

# Challenges of natural language interfaces

Historically, NLIs followed a modular process

## Analyzer

The input is broken down using semantic trees

## Mapper

Converts the raw input into a structure representation, with tokens or embeddings

Tokens: Words, punctuation, etc.

## Translator

Transforms the structure representation into something that the Database Management System (DBMS) can understand

## DMBS

Runs the DBMS-primed input against the database, retrieving the relevant data

Returns the output in a user-friendly and comprehensible format

## Dictionary

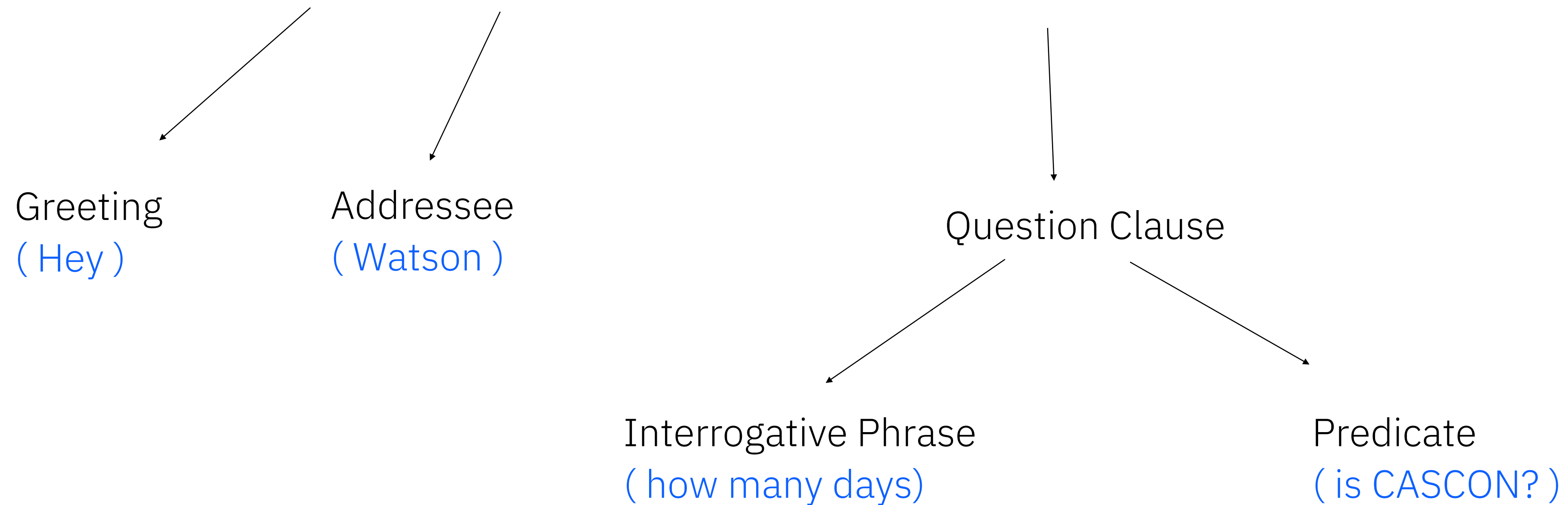
## Mapping Function

## Query Language

## Database

# Semantic tree

Hey Watson, how many days is CASCON?



Historically, NLIs followed a modular process

## Analyzer

The input is broken down using semantic trees

## Dictionary

## Mapper

Converts the raw input into a structure representation, with tokens or embeddings

Tokens: Words, punctuation, etc.

## Mapping Function

## Translator

Transforms the structure representation into something that the Database Management System (DBMS) can understand

## Query Language

## DMBS

Runs the DBMS-primed input against the database, retrieving the relevant data

Returns the output in a user-friendly and comprehensible format

## Database

# Mapping Function

Tokens: [ Find, a, hotel, in Toronto, for, Wednesday ]

- Find (action: reserve)
- hotel (entity: travel)
- Toronto (location: city)
- Wednesday (time: date)

Historically, NLIs followed a modular process

## Analyzer

The input is broken down using semantic trees

## Mapper

Converts the raw input into a structure representation, with tokens or embeddings

Tokens: Words, punctuation, etc.

## Translator

Transforms the structure representation into something that the Database Management System (DBMS) can understand

## DMBS

Runs the DBMS-primed input against the database, retrieving the relevant data

Returns the output in a user-friendly and comprehensible format

## Dictionary

## Mapping Function

## Query Language

## Database

# Translate into Query Language

```
SELECT hotel_name, location, price_per_night, availability
FROM hotels
WHERE city = 'Toronto'
      AND available_date = '2024-11-20'
ORDER BY price_per_night ASC;
```

Historically, NLIs followed a modular process

## Analyzer

The input is broken down using semantic trees

## Mapper

Converts the raw input into a structure representation, with tokens or embeddings

Tokens: Words, punctuation, etc.

## Translator

Transforms the structure representation into something that the Database Management System (DBMS) can understand

## DMBS

Runs the DBMS-primed input against the database, retrieving the relevant data

Returns the output in a user-friendly and comprehensible format

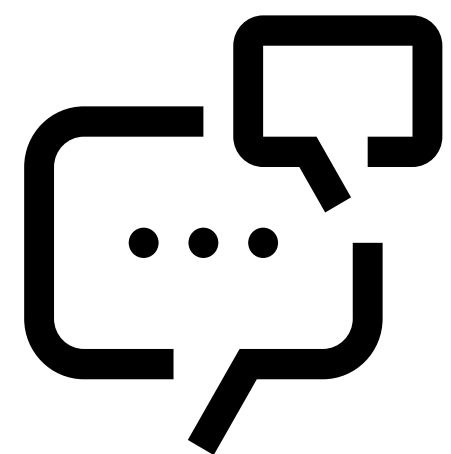
## Dictionary

## Mapping Function

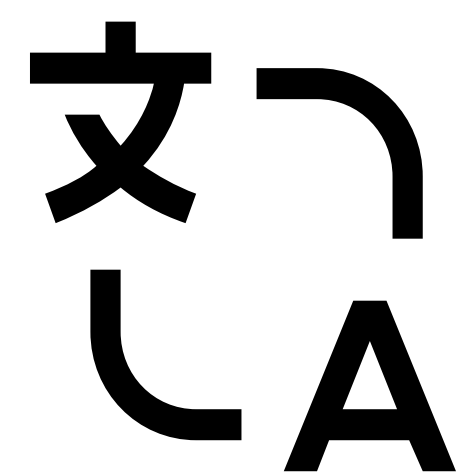
## Query Language

## Database



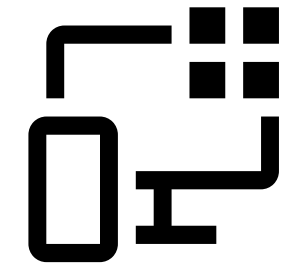


Traditional natural language processes are long, complicated, and require many resources



It is difficult to translate messy, vague natural language inputs into deterministic, specific questions and instructions that a machine can act on.

# Challenges



## Application

## Aa

## Language



## Database

# Application Challenges

## Building a suitable database

Compiling databases takes time, resources, and money

---

## Extracting information

Natural language is inconsistent and unreliable

---

## Formulating judgments and opinions

How can data determine subjectivity and ambiguity?

# Extracting information

Input: "Hey Watson, play my favorite song from last week and remind me to call Mom afterwards."

- 3 verbs: play [music], remind me, call Mom
- 2 time-conditions: last week, afterwards [now]
- 1 opinion-based understanding: favorite

# Application Challenges

## Building a suitable database

Compiling databases takes time, resources, and money

---

## Extracting information

Natural language is inconsistent and unreliable

---

## Formulating judgments and opinions

How can data determine subjectivity and ambiguity?

# Formulating judgments and opinions

Input: "Hey Watson, bake the bread thoroughly and until crispy."

- Thoroughly & crispy could mean burnt, overcooked, or golden

Input: "Hey Watson, which applicants can program in Python? And which applicants specialized in Python?"

- "program" vs "specialize"

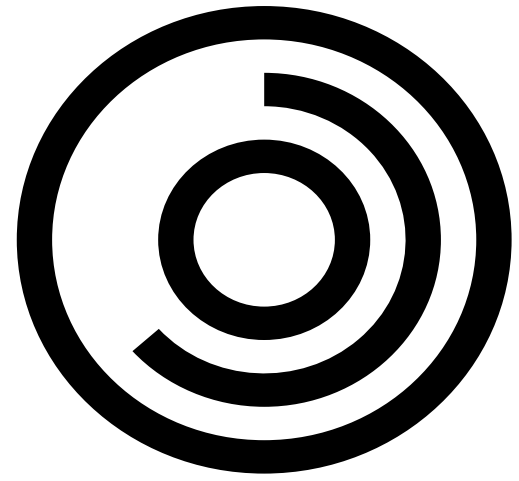
# Language Challenges

## Ambiguity

Natural language relies heavily on context-specific knowledge and cues

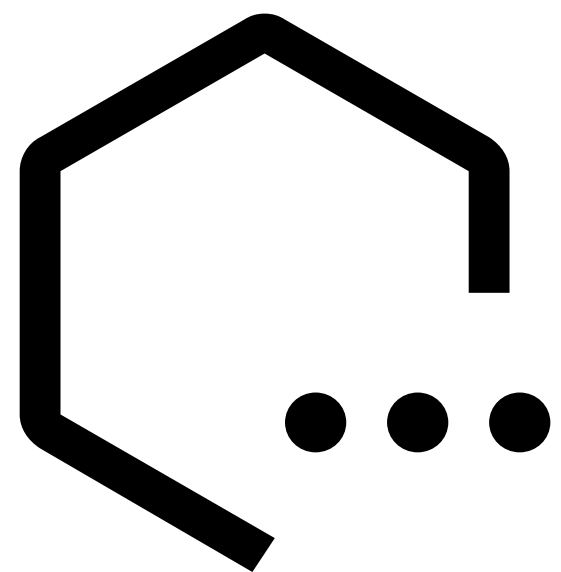
## Input formula

Unrealistic and unreliable to push users towards an unnatural way of speaking



# Anaphora

The use of a word referring to or replacing a word used earlier in a sentence



---

# Ellipsis

The omission from speech or writing of a word or words that are superfluous or able to be understood from contextual clues



# Anaphora & Ellipsis

1

Which applicants  
know Python and  
C?

2

Which of **them**  
also live in  
California?

3

How many [...] **also**  
know Java?

# Yes/No

## Is John Smith being interviewed?

1

No, John Smith  
has applied but  
has not yet  
been interviewed

2

No, John Smith  
has applied but  
was not selected  
for an interview

3

No, John Smith  
has applied and  
was already  
interviewed

# Database Challenges

Most databases are not optimized for NLP

Simple sentences turn into complex, nested queries

---

Not all databases have an interactive query language

Results in slow retrieval and specialized syntax

---

Databases are not updated quick enough

Depending on the industry, information accuracy could be critical

What companies in Canada shipped fur to England in the 1700s?

1. What Canadian companies existed in the 1700s?
  - a. What Canadian companies shipped fur in 1700s?
2. What English companies received fur in the 1700s?
  - a. What English companies received fur from Canadian companies in 1700s?

# Database Challenges

Most databases are not optimized for NLP

Simple sentences turn into complex, nested queries

---

Not all databases have an interactive query language

Results in slow retrieval and specialized syntax

---

Databases are not updated quick enough

Depending on the industry, information accuracy could be critical