Dialogue Modelling

Raquel Fernández
Institute for Logic, Language and Computation (ILLC)

NLP1 guest lecture, December 2022

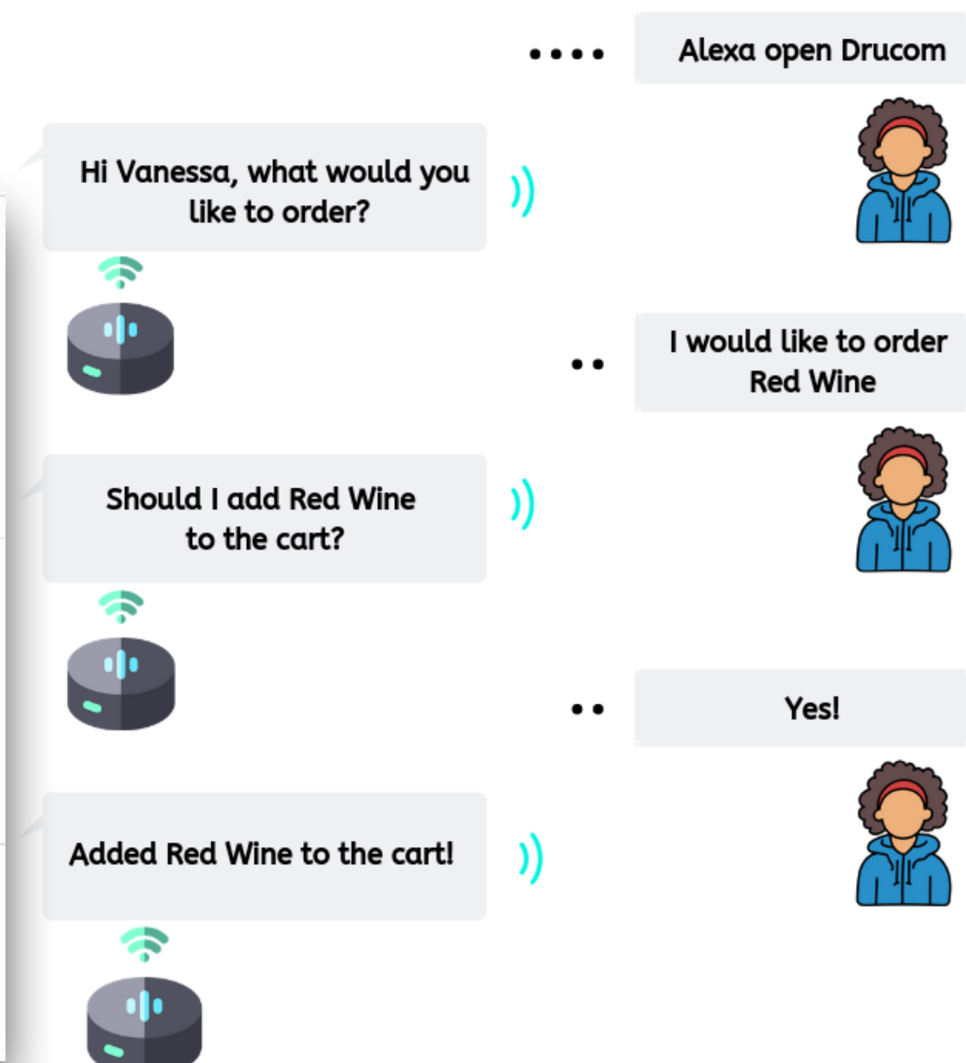
Plan for today

- What is dialogue modelling?
- Current NLP methods to model dialogue systems / chatbots
- Three examples of recent research done by my group

Dialogue

What is it and why do we care

- Using language for cross-speaker communication and interaction
- Primary form of language use and language learning



Dialogue

What is it and why do we care

It is convenient to distinguish between

- Social chit-chat dialogue
- Task-oriented dialogue

A: What's your favorite holiday?
B: I'm a big fan of Christmas.
A: Is that so? Mine is Holloween.
B: I also like Holloween. But I like Christmas most.

PC: Alexa, open plan my trip.
ALEXA: Where are you planning to go?
PC: I'm going to Portland.
ALEXA: What city are you leaving from?
PC: Seattle.
ALEXA: What date are you flying out?
PC: Next Thursday.
ALEXA: This will be fun. You go from Seattle to Portland on April 27th, 2017.

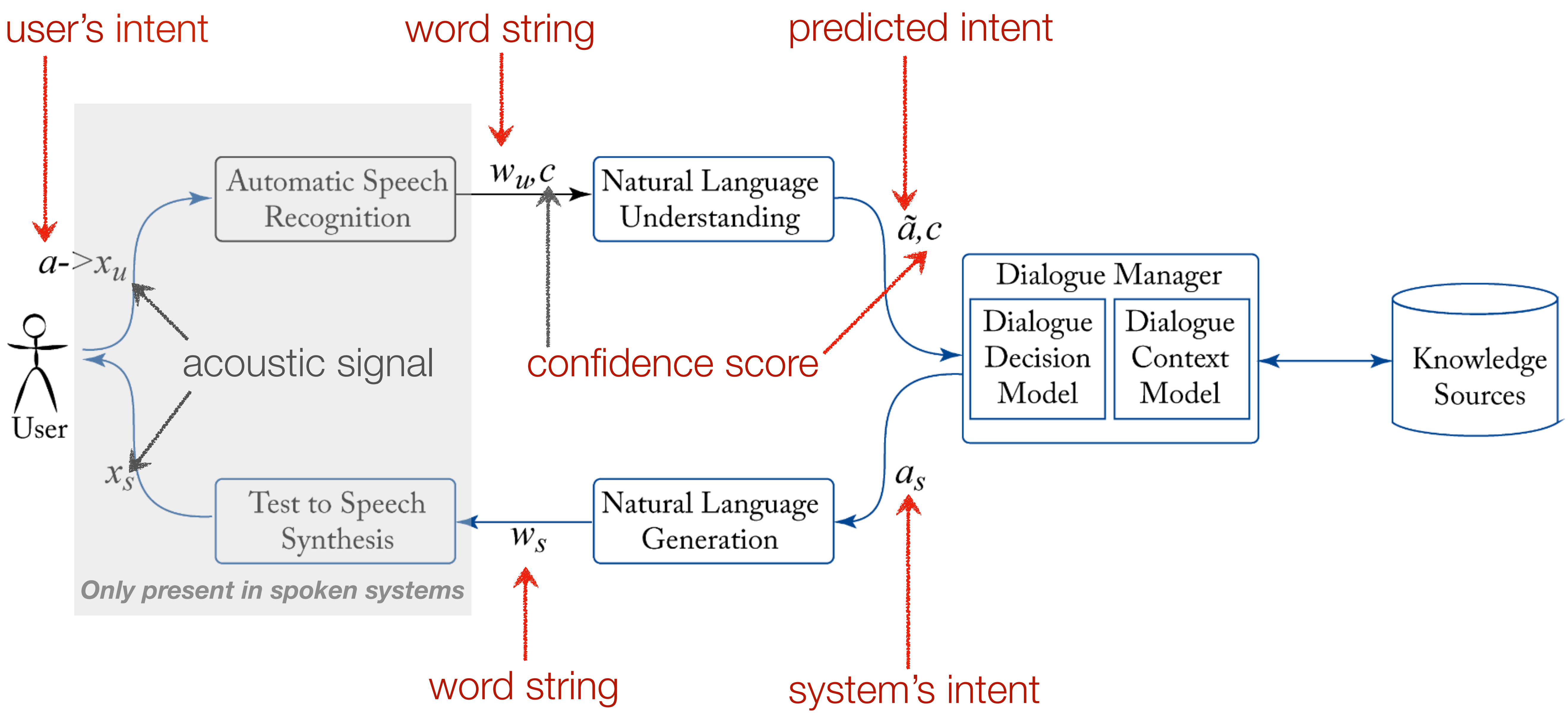
Dialogue modelling

Modelling a dialogue agent involves:

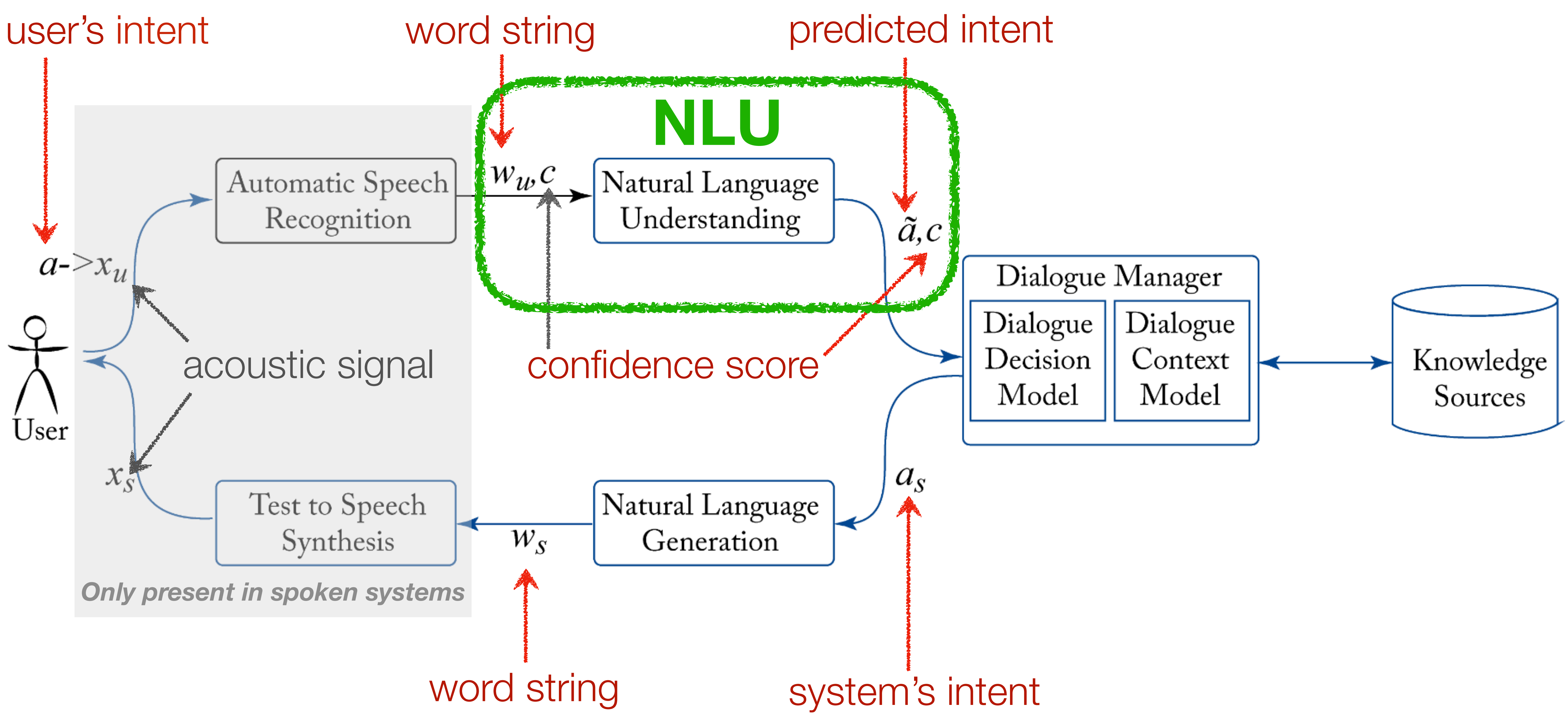
- ☑ Understanding the utterances by the dialogue partner.
- ☑ Keeping track of the dialogue history.
- ☑ Deciding what to say.
- ☑ Generating an utterance that conveys the speaker’s intend.

A dialogue agent (McTear, 2020)

- Task-oriented dialogue agents are typically modelled using a **modular architecture**, with modules for the steps above



A dialogue agent (McTear, 2020)



NLU

Intent prediction: Why is it difficult?

Speech act or **dialogue act**: the function of (or the action performed by) an utterance. The intention of the speaker.

- *statement, question, answer, agreement, request*, ….

- There isn't a one-to-one mapping between form and function (between the word string and the dialogue act)

  The gun is loaded. *Threat? Warning? Statement?*

- It may require inference (e.g., computing a "conversational implicature"):

  A: Are you going to Paul's party?
  B: I have to work.
  (=> I'm not going — *negative answer*)

NLU

Intent prediction: What is it in practice?

Predict a **meaning representation** given the word string.

In task-oriented dialogue, these are usually “frames” consisting of:

- Domain of the conversation (if not pre-defined)
- Each domain, has a set of possible user intents (task goals).
- Each intent, has a set of possible slots and slot values.

What are possible morning flights from Boston to SF on Tuesday?

```
DOMAIN:        AIR-TRAVEL
INTENT:        SHOW-FLIGHTS
ORIGIN-CITY:   Boston
ORIGIN-DATE:   Tuesday
ORIGIN-TIME:   morning
DEST-CITY:     San Francisco
```

Wake me tomorrow at six.

```
DOMAIN:   ALARM-CLOCK
INTENT:   SET-ALARM
TIME:     2017-07-01 0600-0800
```

NLU

Intent prediction: What is it in practice?

- Many of the NLP techniques you have seen in this course are relevant for intent prediction in dialogue:

  - word embeddings, POS tagging, syntactic parsing, compositional semantics, etc.

- This approach requires **annotated dialogue datasets** where utterances are annotated with meaning representations.

What are possible morning flights from Boston to SF on Tuesday?

```
DOMAIN:        AIR-TRAVEL
INTENT:        SHOW-FLIGHTS
ORIGIN-CITY:   Boston
ORIGIN-DATE:   Tuesday
ORIGIN-TIME:   morning
DEST-CITY:     San Francisco
```

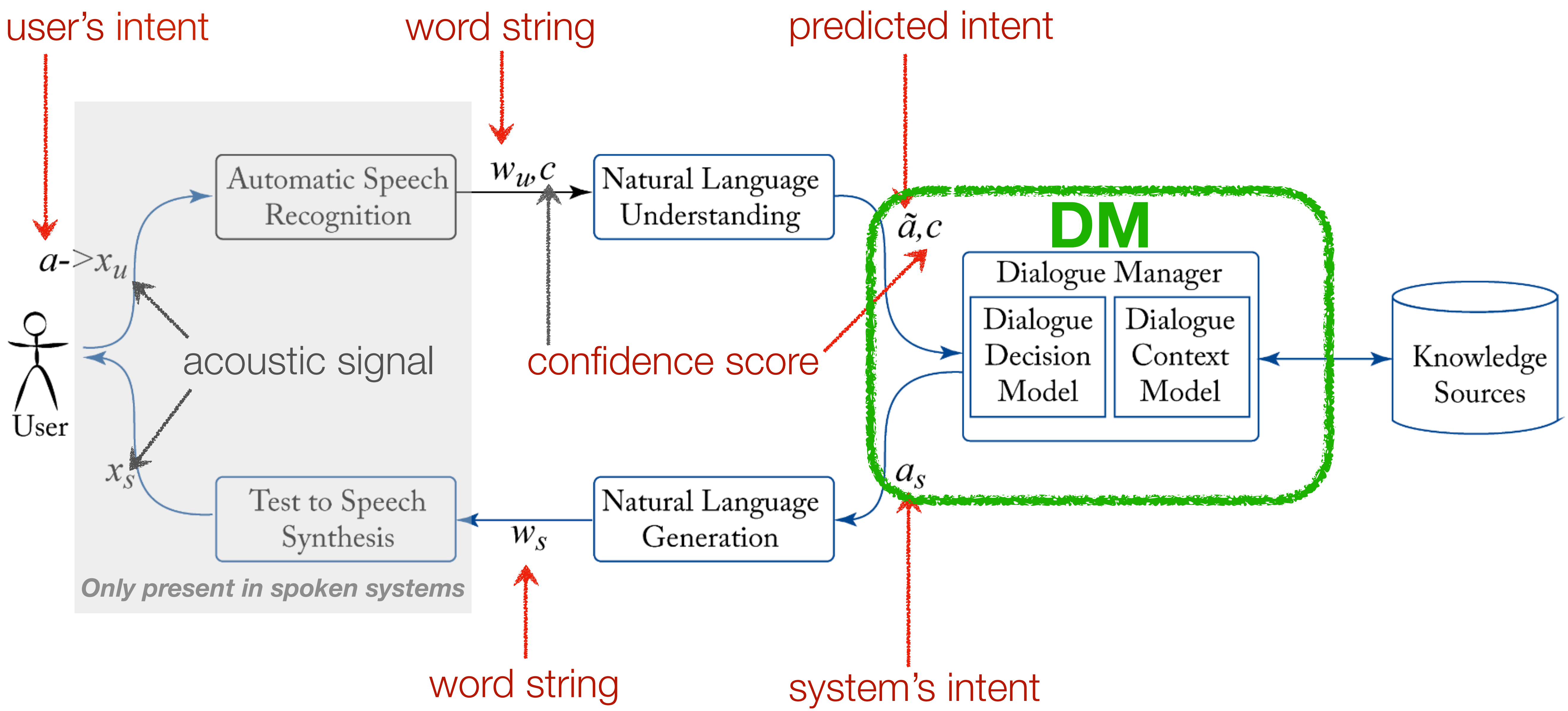
Wake me tomorrow at six.

```
DOMAIN:   ALARM-CLOCK
INTENT:   SET-ALARM
TIME:     2017-07-01 0600-0800
```

Some resources

- https://parl.ai/docs/tasks.html
- https://breakend.github.io/DialogDatasets/references.html
- https://docs.google.com/spreadsheets/d/1N5_5gBKlGR-OrigRNct4jQ6iEqSycyqcoN61JpsHFDQ/htmlview

A dialogue agent (McTear, 2020)



Dialogue management

- The relevant slots may be filled across multiple dialogue turns—the **dialogue context / history** keeps track of this information.
- The **dialogue decision model / policy:** predict the next system action given dialogue context (e.g., slots that are still missing).
  - System intent with the highest probability given the context.

U: Show me morning flights to SF.

```
DOMAIN:        AIR-TRAVEL
INTENT:        SHOW-FLIGHTS
ORIGIN-CITY:   [     ]
ORIGIN-DATE:   [     ]
ORIGIN-TIME:   morning
DEST-CITY:     San Francisco
```

```
DOMAIN:        AIR-TRAVEL
INTENT:        REQUEST(ORIGIN-CITY)
```

S: Where are you flying from?

Dialogue management

Confirmation and rejection

- How likely is the system to have understood the user?
- We can exploit NLU confidence scores to decide on a confirmation/rejection policy:

$< \alpha$	low confidence	reject
$\geq \alpha$	above the threshold	confirm explicitly
$\geq \beta$	high confidence	confirm implictly
$\geq \gamma$	very high confidence	don't confirm at all

`CONFIRM_EXPLICIT(ORIGIN-CITY)`

S: Which city do you want to leave from?
U: Baltimore.
S: **Do you want to leave from Baltimore?**
U: Yes.

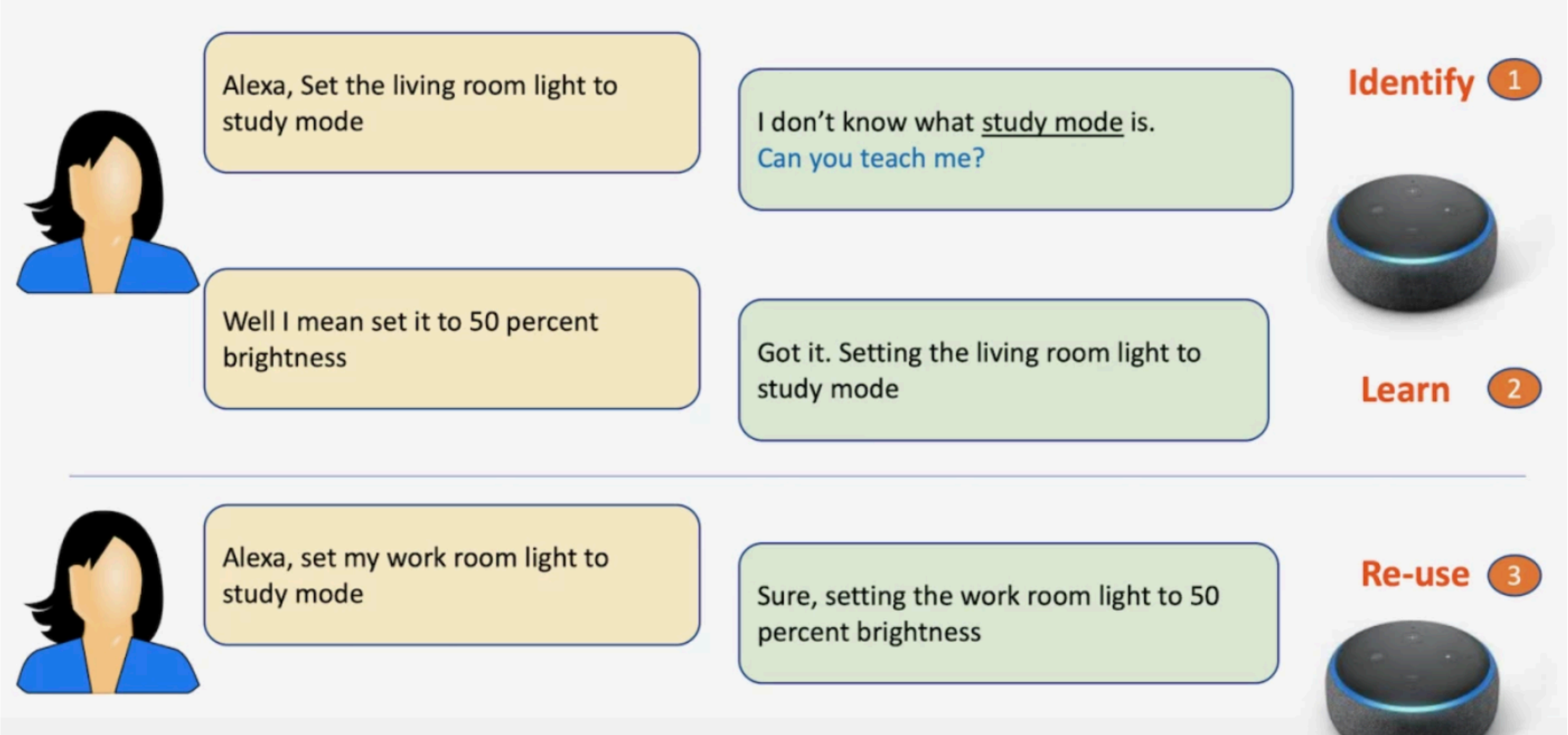
`CONFIRM_IMPLICIT(DEST-CITY)`

U: I want to travel to Berlin
S: **When do you want to travel to Berlin?**

Dialogue management

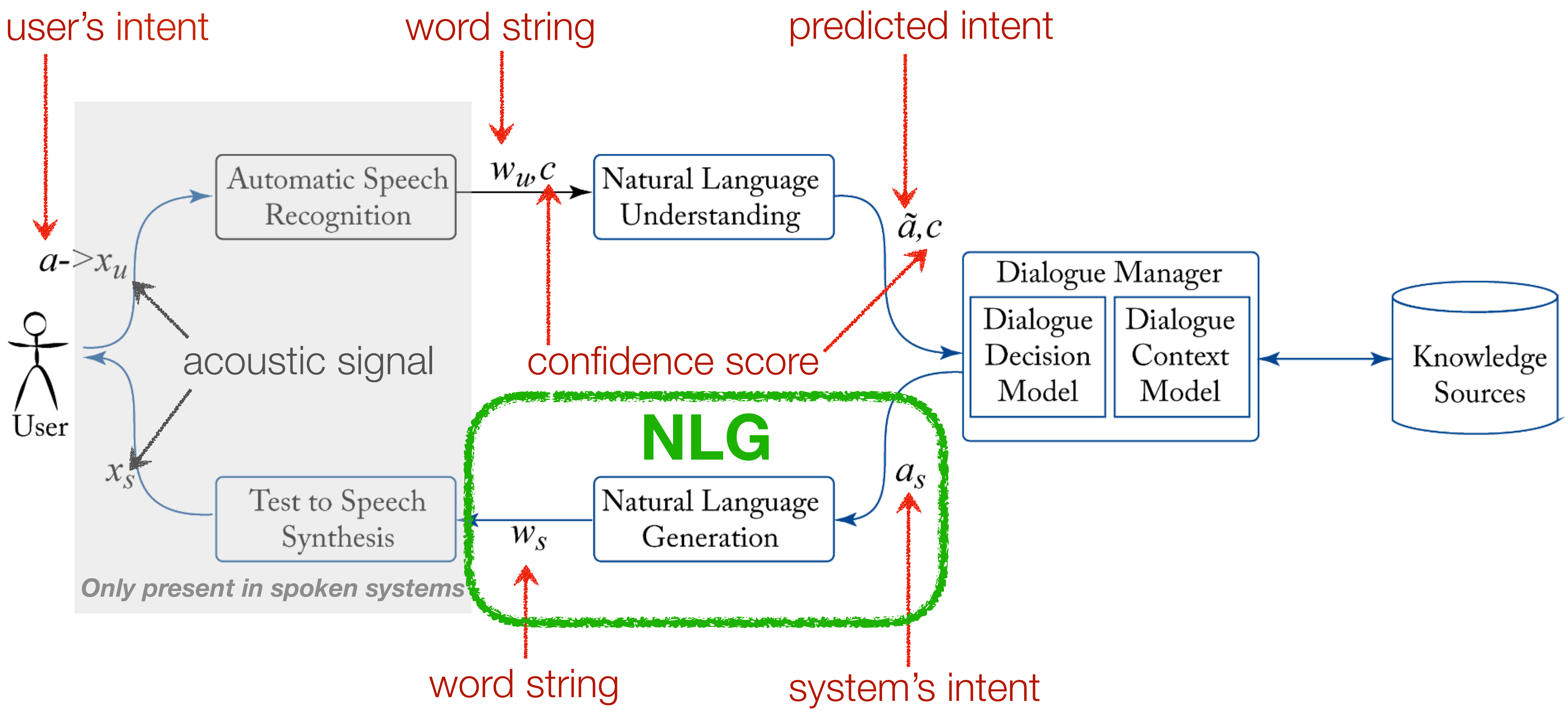
Advanced: Learning and generalisation

- Confidence scores can also be exploited to identify unknown slots and learn to generalise to new situations



https://www.amazon.science/blog/new-alexa-features-interactive-teaching-by-customers

A dialogue agent (McTear, 2020)



NLG

Assuming the DM has chosen a next system action/intent…

- The goal of the NLG module is to learn to generate sentences by training on many representation/sentence pairs from an annotated dialogue corpus

- Some examples:

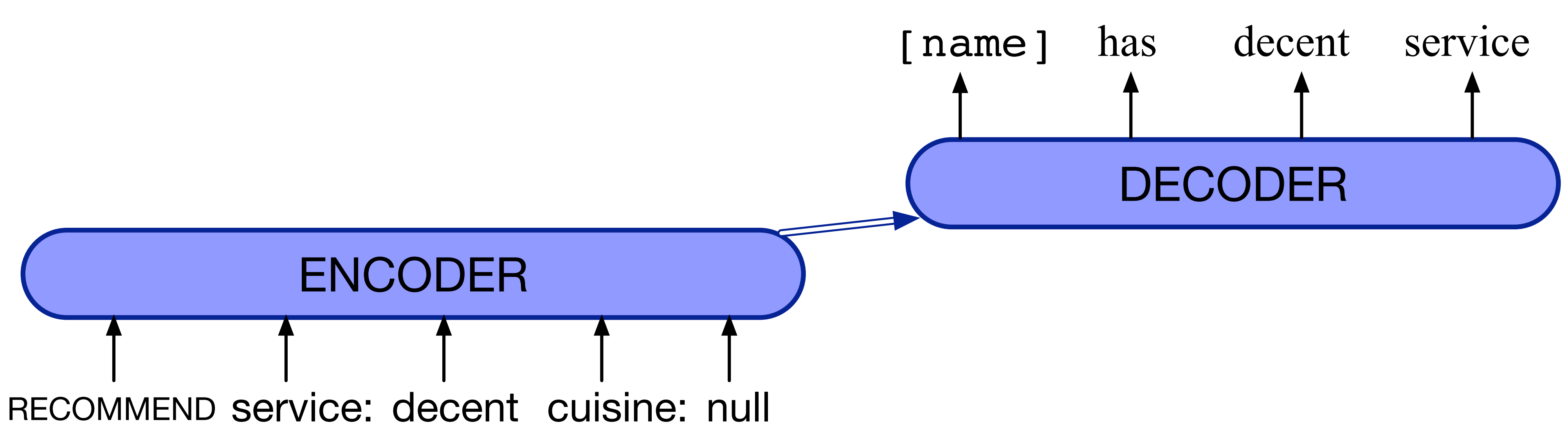
	`recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french`
1	Au Midi is in Midtown and serves French food.
2	There is a French restaurant in Midtown called Au Midi.
	`recommend(restaurant name= Loch Fyne, neighborhood = city centre, cuisine = seafood)`
3	Loch Fyne is in the City Center and serves seafood food.
4	There is a seafood restaurant in the City Centre called Loch Fyne.

NLG

Sequence-to-sequence prediction (cf. previous lecture):

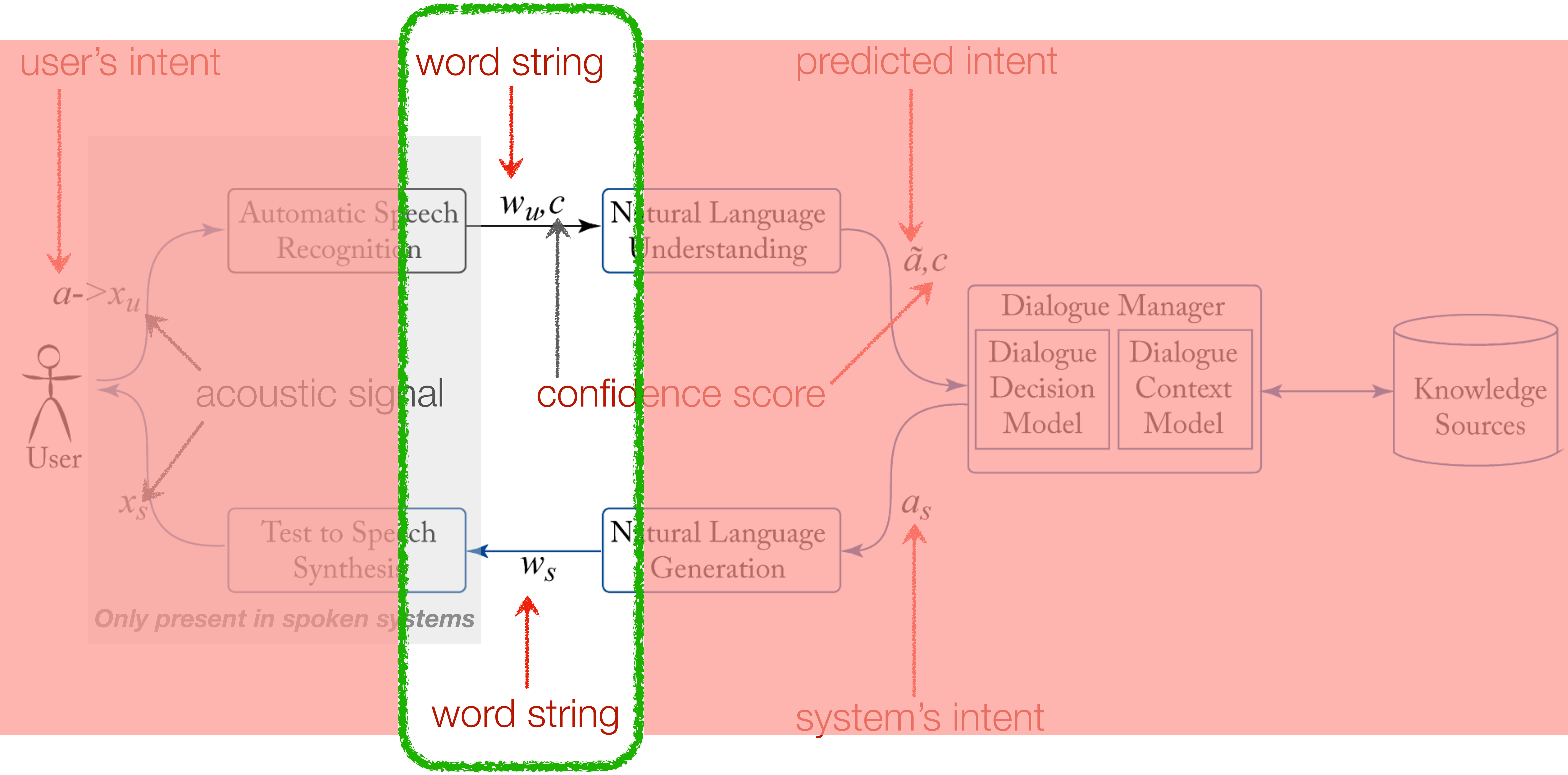
- Input: linearised meaning representation
- Output: word string (system utterance)

	`recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french`
1	Au Midi is in Midtown and serves French food.
2	There is a French restaurant in Midtown called Au Midi.
	`recommend(restaurant name= Loch Fyne, neighborhood = city centre, cuisine = seafood)`
3	Loch Fyne is in the City Center and serves seafood food.
4	There is a seafood restaurant in the City Centre called Loch Fyne.



(NB: Delexicalised representation where entities are replaced with general placeholders to help with generalisation)

Non-modular systems



Non-modular systems

Chatbots

- Dialogue response generation from previous turn(s), without intermediate meaning representations.
- Typically used to model social **chit-chat dialogue** (no need to make progress towards task completion)
- Two methods: Retrieval vs generation

A: What's your favorite holiday?
B: I'm a big fan of Christmas.
A: Is that so? Mine is Holloween.
B: I also like Holloween. But I like Christmas most.

Non-modular systems

Retrieval

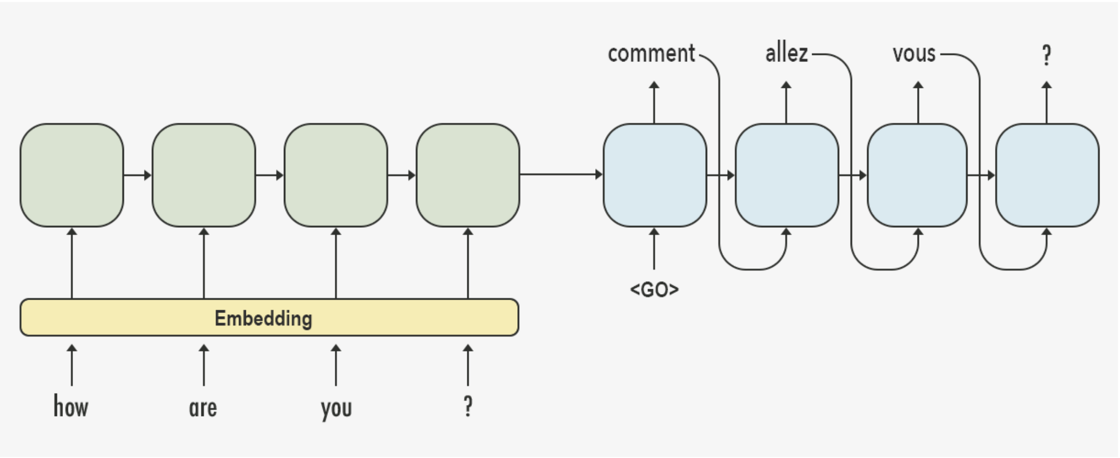
- Given a user turn q and a dialogue corpus C
- Find in C a turn p that is most similar to q
- Retrieve the turn r following p in C
- Use r as a response to q

Non-modular systems

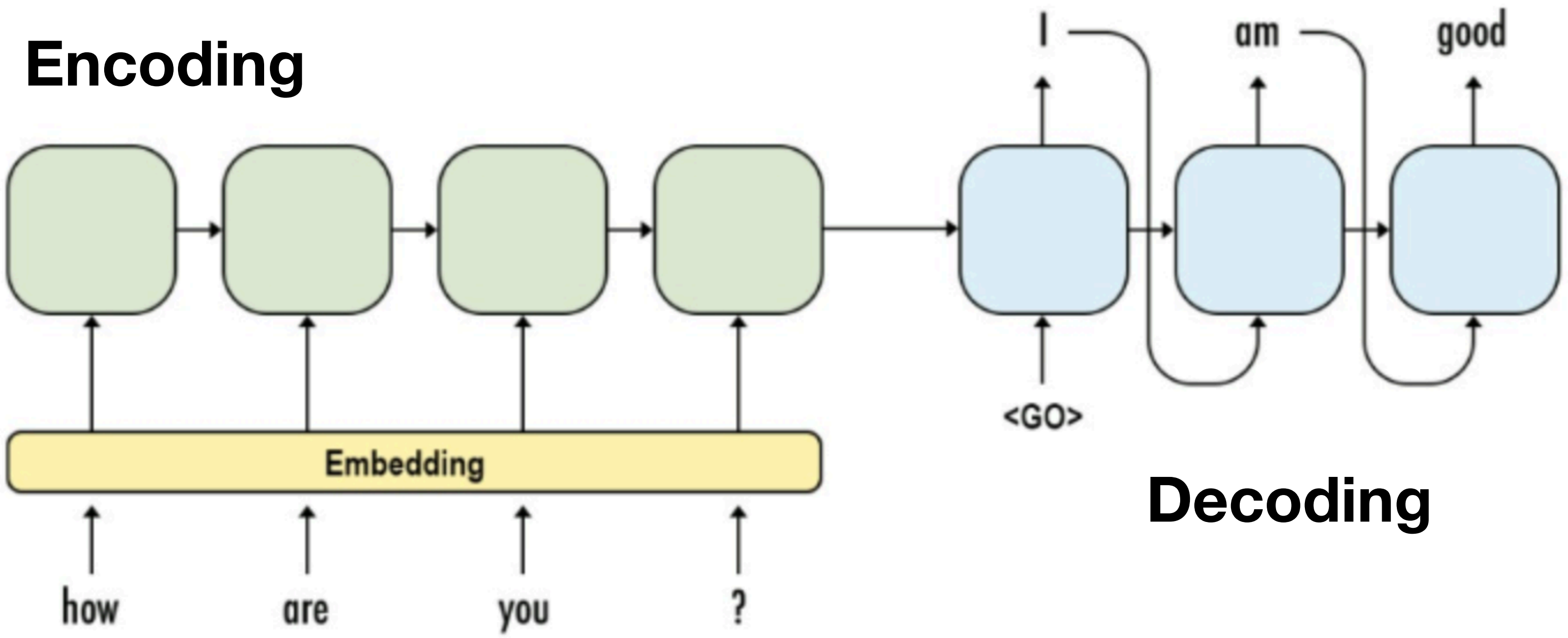
Generation

- **Sequence-to-sequence** models:

  - Inspired by machine translation



- Encoder RNN to produce a representation of the previous turns
- Decoder RNN to generate the response word-by-word by conditioning on the context and the response so far



Non-modular systems

Generation

An alternative to the encoder-decoder architecture:

- Use a large pre-trained **language model** (e.g., GPT-2)
- Fine-tune it on conversational data
- Use the language model directly as a response generator

Non-modular systems

PROS

- No annotations needed
- No finite, predefined set meaning representation

CONS

- Very data-hung: trained on dialogue corpora with hundreds of millions or words
- No real understanding
- Tendency to output generic utterances (uninformative, bland, repetitive)
- Simplistic encoding of the dialogue history, leading to semantic inconsistency

U:	tell me ... how long have you had this falling sickness ?
S:	i'm not a doctor.
U:	How long does it take you guys to learn the drill?
S:	Like 2 weeks ago!!
U:	Where do you live?
S:	I live in England.
U:	In which city do you live?
S:	I live in Los Angeles.

[adapted from Jurafsky & Martin, SLP-3]

Other important topics

… that we won’t be able cover today

- **Evaluation**
  - Very complex and difficult to operationalise
  - Easier for task-oriented dialogue (task completion)
- **Ethical considerations**
  - Implicit biases and reinforcement of stereotypes present in the training data
  - Deception: being perceived as human (anthropomorphism) may be problematic

Plan for today

- What is dialogue modelling?
- Current NLP methods to model dialogue systems / chatbots
- Three examples of recent research done by my group

Example 1

- Language style varies across sociolinguistic factors:

  - In this case, generation adapted to the age group of the user.

- Use of GPT-2 with “control” module trained age-annotated data.

- Extensive evaluation with humans judgements and discussion of ethical considerations.

GEM Workshop @EMNLP-2022

Controllable Text Generation *for All Ages*: Evaluating a Plug-and-Play Approach to Age-Adapted Dialogue

Lennert Jansen[†], Štěpán Lars Laichter[†], Arabella Sinclair[‡], Margot J. van der Goot[†], Raquel Fernández[†], Sandro Pezzelle[†]
[†]University of Amsterdam, [‡]University of Aberdeen
{lennertjansen95|lars.laichter}@gmail.com
arabella.sinclair@abdn.ac.uk
{m.j.vandergoot|raquel.fernandez|s.pezzelle}@uva.nl

Abstract

To be trusted and perceived as natural and coherent, conversational systems must adapt to the language of their users. While personalized dialogue is a promising direction, controlling generation for fine-grained language features remains a challenge in this approach. A recent line of research showed the effectiveness of leveraging pre-trained language models toward adapting to a text's topic or sen-

geographic location, etc. This is achieved by training systems with either implicit (Kottur et al., 2017; Li et al., 2016) or explicit (Qian et al., 2018; Zhang et al., 2018; Zheng et al., 2019) representations of a speaker. These approaches are generally shown to produce multi-turn conversations that are deemed of better quality by humans, but they pay little attention to understanding what factors determine human judgements. Recently, See et al. (2019) showed that linguistic aspects such as specificity

Example 2

- Many conversations involve more than the linguistic modality.
- Language + vision encoder-decoder architecture.
- Evaluation against linguistic properties of human dialogues.

NAACL-2019

Beyond Task Success:
A Closer Look at Jointly Learning to See, Ask, and GuessWhat

Ravi Shekhar[†], **Aashish Venkatesh**[*], **Tim Baumgärtner**[*], **Elia Bruni**[*],
Barbara Plank[♥], **Raffaella Bernardi**[†] and **Raquel Fernández**[*]
[†]University of Trento, [*]University of Amsterdam, [♥]IT University of Copenhagen
ravi.shekhar@unitn.it aashishv@outlook.com
baumgaertner.t@gmail.com elia.bruni@gmail.com bapl@itu.dk
raffaella.bernardi@unitn.it raquel.fernandez@uva.nl

Abstract

We propose a grounded dialogue state encoder which addresses a foundational issue on how to integrate visual grounding with dialogue system components. As a test-bed, we focus on the *GuessWhat?!* game, a two-player game where the goal is to identify an object in a complex visual scene by asking a sequence of yes/no questions. Our visually-grounded encoder leverages synergies between guessing and asking questions, as it is trained jointly

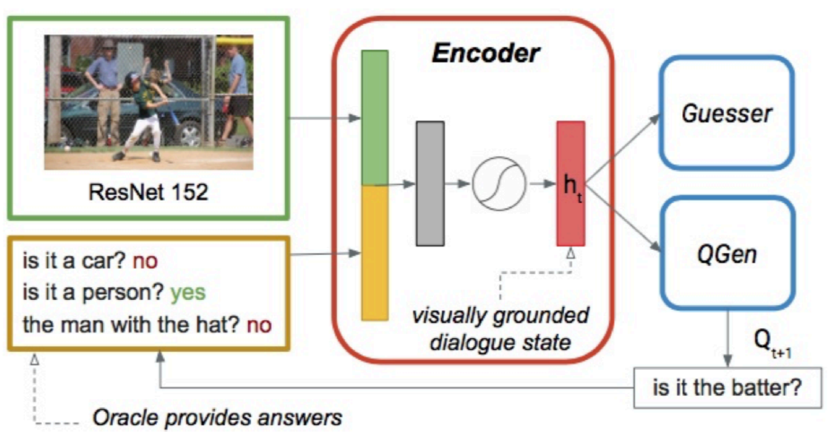


Figure 1: Our questioner model with a single visually grounded dialogue state encoder.

Example 3

- Use of NLP techniques to analyse human-human dialogue.
- Information theoretic perspective: estimate information content / processing effort with a large language model.
- Analyse patterns of information dynamics: interesting from a psycholinguistic point of view and informative for AI modelling.

EMNLP-2021

Is Information Density Uniform in Task-Oriented Dialogues?

Mario Giulianelli, Arabella Sinclair, Raquel Fernández
Institute for Logic, Language and Computation
University of Amsterdam
{m.giulianelli|a.j.sinclair|raquel.fernandez}@uva.nl

Abstract

The Uniform Information Density principle states that speakers plan their utterances to reduce fluctuations in the density of the information transmitted. In this paper, we test whether, and within which contextual units this principle holds in task-oriented dialogues. We show

fluctuations in the density of the information transmitted. Evidence for the principle of uniform information density (UID; Jaeger and Levy, 2007; Jaeger, 2010) has been found at many levels of language production: speakers tend to reduce the duration of more predictable sounds (Aylett and Turk, 2004, 2006; Bell et al., 2003; Demberg et al.,

For these and other papers

(on dialogue and beyond)

Dialogue Modelling Group @ UvA

https://dmg-illc.github.io/dmg/