	dobj	NN
5	nsubj	VB	dobj	NN
6	nsubj	DT	ccomp	NN
7	nsubj	NN	dobj	NNP
8	acl:relcl	WDT	dobj	NN
9	nsubj:xsubj	NN	dobj	NN

Rule #3 matches this example





Results – Test

- Apply the rules to 20 new sentences for each keyword
- Tables on right show the test result (85% accuracy for each keyword)

Caused

Number of sentences	Corresponding rule NO.
15	3
2	1
3	No rule matches

Caused by

Number of sentences	Corresponding rule NO.
11	3
5	1
1	8
3	No rule matches



Summary

- We developed a method for identifying causal relationships from text-based event reports
- Natural language processing techniques are used to facilitate the analysis
- The method was applied to the NRC LER reports as a case study
 - Keywords identified
 - Rules for two keywords developed
 - Rules developed tested on new sentences
- The proposed method can be readily extended to other event report databases



Acknowledgements

- This research is being performed using funding received from the DOE Office of Nuclear Energy's Nuclear Energy University Programs. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the view of the US Department of Energy.
- The authors would also like to acknowledge the efforts of the undergraduate students involved in this research: Michael Celesti and Charles Cummings for their assistance in analysing the reports and summarizing the rules; and Edillower Wang for his work at the beginning of the project and the introduction of the Stanford CoreNLP API.



Thank you!
Any question?

Yunfei Zhao: zhao.2263@osu.edu