Synergy of Linguistic Science and Technology: NLP's Linguistic Endeavors

Presented by : Hemanta Baruah

Centre for Linguistic Science & Technology

IIT Guwahati

OUTLINE:

1. Introduction
2. What is NLP ?
3. The Linguistic Landscape
4. NLP Tasks and Applications
5. Importance of NLP in Multidisciplinary Context
6. Synergy of Linguistic Science and Technology
7. Challenges and Future Directions
8. Conclusion and Q&A

1. Introduction

- Welcome to the presentation

- Overview
  - Explore the intersection of linguistic science and technology

- Agenda

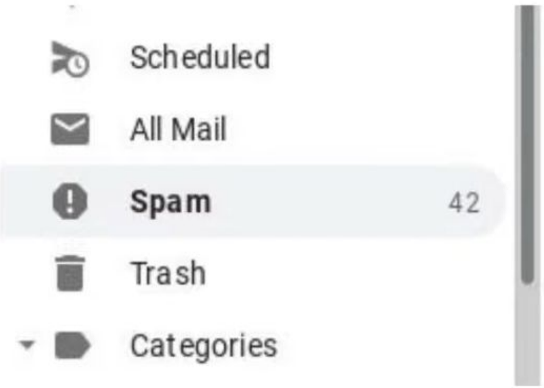
  - Journey through key NLP applications

  - Significance of these applications in addressing everyday linguistic challenges

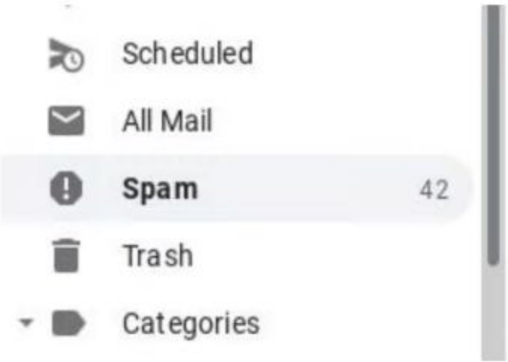

Scheduled
All Mail
Spam
42
Trash
Categories

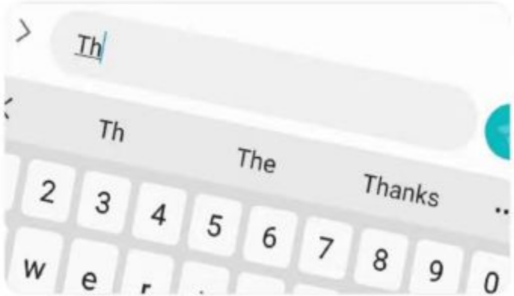

Scheduled
All Mail
Spam 42
Trash
Categories

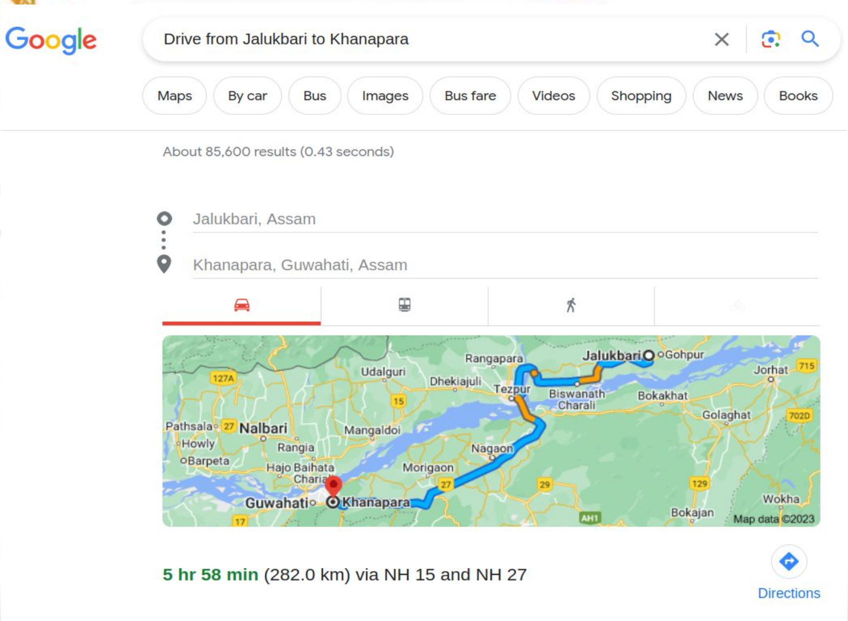

Th
Th
The
Thanks


Scheduled
All Mail
Spam
42
Trash
Categories

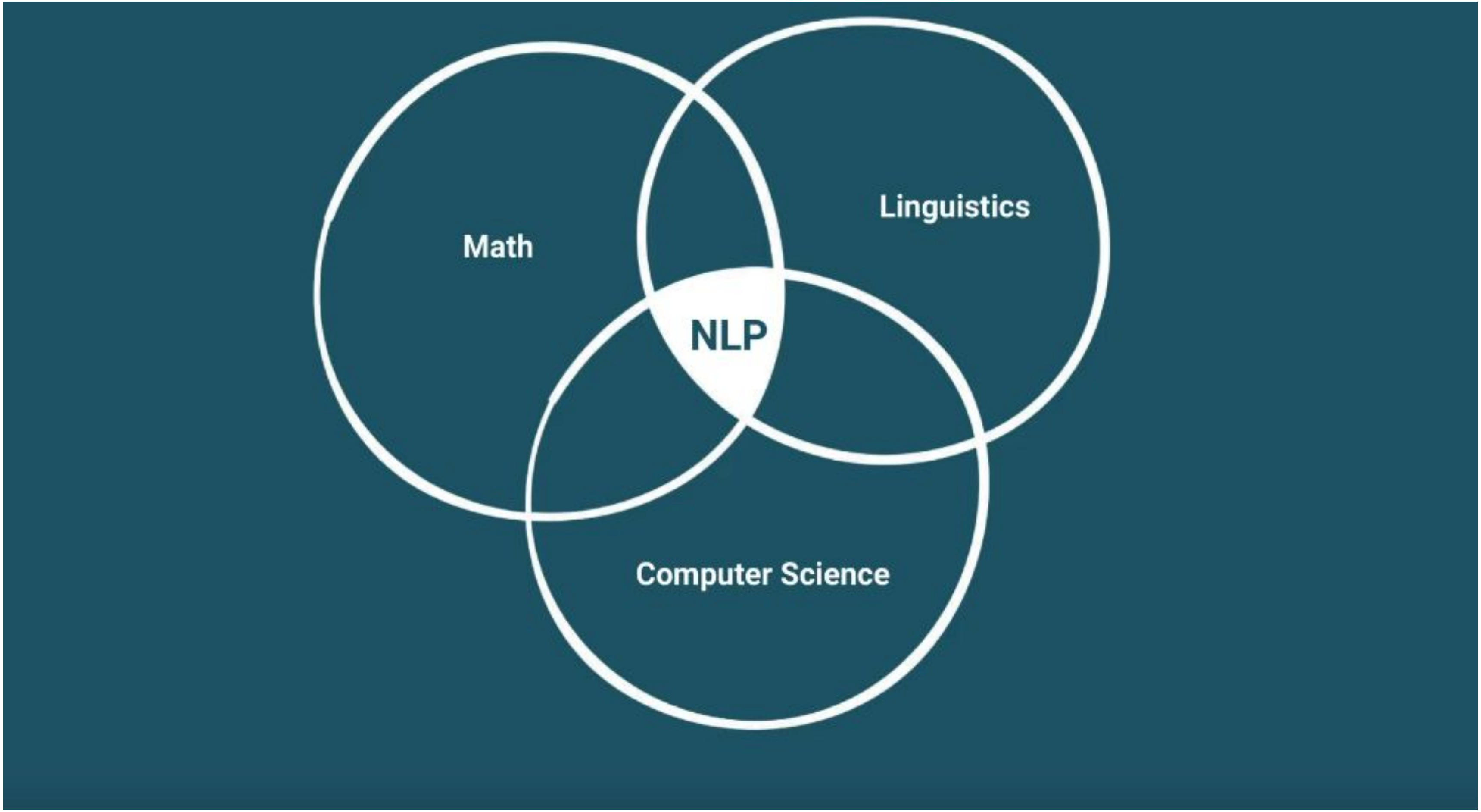

Th
Th
The
Thanks
2 3 4 5 6 7 8 9 0
w e


Scheduled
All Mail
Spam
42
Trash
Categories

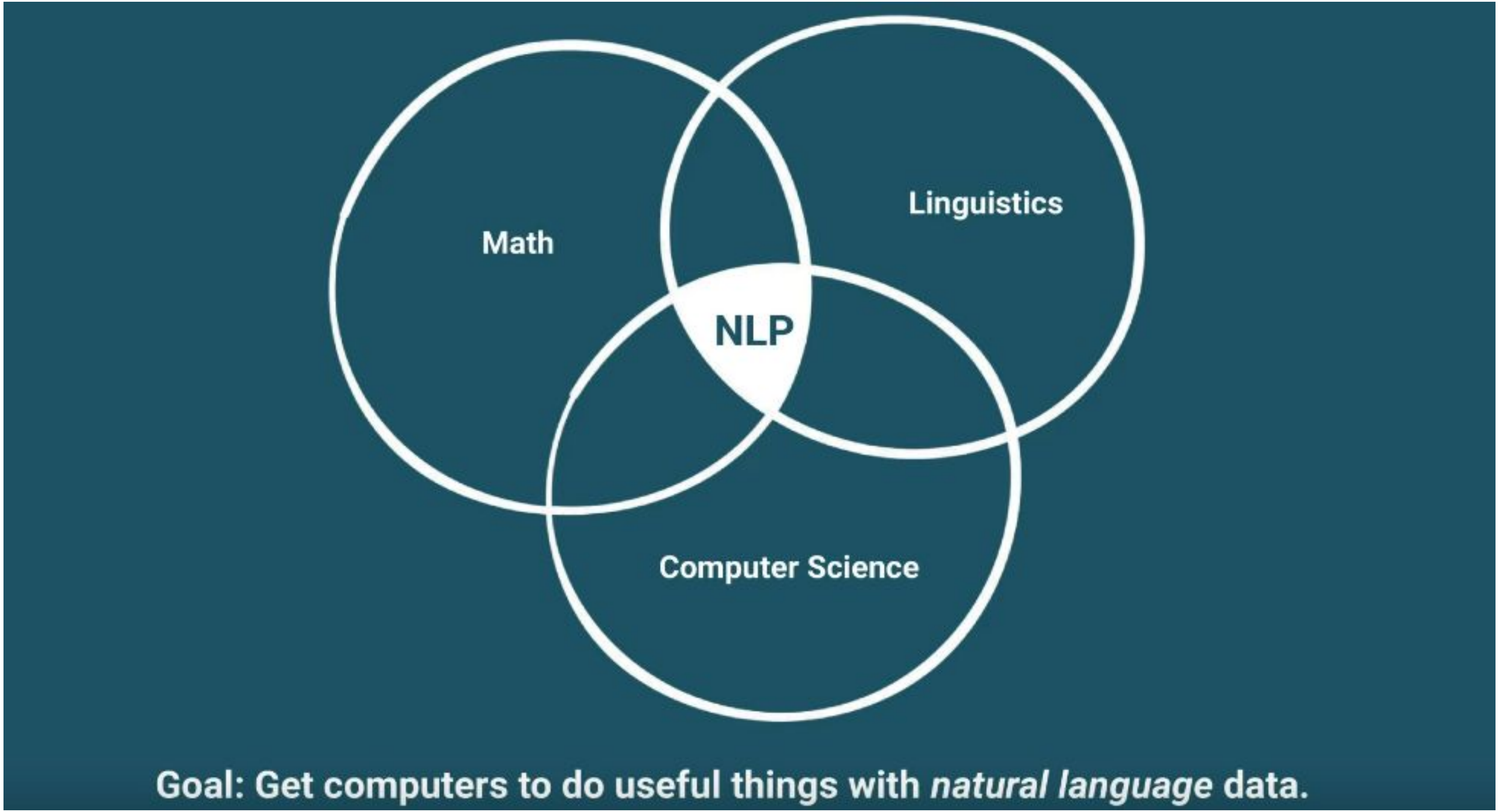

Th
Th
The
Thanks
2 3 4 5 6 7 8 9 0
w e r


Google
Drive from Jalukbari to Khanapara
Maps
By car
Bus
Images
Bus fare
Videos
Shopping
News
Books
About 85,600 results (0.43 seconds)
Jalukbari, Assam
Khanapara, Guwahati, Assam
Jalukbari
Gohpur
Rangapara
Udalguri
Dhekiajuli
Tezpur
Biswanath Charali
Bokakhat
Jorhat
Golaghat
Pathsala
Nalbari
Howly
Barpeta
Rangia
Mangaldoi
Nagaon
Hajo
Baihata Chariali
Morigaon
Guwahati
Khanapara
Bokajan
Wokha
Map data ©2023
5 hr 58 min (282.0 km) via NH 15 and NH 27
Directions

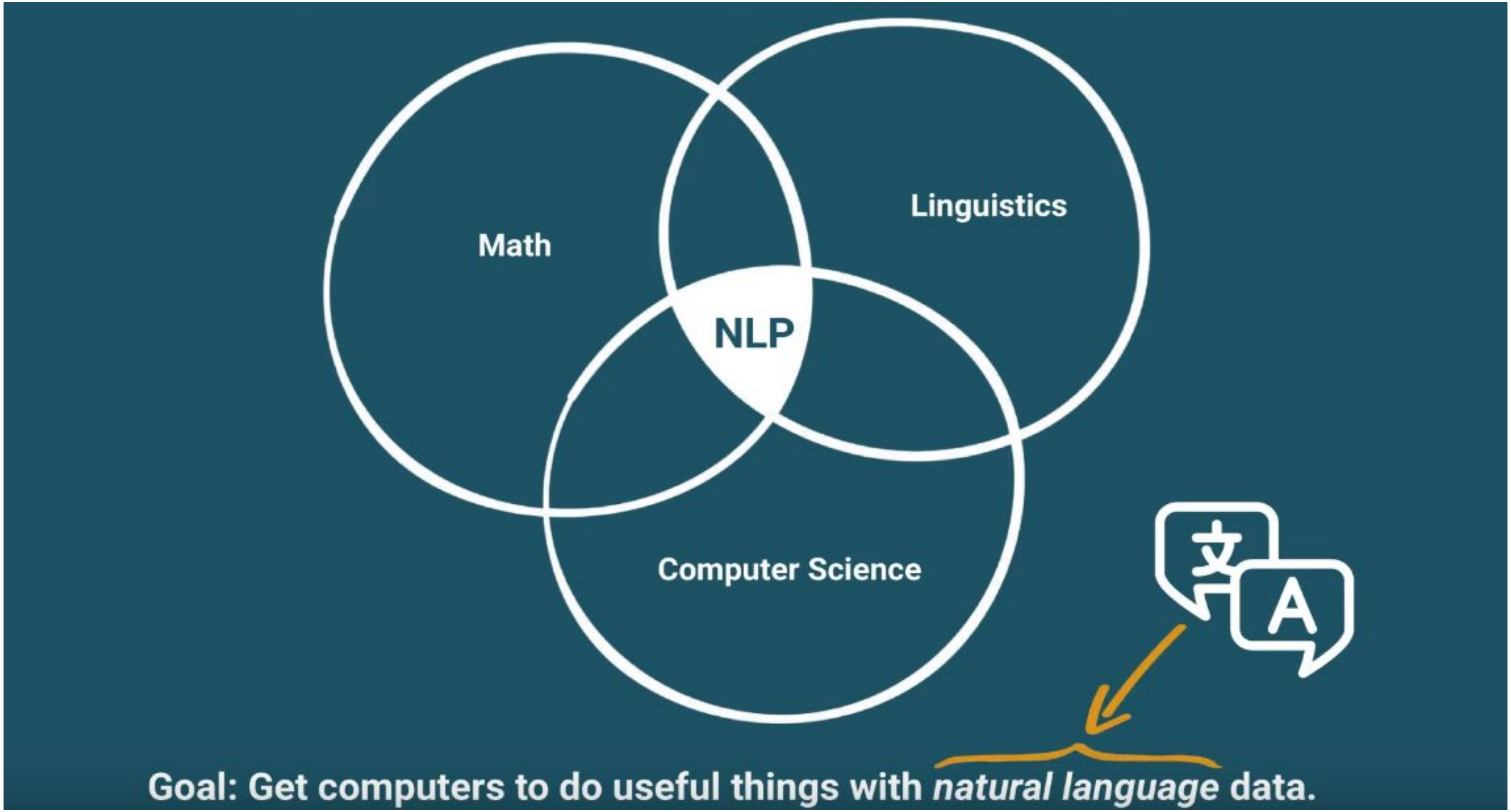
2. What is NLP ?



2. What is NLP ?


Math
Linguistics
NLP
Computer Science
Goal: Get computers to do useful things with *natural language* data.

2. What is NLP ?


Math
Linguistics
NLP
Computer Science
文
A
Goal: Get computers to do useful things with *natural language* data.

3. The Linguistic Landscape

Why is NLP difficult?

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

By a tree, a stolen painting
was found.

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

By a tree, a stolen painting was found.	vs.	A tree found a stolen painting.

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

Prepositional Phrase

By a tree, a stolen painting was found.

vs.

A tree found a stolen painting.

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

Prepositional Phrase

By a tree, a stolen painting was found. vs. A tree found a stolen painting.

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

Prepositional Phrase

By a tree, a stolen painting was found. vs. A tree found a stolen painting.

3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"Stolen painting found by tree"

Prepositional Phrase

By a tree, a stolen painting was found.

vs.

A tree found a stolen painting.

PP-attachment problem

syntactic ambiguity

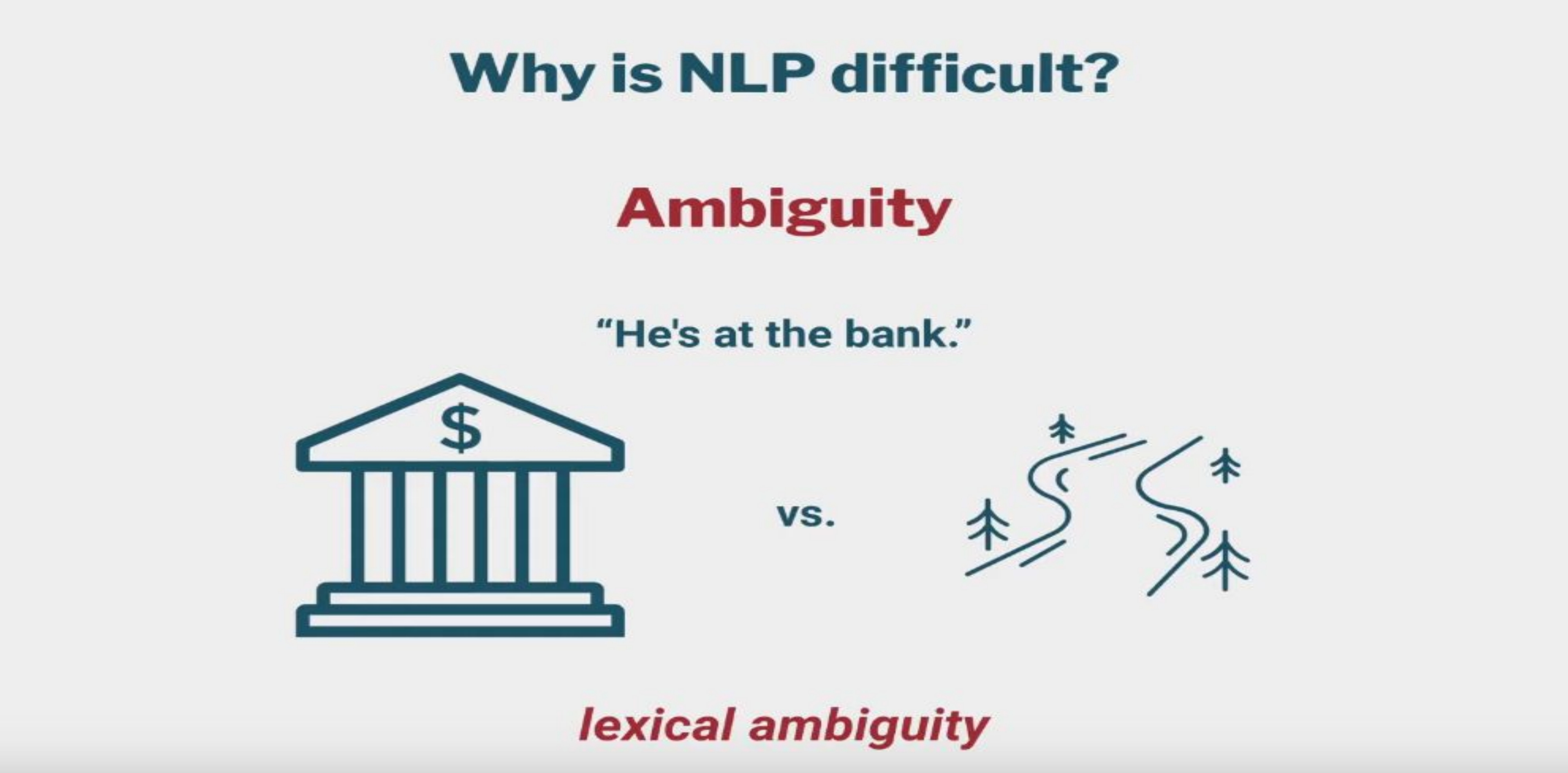
3. The Linguistic Landscape

Why is NLP difficult?

Ambiguity

"He's at the bank."

3. The Linguistic Landscape



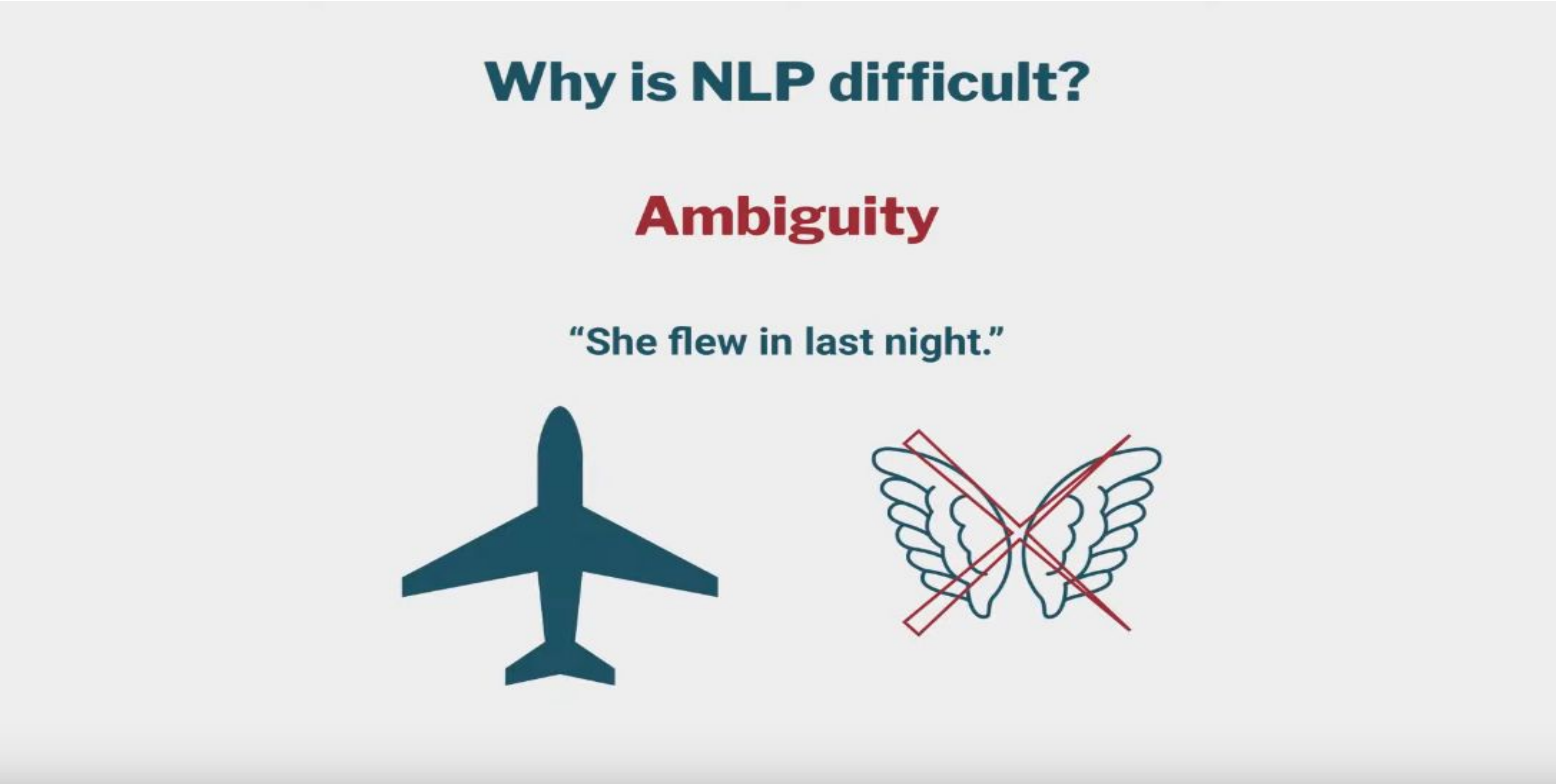
3. The Linguistic Landscape

Why is NLP difficult?

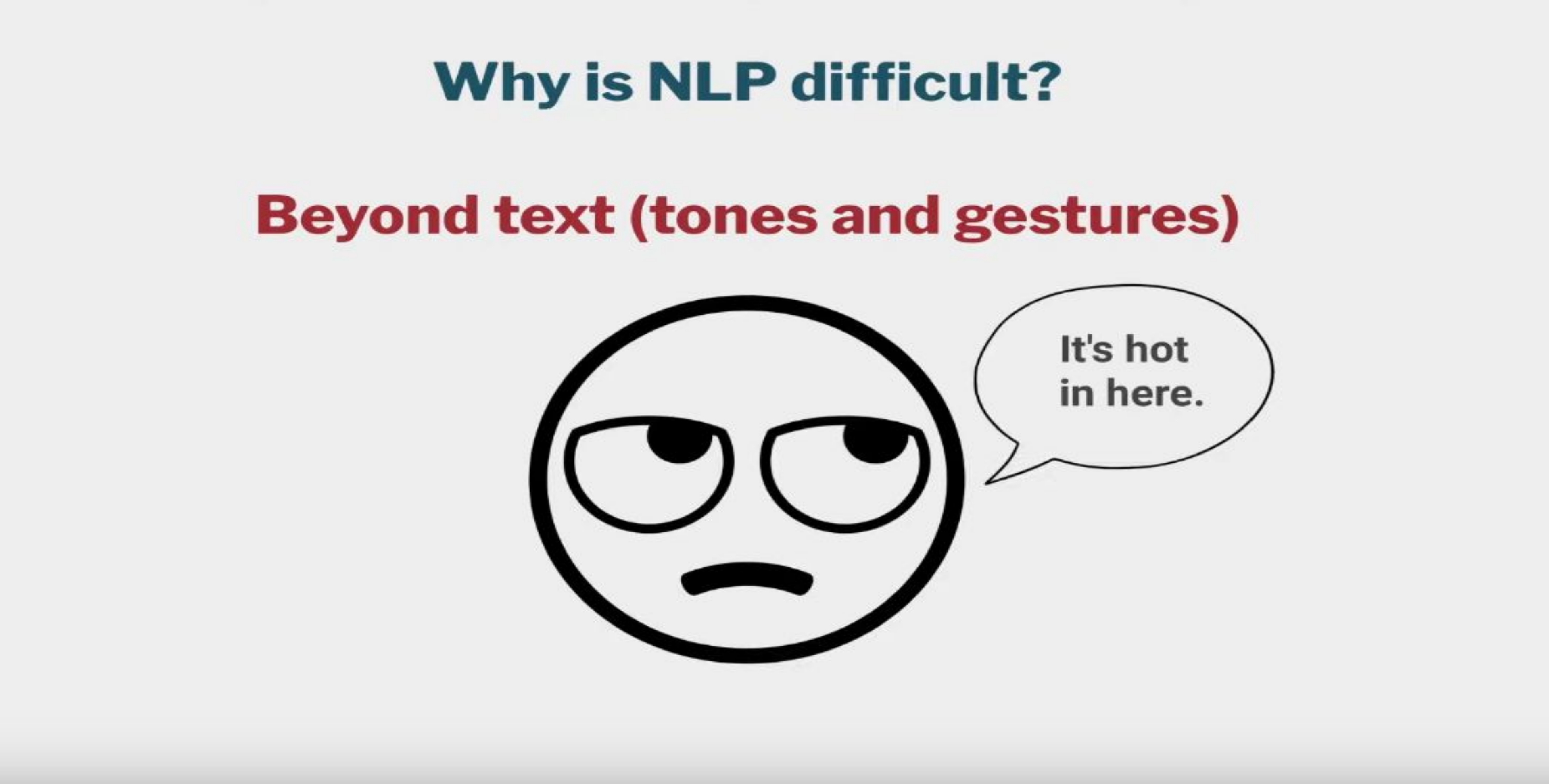
Ambiguity

“She flew in last night.”

3. The Linguistic Landscape



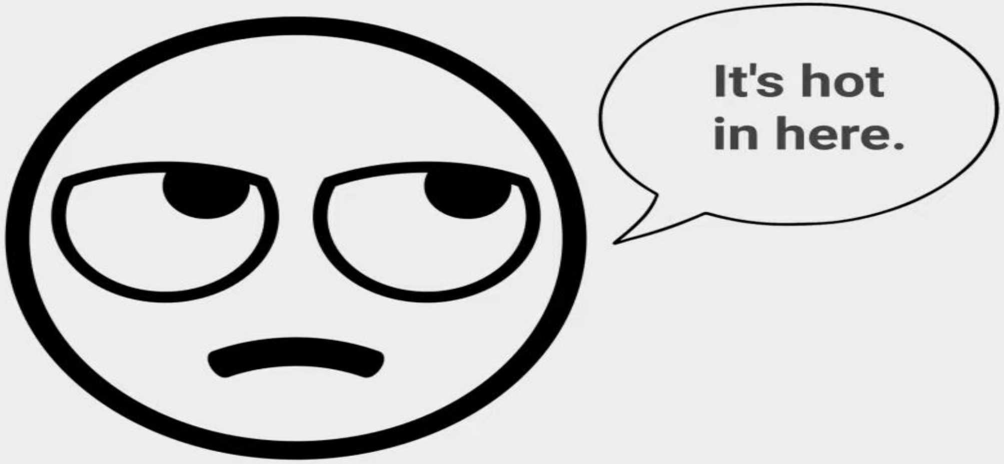
3. The Linguistic Landscape



3. The Linguistic Landscape

Why is NLP difficult?

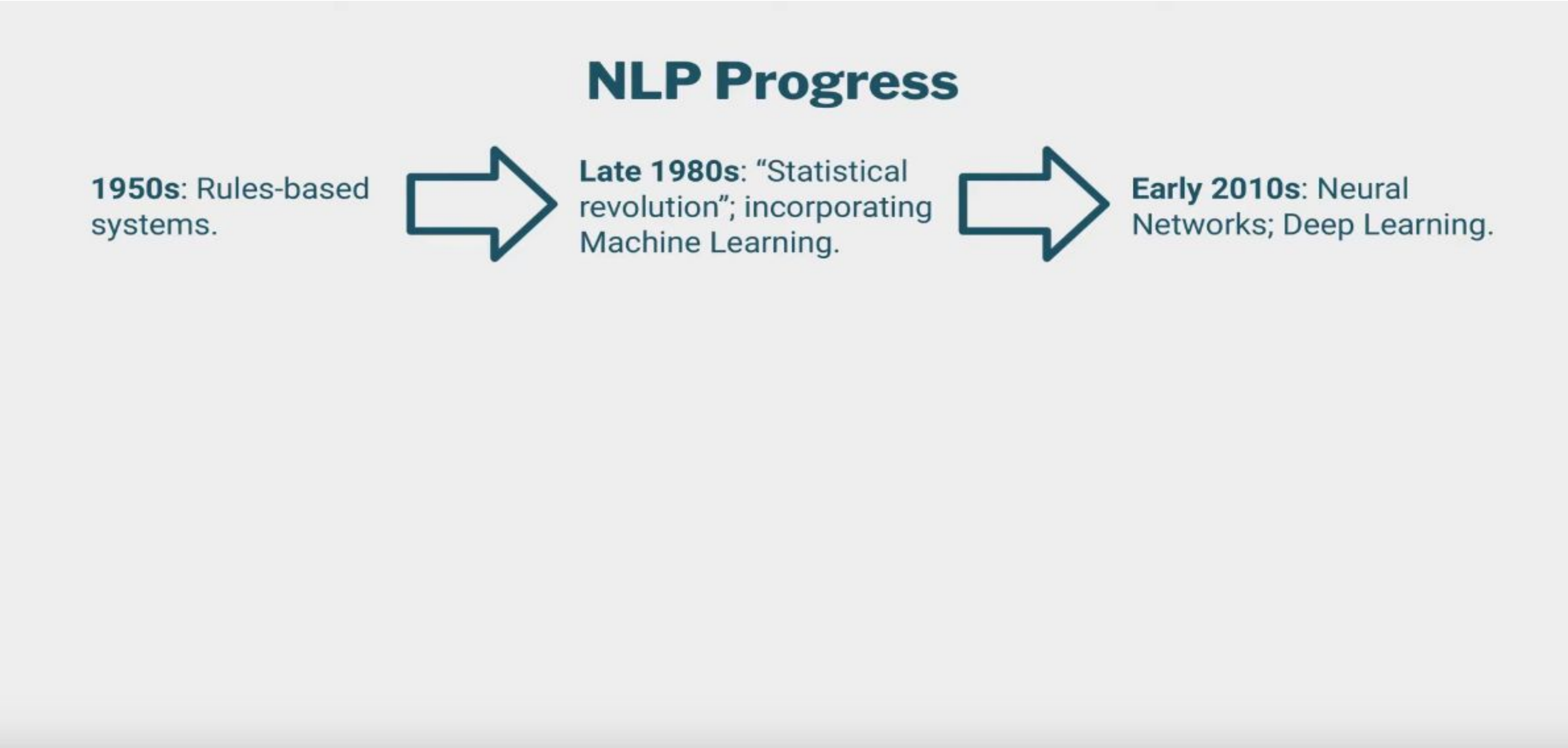
Beyond text (tones and gestures)



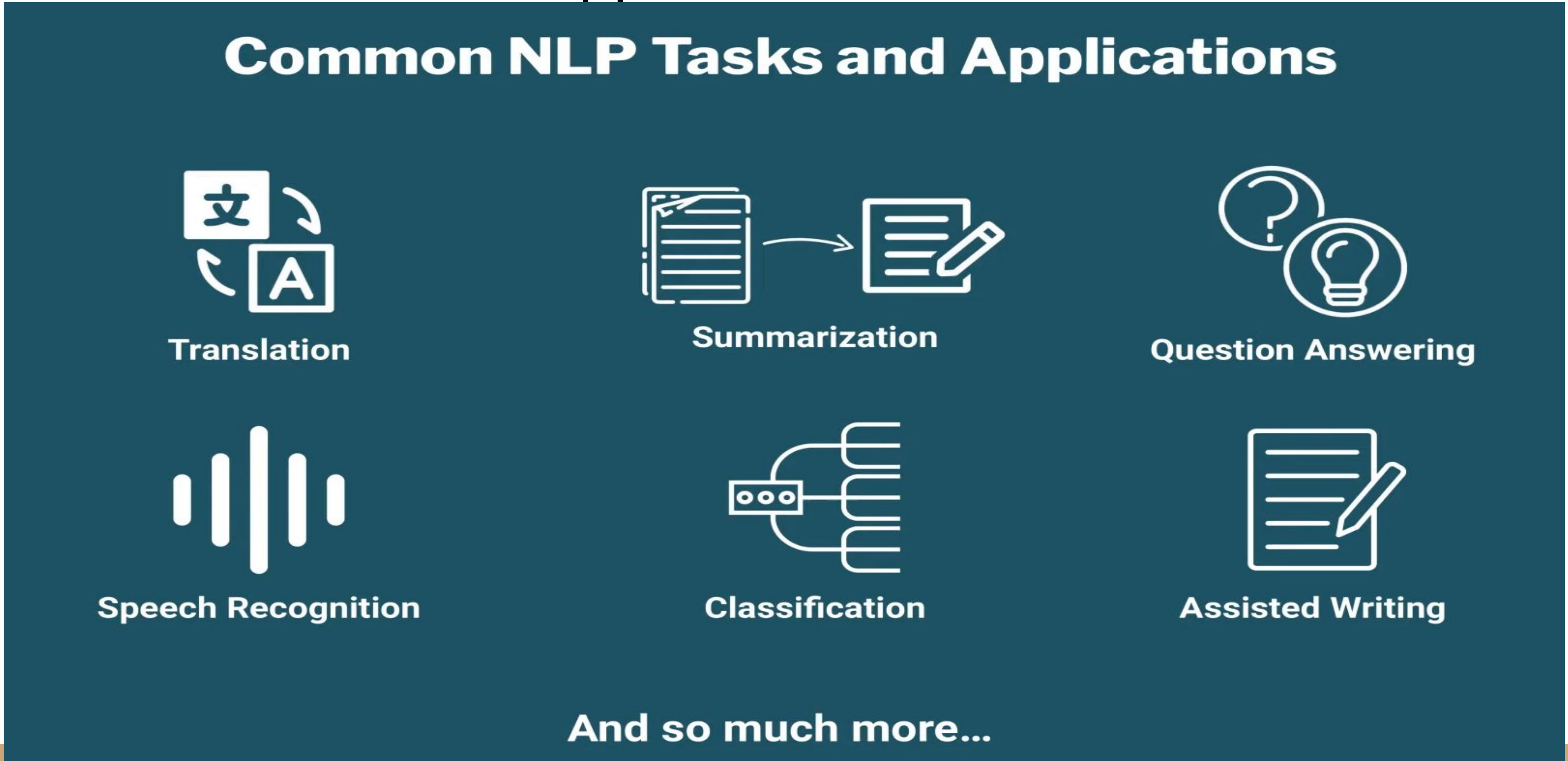
Really: "I'm bored and tired and I hate you for bringing me here."

Pragmatics (how people use language)

Progress in NLP



4. NLP Tasks and Applications



4.1 Few Classic NLP Problems: POS

Part-of-Speech (POS) Tagging

Classifying how a word is used in a sentence.

{ noun, verb, adjective, ... }

4.1 Few Classic NLP Problems: POS

Part-of-Speech (POS) Tagging

Classifying how a word is used in a sentence.

{ noun, verb, adjective, ... }

POS tagging assigns each word in a sentence its corresponding POS.

[John, watched, an, old, movie, at, the, cinema, .]

4.1 Few Classic NLP Problems: POS

Part-of-Speech (POS) Tagging

Classifying how a word is used in a sentence.

{ noun, verb, adjective, ... }

POS tagging assigns each word in a sentence its corresponding POS.

[John, watched, an, old, movie, at, the, cinema, .]

['PROPN', 'VERB', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'PUNCT']

4.1 Few Classic NLP Problems: POS

Part-of-Speech (POS) Tagging

Classifying how a word is used in a sentence.

{ noun, verb, adjective, ... }

Parts of speech can help discover intent or action.

"How can I help you today?"

"I want to book a hotel room."

VERB

NOUN

e.g. you could scan for *VERB-NOUN* patterns. Scanning for POS patterns can also help with other tasks such as information extraction.

4.1 Few Classic NLP Problems: POS

Part-of-Speech (POS) Tagging

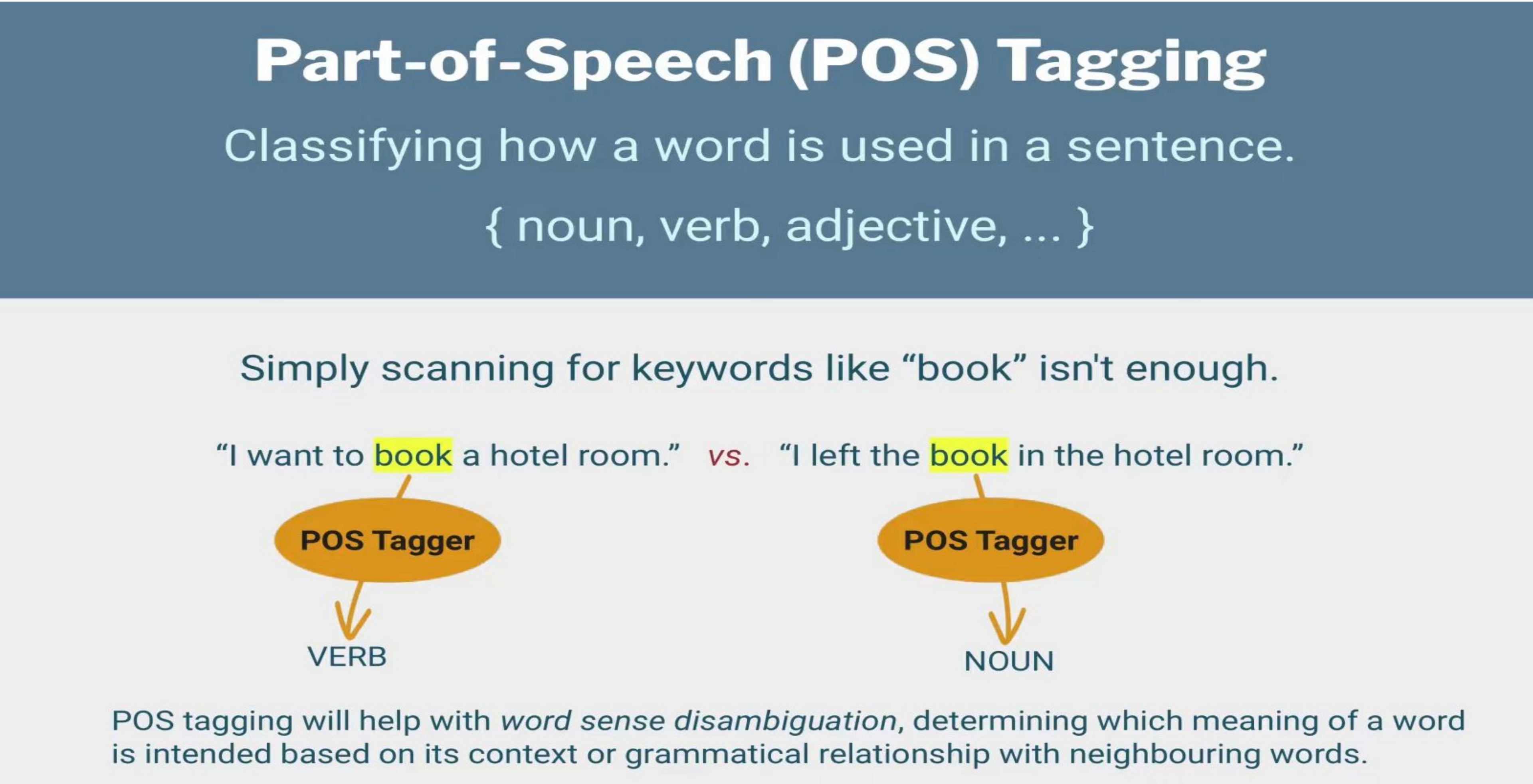
Classifying how a word is used in a sentence.

{ noun, verb, adjective, ... }

Simply scanning for keywords like "book" isn't enough.

"I want to book a hotel room." *vs.* "I left the book in the hotel room."

4.1 Few Classic NLP Problems: POS



4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Named Entity: (roughly) anything that can be referred by a proper name. They often have a *Proper Noun (PROPN)* POS tag.

4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Named Entity: (roughly) anything that can be referred by a proper name. They often have a *Proper Noun (PROPN)* POS tag.

Most Common

Person (PER)
"Alice", "Isaac Newton"

Location (LOC)
"Bay Area", "Rocky Mountains"

Geopolitical Entity (GPE)
"Canada", "Chicago"

Organization (ORG)
"Microsoft", "Porsche"

4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Named Entity: (roughly) anything that can be referred by a proper name.
They often have a *Proper Noun (PROPN)* POS tag.

But also often extended to other types

- **Money**
- **Time**
- **Date**

.
.
.

4.2 Few Classic NLP Problems: NER

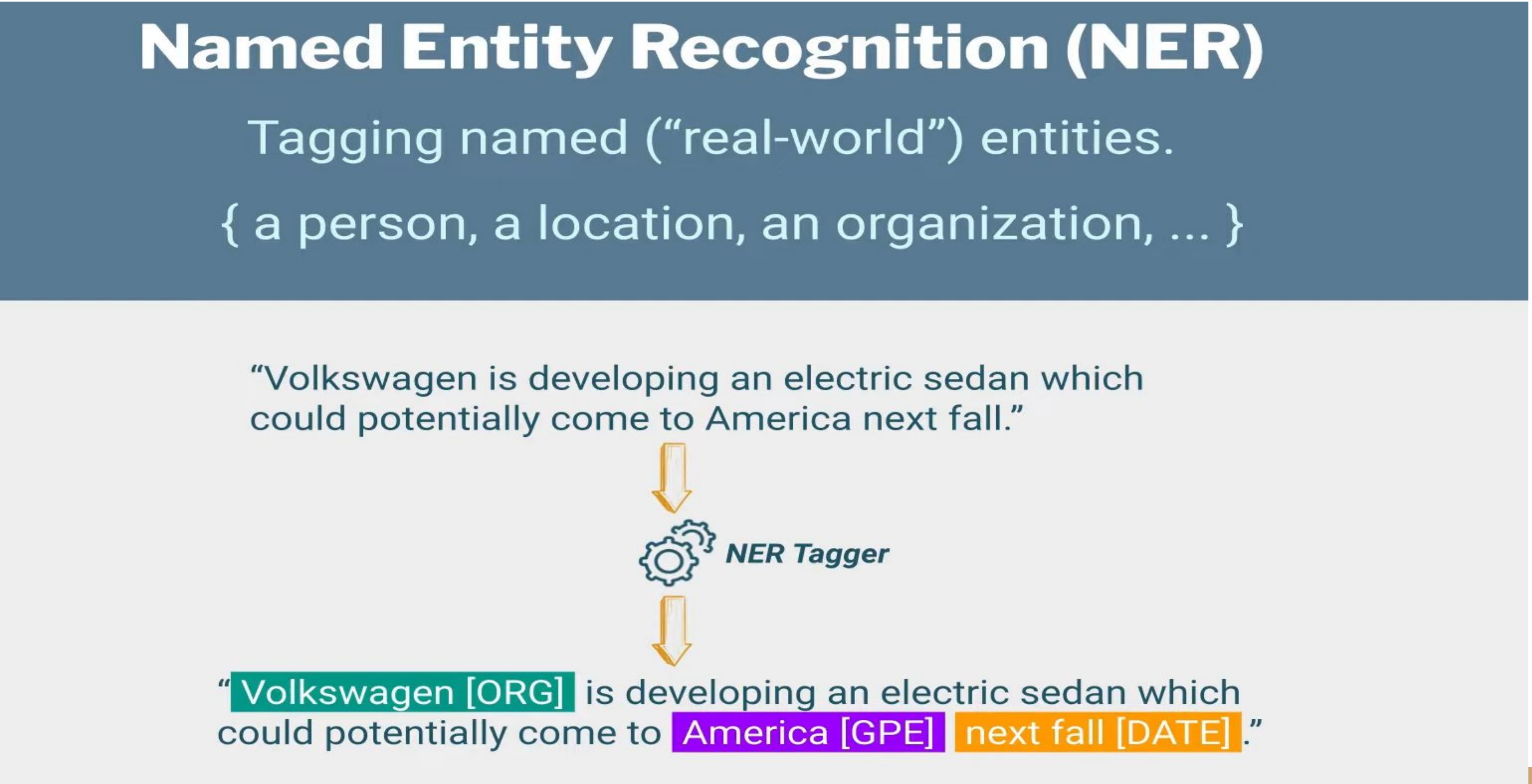
Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

"Volkswagen is developing an electric sedan which could potentially come to America next fall."

4.2 Few Classic NLP Problems: NER



4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Useful in a variety of tasks and applications...

Organizing/Categorizing corpus

e.g. identify medical procedures or diseases in research, or categorize support tickets based on entities mentioned.

4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Useful in a variety of tasks and applications...

Organizing/Categorizing corpus

e.g. identify medical procedures or diseases in research, or categorize support tickets based on entities mentioned.

Question answering

e.g. extract entities from a question and use NER to narrow down possible candidate answers. A question about a country's capital is going to result in an answer that's either LOC or GPE.

4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

Useful in a variety of tasks and applications...

Organizing/Categorizing corpus

e.g. identify medical procedures or diseases in research, or categorize support tickets based on entities mentioned.

Question answering

e.g. extract entities from a question and use NER to narrow down possible candidate answers. A question about a country's capital is going to result in an answer that's either LOC or GPE.

Critical in information extraction

e.g. extracting events and relationships between entities.

4.2 Few Classic NLP Problems: NER

Named Entity Recognition (NER)

Tagging named ("real-world") entities.

{ a person, a location, an organization, ... }

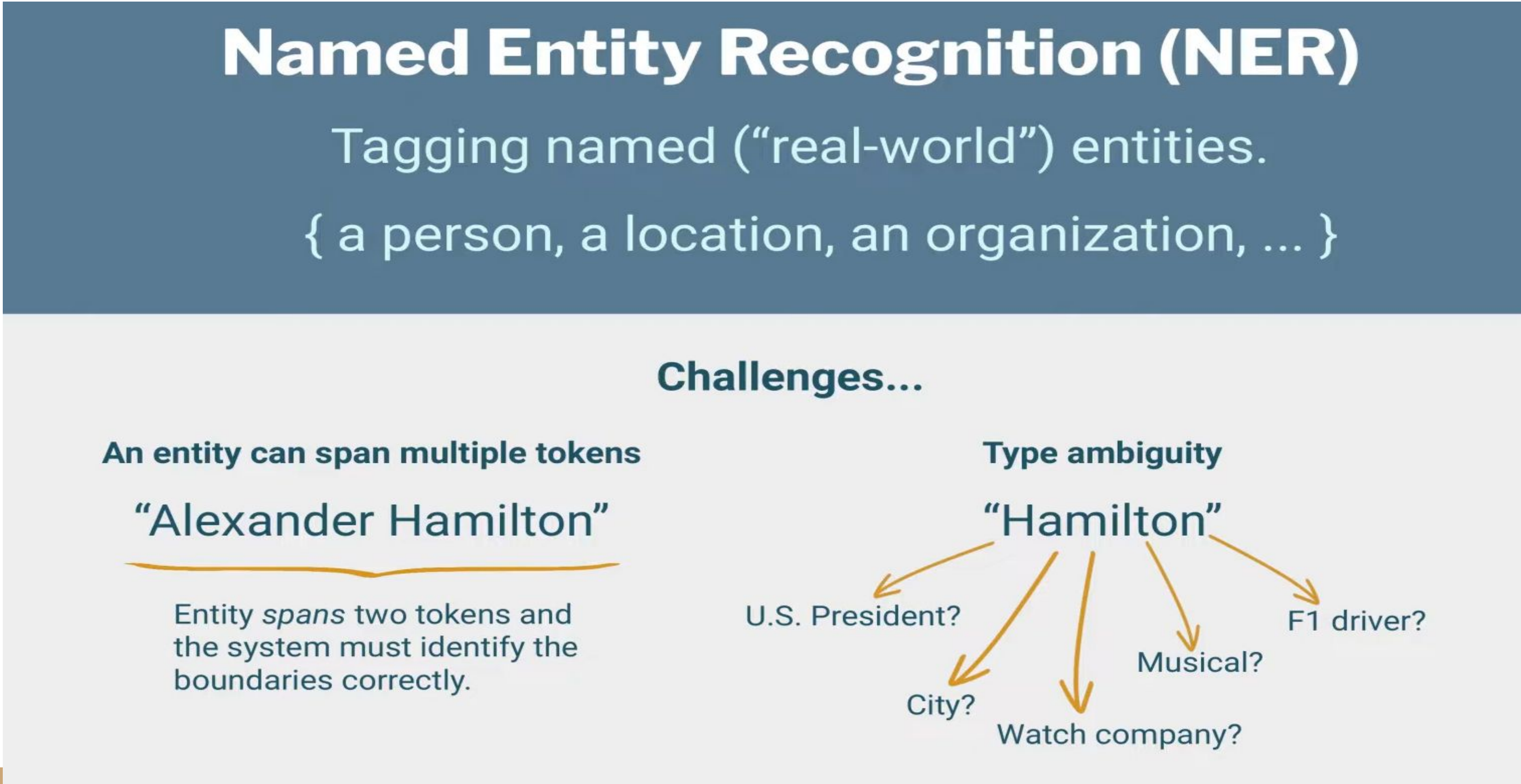
Challenges...

An entity can span multiple tokens

"Alexander Hamilton"

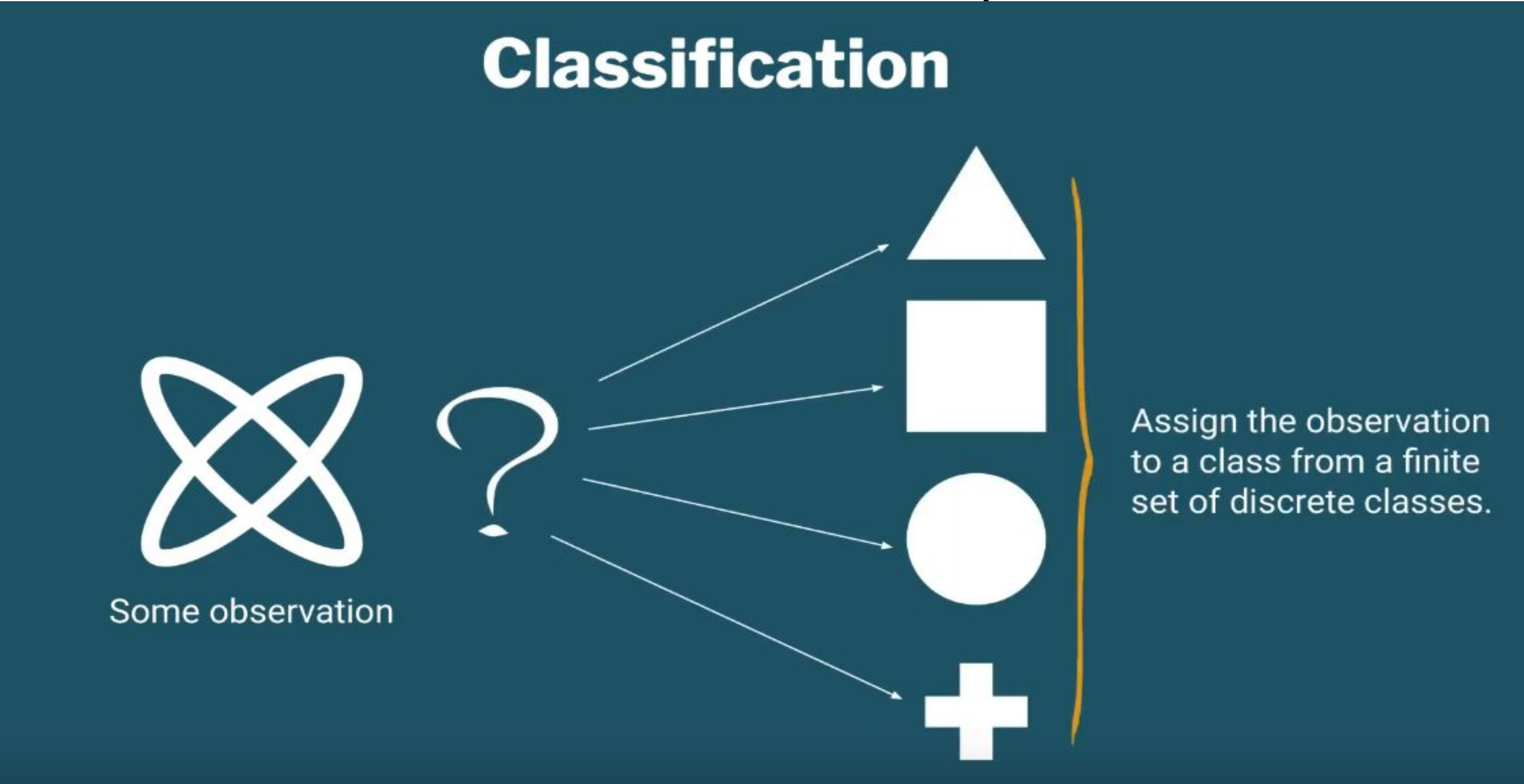
Entity *spans* two tokens and the system must identify the boundaries correctly.

4.2 Few Classic NLP Problems: NER



4.3 Few Common NLP Problems: Classification

4.3 Few Common NLP Problems: Classification



4.3 Few Common NLP Problems: Classification

4.3 Few Common NLP Problems: Classification

4.3 Few Common NLP Problems: Classification

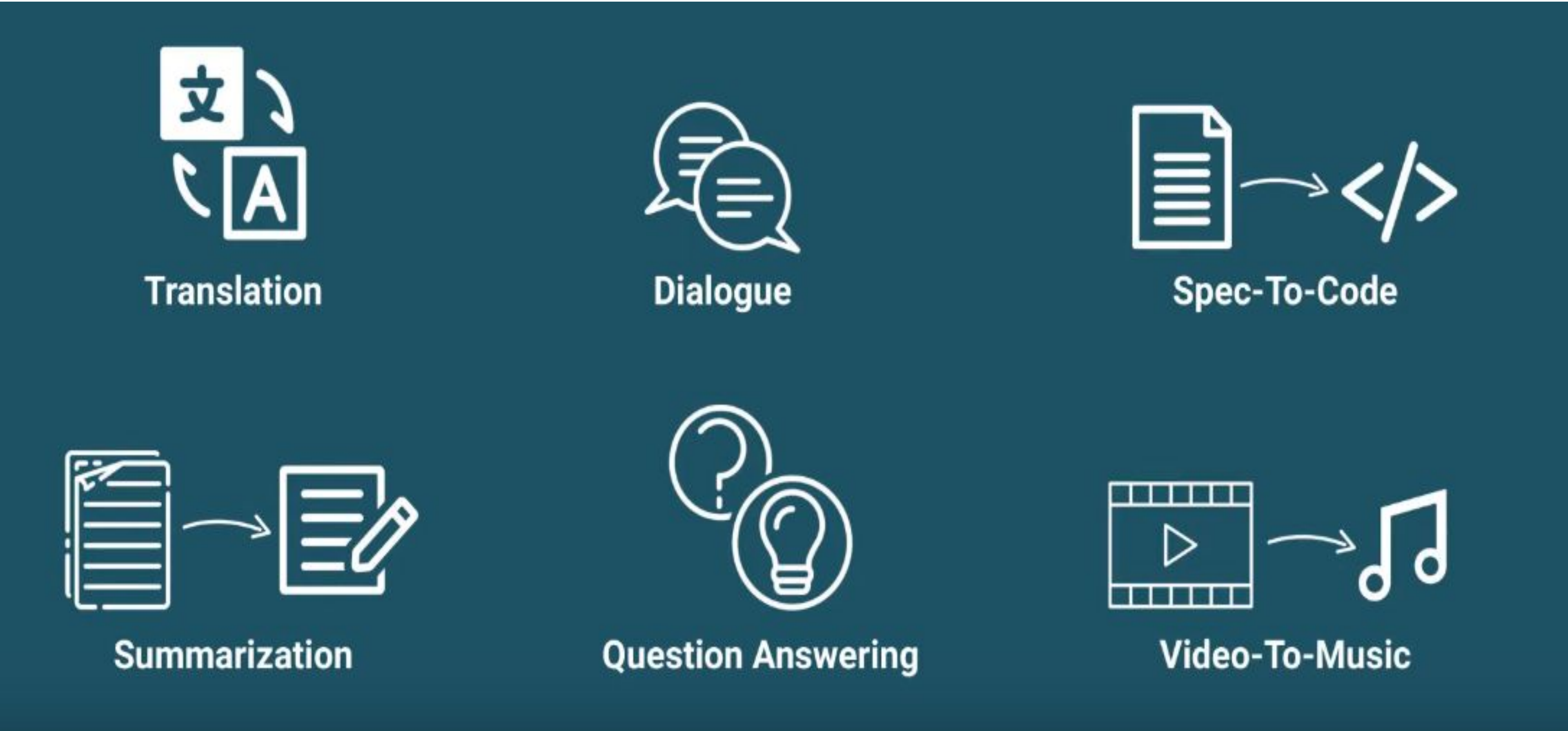
4.4 Few Advanced NLP Problems: Sequence2Sequence

Translation

Summarization

4.4 Few Advanced NLP Problems: Sequence2Sequence

4.4 Few Advanced NLP Problems: Sequence2Sequence



5. Importance of NLP in Multidisciplinary Context

- **Cross-Industry Applications: Similar NLP approaches can be applied in various industries:**
  - **Healthcare:** Accurate translation in medical contexts for better patient care.

5. Importance of NLP in Multidisciplinary Context

- **Cross-Industry Applications: Similar NLP approaches can be applied in various industries:**
  - **Law:** Help with the precedence search in Law.

5. Importance of NLP in Multidisciplinary Context

- **Cross-Industry Applications: Similar NLP approaches can be applied in various industries:**
    - **Finance:** Report and News analysis in Finance.

5. Importance of NLP in Multidisciplinary Context

- **Cross-Industry Applications: Similar NLP approaches can be applied in various industries:**
  - **Education:** Language learning apps that adapt to individual needs.

6. Synergy of Linguistic Science and Technology

- **Linguistic Expertise Enhances NLP**

  - **Rule Formulation:**
    - **Linguistic experts formulate rules and guidelines. For example, understanding that names can have different forms, titles, or nicknames helps improve the precision of NER models.**
  - **Training Data Creation:**
    - **Expertise crucial for creating training data in tasks like Machine Translation, POS Tagging, Dependency Parsing etc. for state-of-the-art models.**
  - **Outcome Validation:**
    - **Only linguistic experts can effectively validate outcomes of certain linguistic experiments.**
  - **Key to Research:**
    - **Well-versed linguistic experts are essential for multidisciplinary computational linguistic research.**

6. Synergy of Linguistic Science and Technology

- **Technology Facilitates Linguistic Research**
  - **Automatic POS and NER Tagger:**
    - **Build an automatic system for part-of-speech (POS) tagging and Named Entity Recognition (NER) to enhance linguistic analysis**
  - **Automatic Translator and Transliterator:**
    - **Create an automated system for translation and transliteration tasks.**
  - **Automatic Sentiment Analyzer, Language Identifier:**
    - **Build an automated system for sentiment analysis and language identification.**
  - **Chatbots, Conversational AI models:**
    - **Develop intelligent Chatbots and state-of-the-art Conversational AI model like ChatGPT, GPT4 for interactive linguistic interactions.**

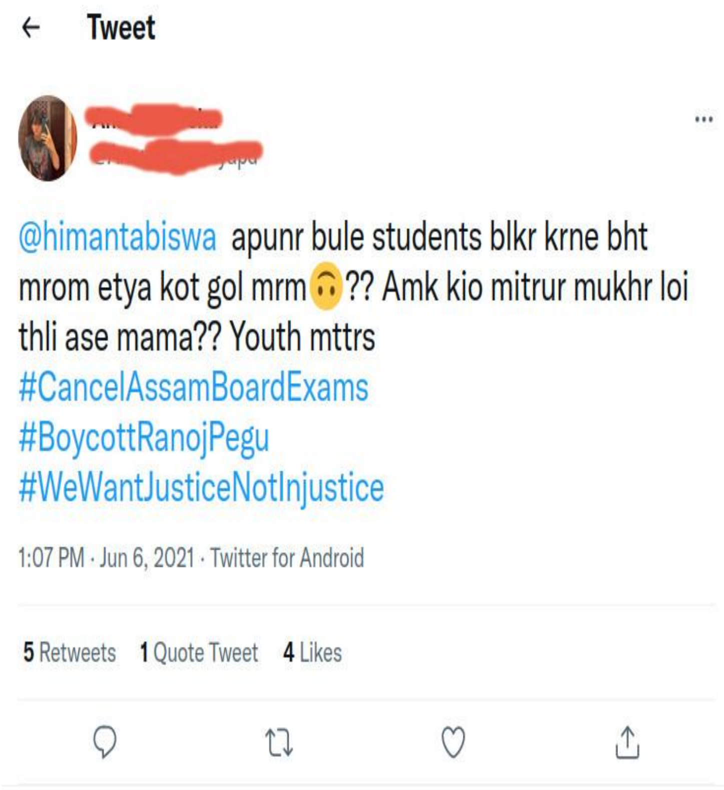
7. Challenges and Future Directions:

- Sentiment Analysis & Sarcasm Detection

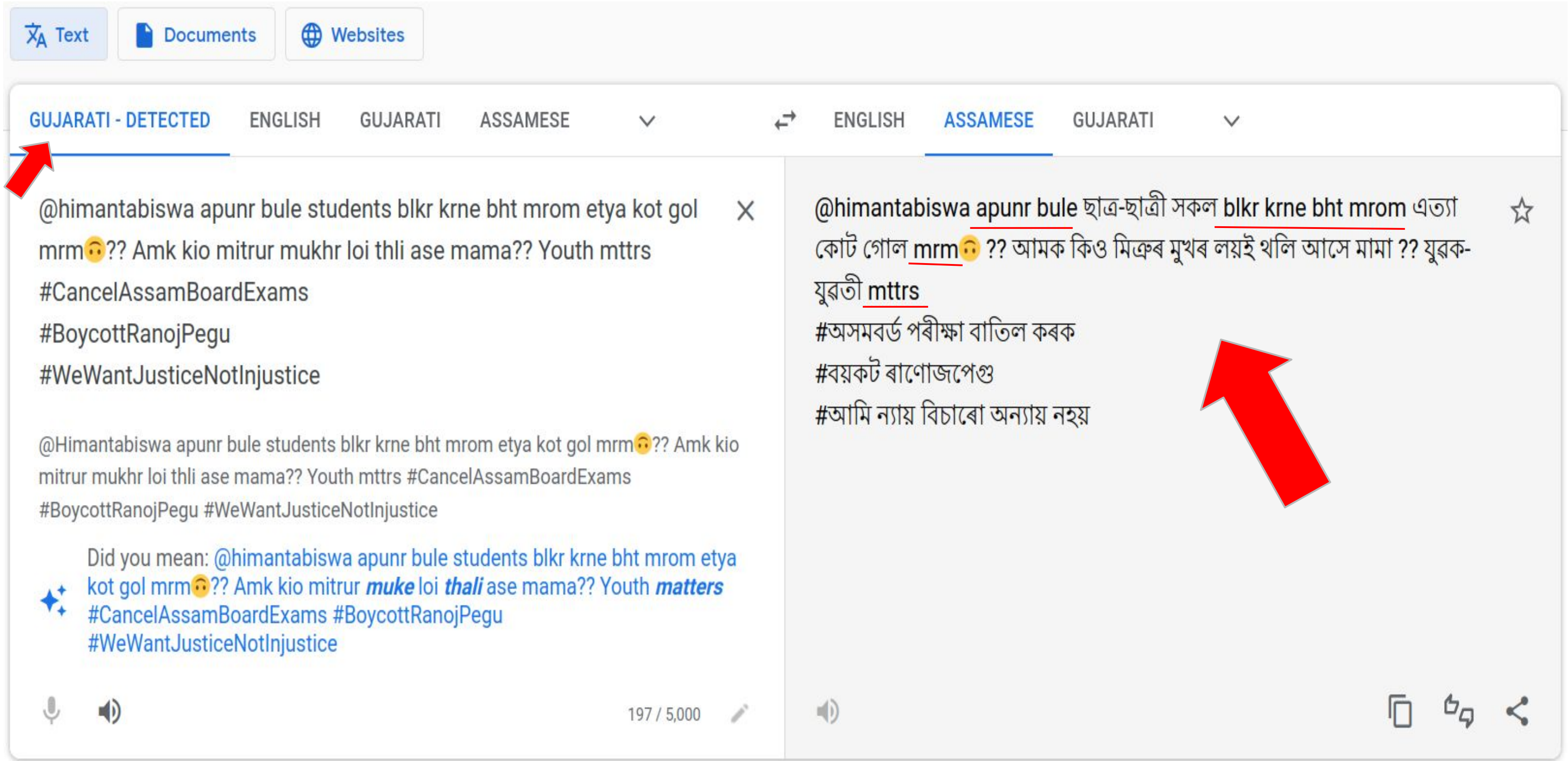
  On user generated Social Media Content

7. Challenges and Future Directions:

- Language Identification, Machine Translation & Transliteration in code-mixed Social Media Data



7. Challenges and Future Directions:



8. Conclusion

- **Linguistic-Technology Fusion**
- **Diverse NLP Applications**
- **Significance in Multidisciplinary Context**
- **Ongoing Challenges and Future Directions**
- **Call to Collaborate**

Q&A

THANK YOU

FOR

LISTENING

References

https://krishnaik.in/

http://jalammar.github.io/

https://colah.github.io/

https://www.nlpdemystified.org

8. Conclusion and Q&A

1. **Linguistic-Technology Fusion:**
   - **The presentation showcased the powerful fusion of linguistic insights with technological advancements in the realm of Natural Language Processing (NLP).**
2. **Diverse NLP Applications:**
   - **From automatic POS and NER tagging to translation, sentiment analysis, and chatbots, NLP applications demonstrated the versatility of linguistic science and technology collaboration.**
3. **Significance in Multidisciplinary Context:**
   - **Highlighted the relevance of NLP in diverse fields, emphasizing its importance in healthcare, business, education, and beyond.**
4. **Ongoing Challenges and Future Directions:**
   - **Acknowledged current challenges in tasks like sentiment analysis, processing user-generated data and looked ahead to future directions, indicating the continuous evolution of NLP.**
5. **Call to Collaborate:**
   - **Encouraged ongoing collaboration between linguists and technologists, underlining the collective impact in advancing natural language understanding and technology integration.**