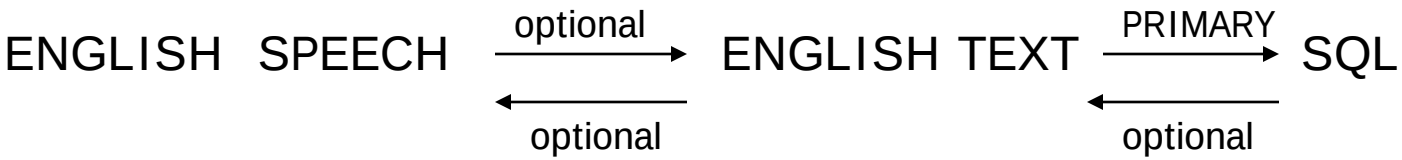
Natural Language Interfaces (NLI's)

- Mapping from (free-form) English text (or speech) to SQL
- English text or speech response to query rather than responding with a table



Natural Language Interfaces (Example)

Which females with PhD's in Department27 earn over $50,000?

SELECT Fname, Lname

FROM Employee, Department

WHERE Gender = 'F'

AND Degree = 'PhD'

AND DNO = 27

AND Salary > 50000

Natural Language Interfaces - History

- 1970's and 1980's
  - A flurry of interest in commercial systems

    (one of the first applications of NLP in the marketplace)

    PROBLEM: Difficult to adapt to new domains

    (especially by novice users)

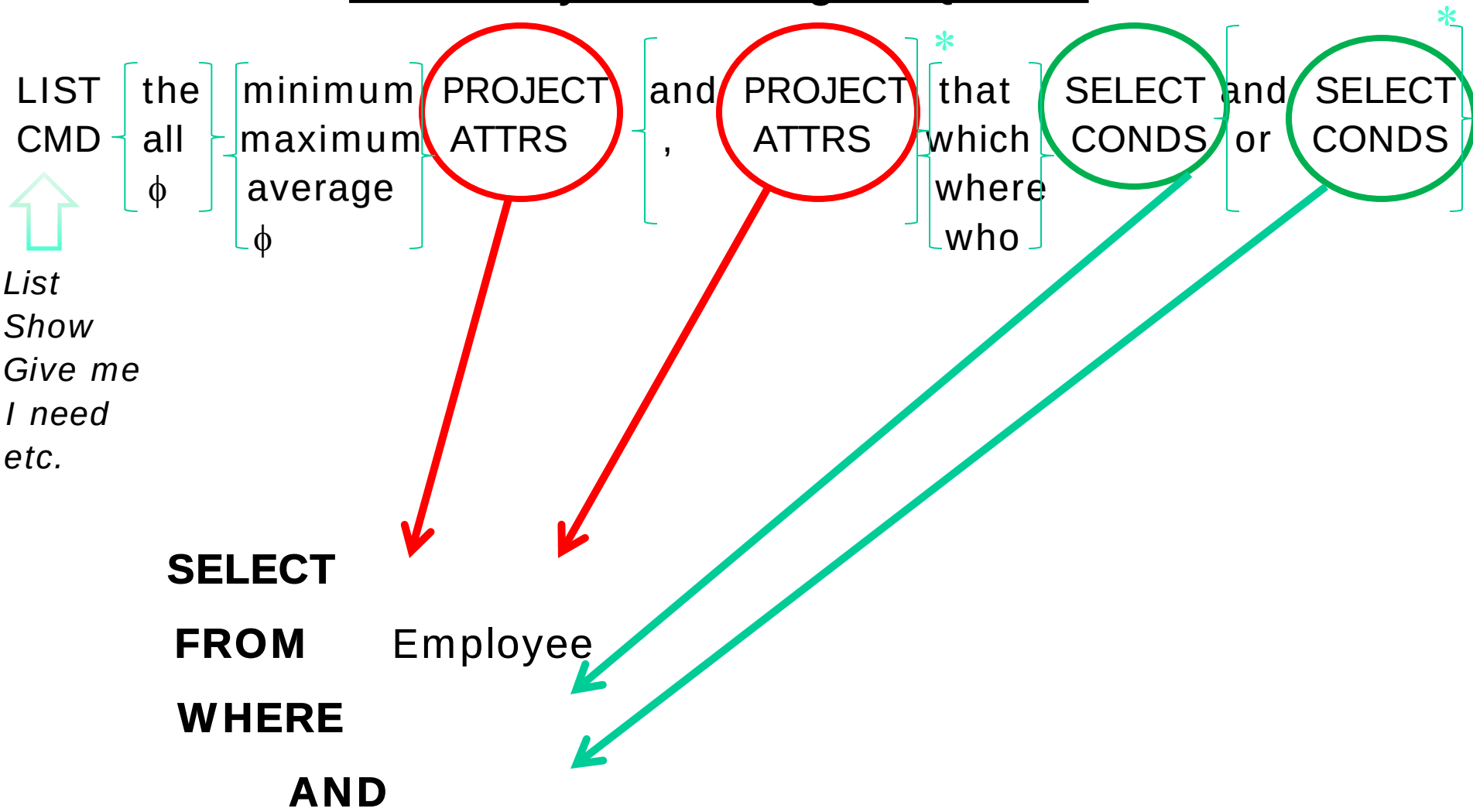
- 1988-2000+
  - Speech-based natural language database interfaces (ATIS)

Natural Language Interfaces - History

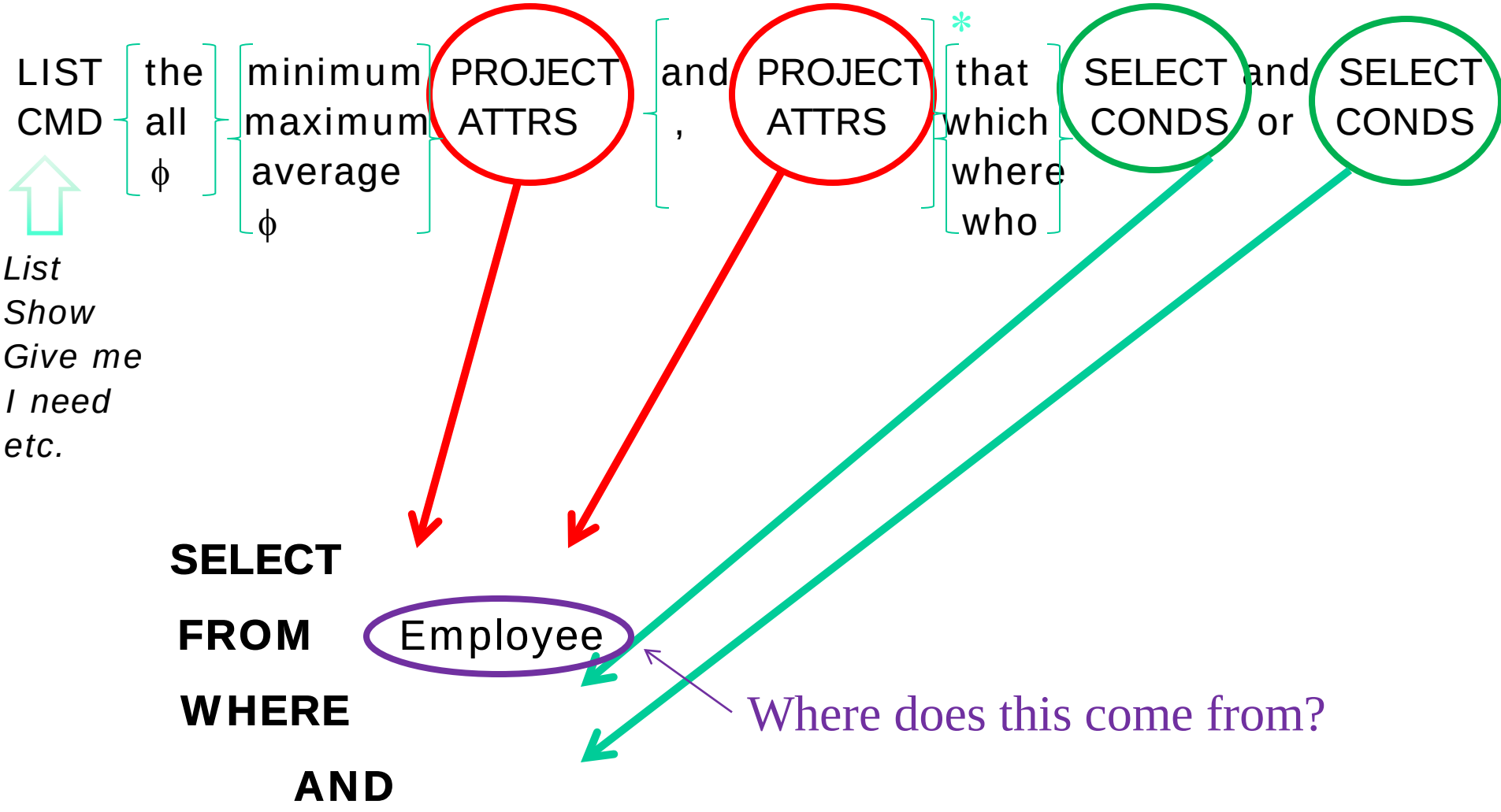
- HAL – Lotus
- NLI Inc. – Data talker
- NLQ (Battelle) – Natural Language Query
- Intelligent Query

==> WIDE USE WILL DEPEND ON SPEECH RECOGNITION TECHNOLOGY

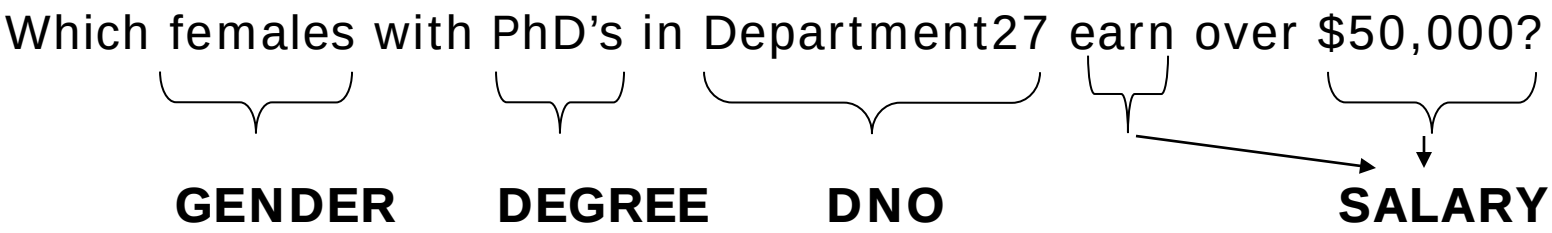
General Syntax of English Queries



General Syntax of English Queries



Mapping from Words to Attributes



SELECT Fname, Lname

FROM Employee, Department

WHERE Gender = 'F'

AND Degree = 'PhD'

AND DNO = 27

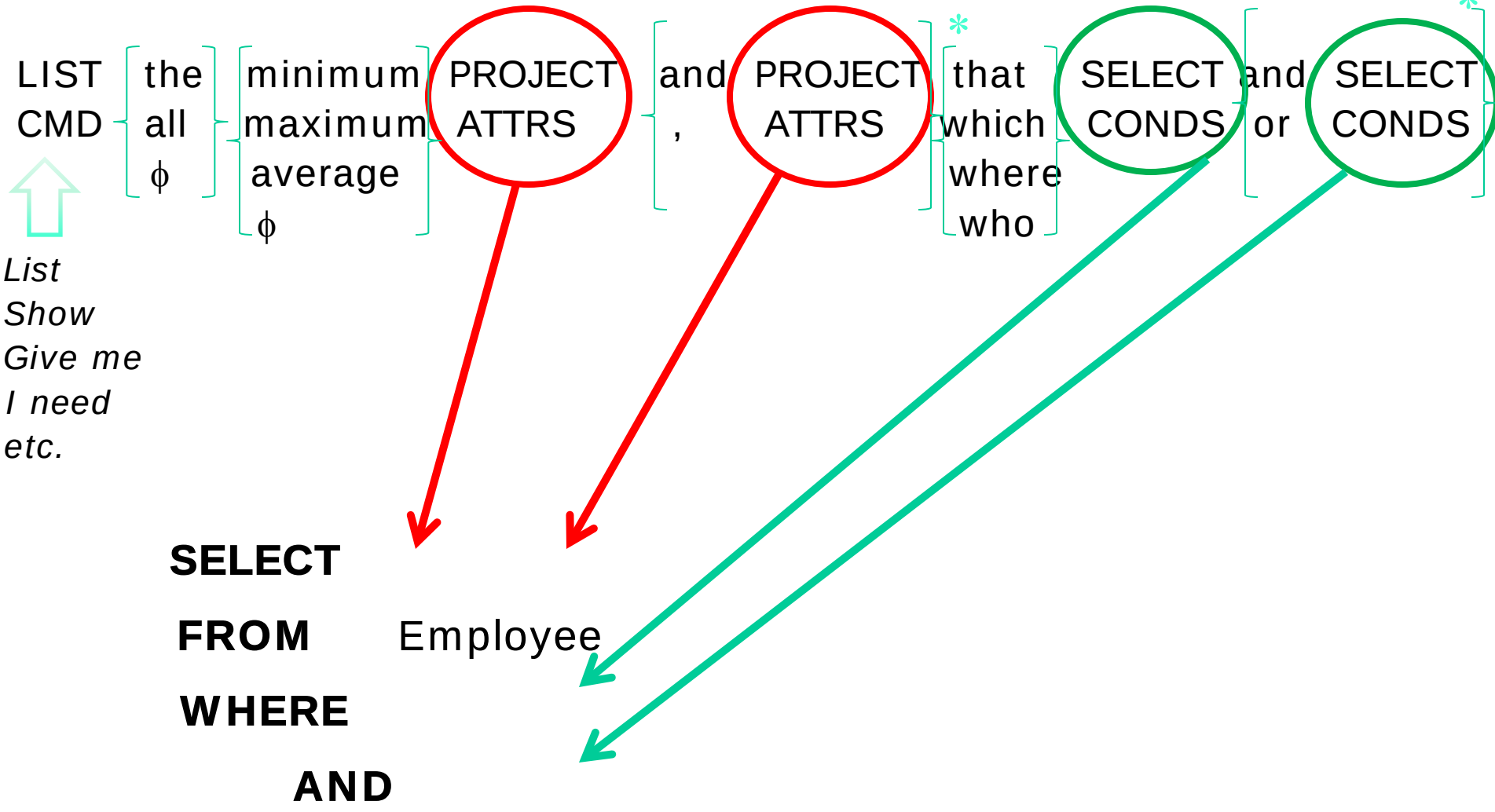
AND Salary > 50000

Selecting Attribute Names from Seen Values

ATTRIBUTE NAME	SEEN VALUES
Department	Accounting Sales Research
Lname	Smith Jones Madison
Locations	Skaggsville Boston Madison

NAMED ENTITY CLASSIFICATION: Proper name->Type

General Syntax of English Queries



We can build complex finite-state English grammars and maps to SQL.

Range Qualifiers

(V)	**<QUAL>**	**<X>**
earns	*more than*	*50,000*
is	*over*	*30 years old*

ENGLISH	**SQL**
more than	>
at least	>=
over	>
less than	<
under	<
below	<
not more than	<=
<null>	=

Issues – Attachment Ambiguity

- Find all female employees and managers …

Issues – Attachment Ambiguity

- Find all (female employees) and managers …

Issues – Attachment Ambiguity

- Find all female (employees and managers)…

Conjunction Ambiguity

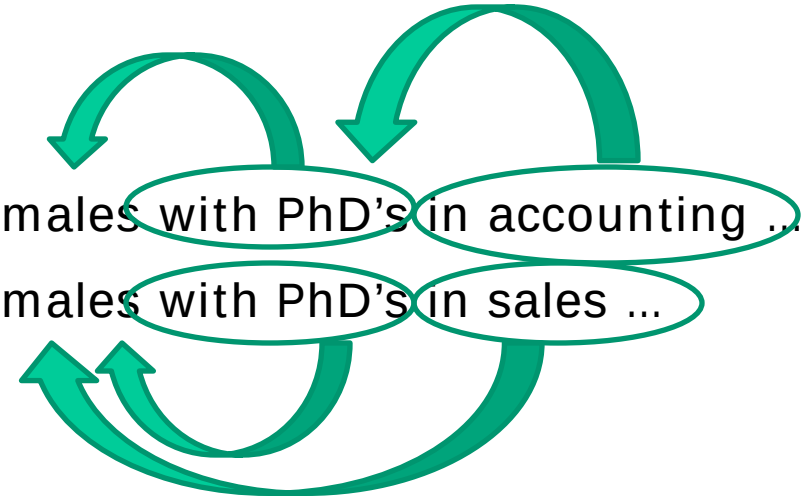
Issues – Attachment Ambiguity

- Find all female employees and managers …

- Which females with PhD's in accounting …
- Which females with PhD's in sales …

Issues – Attachment Ambiguity

- Find all female employees and managers …



- Which females with PhD's in accounting …
- Which females with PhD's in sales …

Issues Continued – What to Output?

Which females with PhD's working in accounting earn over $50,000?

Issues Continued – What to Output?

Which females with PhD's working in accounting earn over $50,000?

- Default attributes to return for each target entity:

  Employee → Lname, Fname

  Department -> Dname (DNO)

- **Plus** attribute values computed in query and not specified directly

  e.g. Salary

Problems with NLI's

- People don't know what they can ask for

  unless (and possibly if) they are expert users

- Menus (or menu dialog) may be more helpful

Key Opportunity for NLI's - SPEECH

- If you have a computer in front of you, menus or point-and-click are often easier/more efficient than typing a full sentence

  (e.g. * Show me all earthquakes in Indonesia from 1950 to 1960 that were over 7.0 on the Richter scale

  * List all stocks in my portfolio which were down more than 20% in today's trading.

  vs. a QBE or fill-in-the-blanks interface

- If you are in the car on a cellphone, **without** text/GUI-based input or output then verbal questions and answers are much more natural

ATIS (air travel) Speech Interface Example

- Please tell me all the flights that leave for Boston today.

```
SELECT   FNO
FROM     FLIGHT
WHERE    DestinationCity = 'Boston'
         AND   Date = '11-19-2002'
         AND   OriginCity = 'Atlanta'
```

ATIS (air travel) Speech Interface Example

- Please tell me all the flights that leave for Boston today.

```
SELECT   FNO
FROM     FLIGHT
WHERE    DestinationCity = 'Boston'
         AND  Date = '11-19-2002'
         AND  OriginCity = 'Atlanta'
```

Dialog/Context-based assumed information

e.g. OriginCity or today's date

Speech-based NLI's

- **Open issue is still what to display / answer?**
  - Please tell me all the flights that leave for Boston today.
    - ** just the flight numbers (literal answer to their question)?*
    - ** the flight departure times also?*
    - ** flight departure and arrival times?*
    - ** the lowest fare(s)?*

- **computer-driven NLI dialogues**
  - asking users targeted questions based on likely needs
  - more constrained knowledge of what the user is likely to say (i.e. the answer domain) much improves speech recognition accuracy

    (i.e. if system is expecting a city name, time or airline the entropy of the input is much reduced)