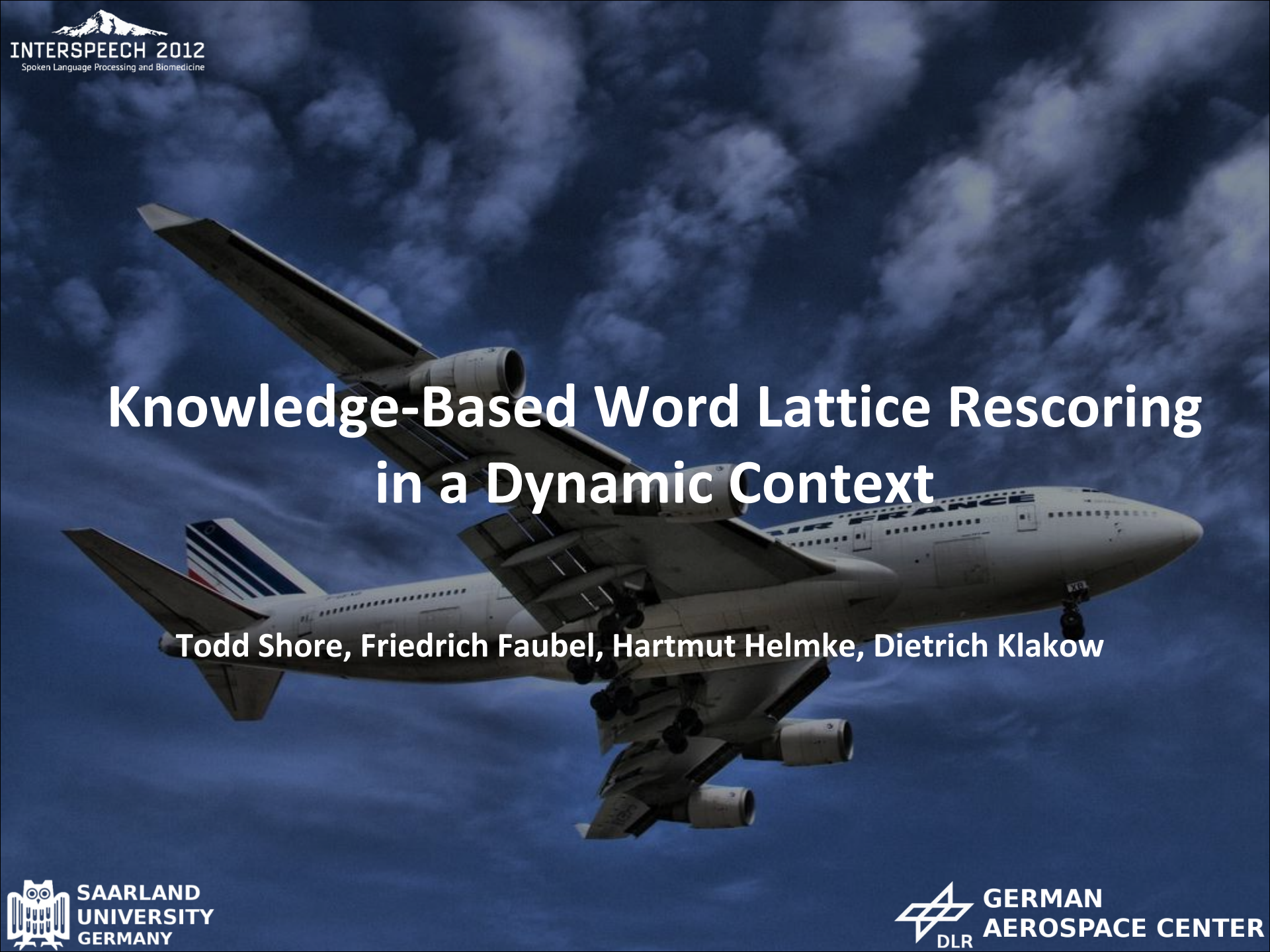


Knowledge-Based Word Lattice Rescoring in a Dynamic Context

Todd Shore, Friedrich Faubel, Hartmut Helmke, Dietrich Klakow

SAARLAND UNIVERSITY GERMANY

DLR GERMAN AEROSPACE CENTER

Section I

Motivation

Motivation

- **Problem:** difficult to incorporate higher-level knowledge sources into automatic speech recognition (ASR)

Motivation

- **Problem:** difficult to incorporate higher-level knowledge sources into automatic speech recognition (ASR)

- **However:** there are domains in which the situational context of utterances is available, e.g. air traffic control (ATC) or command and control tasks

Motivation

- **Problem:** difficult to incorporate higher-level knowledge sources into automatic speech recognition (ASR)

- **However:** there are domains in which the situational context of utterances is available, e.g. air traffic control (ATC) or command and control tasks

- **Approach taken in this work:** incorporate contextual knowledge into ASR by rescoring word lattice output of an ATC task

Section II

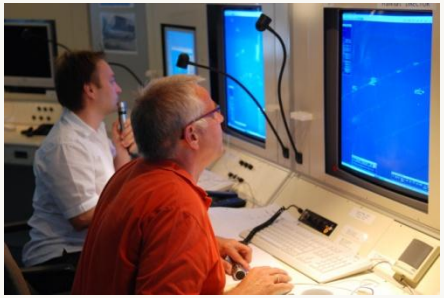
The Air Traffic Control Task

The Air Traffic Control Task

- Air traffic controllers at their workstations

The Air Traffic Control Task

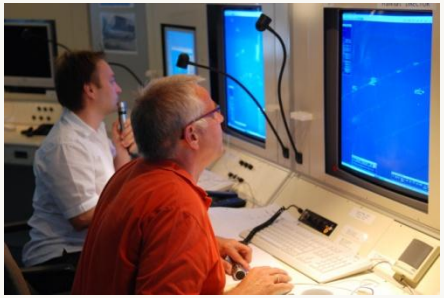
- Air traffic controllers at their workstations



- **Primary objective:**
  - maintain aircraft separation
  - safely guide approaching aircraft to their runway threshold
  - integrate departing and passing aircraft

The Air Traffic Control Task

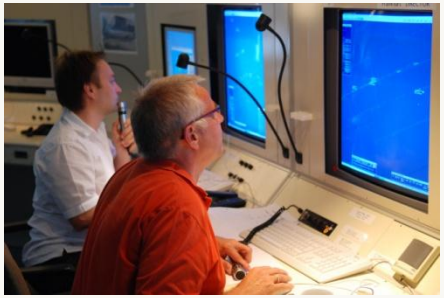
- Air traffic controllers at their workstations



- **Have access to:**
  - radar screens revealing aircraft positions and speeds
  - flight plans
  - weather reports, indicators for speed of wind, etc.

The Air Traffic Control Task

- Air traffic controllers at their workstations



- **Implementation:**
  - issue verbal commands to aircraft pilots
  - use of a standardized subset of English which is formally specified by the International Civil Aviation Organization (ICAO)

The Air Traffic Control Task

- **ICAO Phraseology** for ATC commands comprises
  - single aircraft callsign (identifying the aircraft)
  - goal actions to execute
  - goal values to be achieved

The Air Traffic Control Task

- **ICAO Phraseology** for ATC commands comprises
  - single aircraft callsign (identifying the aircraft)
  - goal actions to execute
  - goal values to be achieved

Type	Values	Example
DESCENT	ALT	*descend altitude* ALT *feet*
DESCENT	FL	*descend flight level* FL
REDUCE	SPD	*reduce speed* SPD *knots*
TURN	DIR, HDG	*turn* DIR *heading* HDG

The Air Traffic Control Task

- **ICAO Phraseology** for ATC commands comprises
  - single aircraft callsign (identifying the aircraft)
  - goal actions to execute
  - goal values to be achieved

- **Example:** *„Delta four three niner turn right heading two two zero“*

The Air Traffic Control Task

- **ICAO Phraseology** for ATC commands comprises
  - single aircraft callsign (identifying the aircraft)
  - goal actions to execute
  - goal values to be achieved

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **Relevant Information:**
  - DL 439
  - TURN, DIR=right, HDG=220

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **Use of a recognition grammar**
  - is sufficient for recognizing ICAO phraseology
  - simplifies extraction of semantic content

- **Example:** *„Delta four three niner turn right heading two two zero“*

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **Main idea of this work:** rescore recognized utterances through use of context knowledge about the situation in the airspace

The Air Traffic Control Task

- **Main idea of this work:** rescore recognized utterances through use of context knowledge about the situation in the airspace

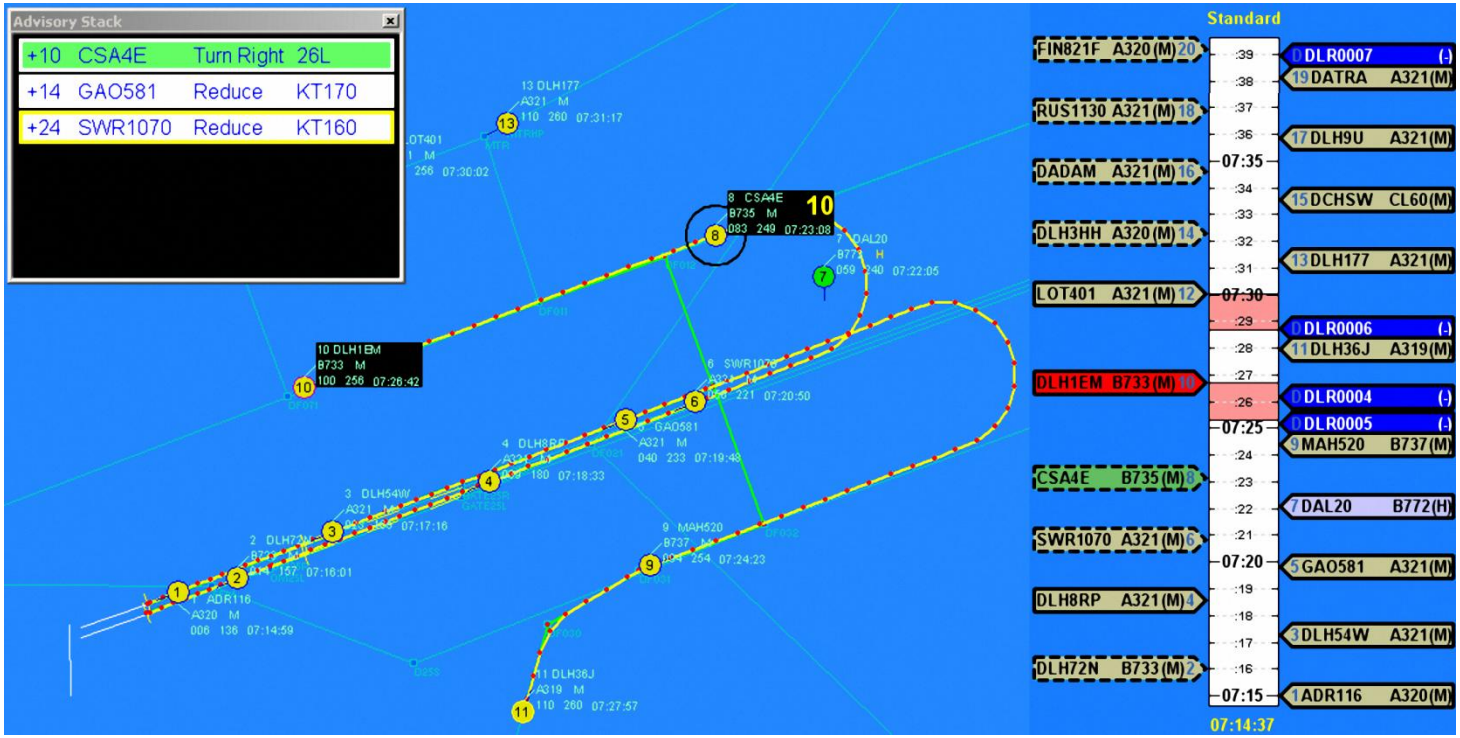
- But where does this information come from?

The Air Traffic Control Task

- Modernized ATC workspace as envisioned by the German Aerospace Center (DLR)

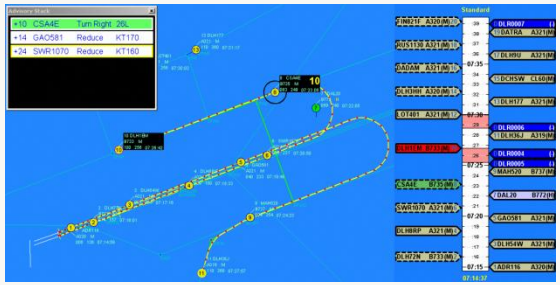
The Air Traffic Control Task

- … including the arrival manager **4D-CARMA**, which assists controllers in managing aircraft arrivals



The Air Traffic Control Task

- … including the arrival manager **4D-CARMA**, which assists controllers in managing aircraft arrivals



- **This system allows us to extract:**
  - the callsigns of aircraft in the airspace
  - the aircraft positions relative to the radar
  - Their speeds, altitudes, climb/descend rates, reduce rates, etc.

Section III

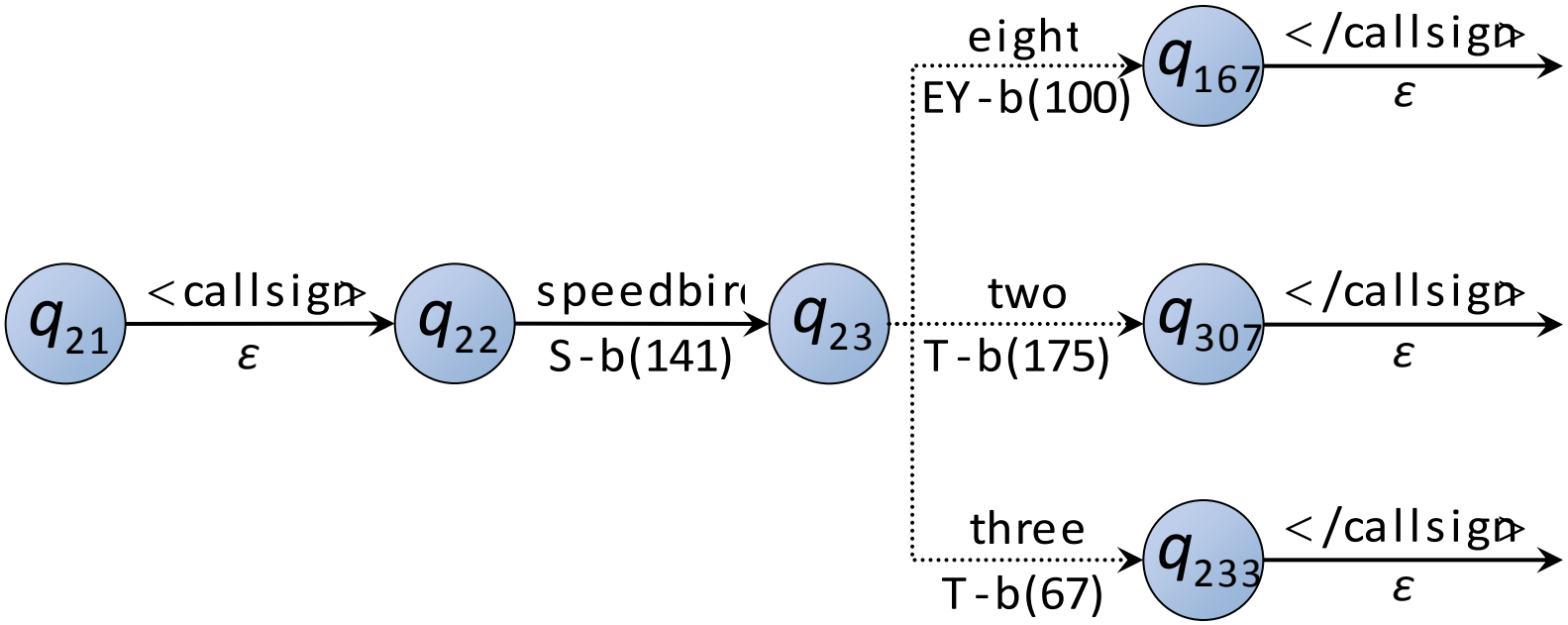
Knowledge-Based Lattice Rescoring

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search



Knowledge-Based Lattice Rescoring

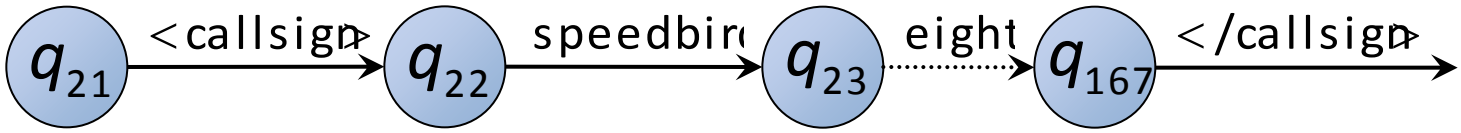
- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

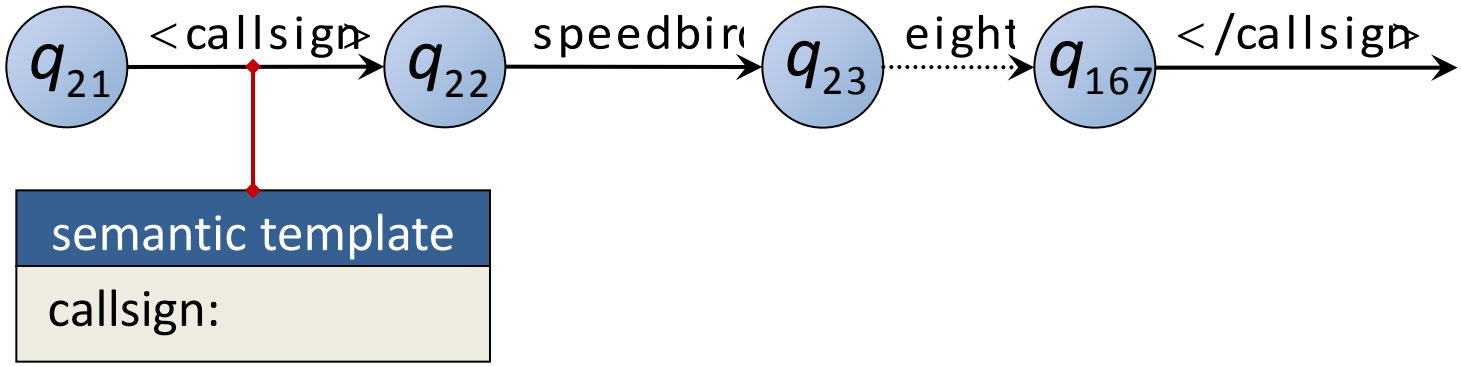
- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

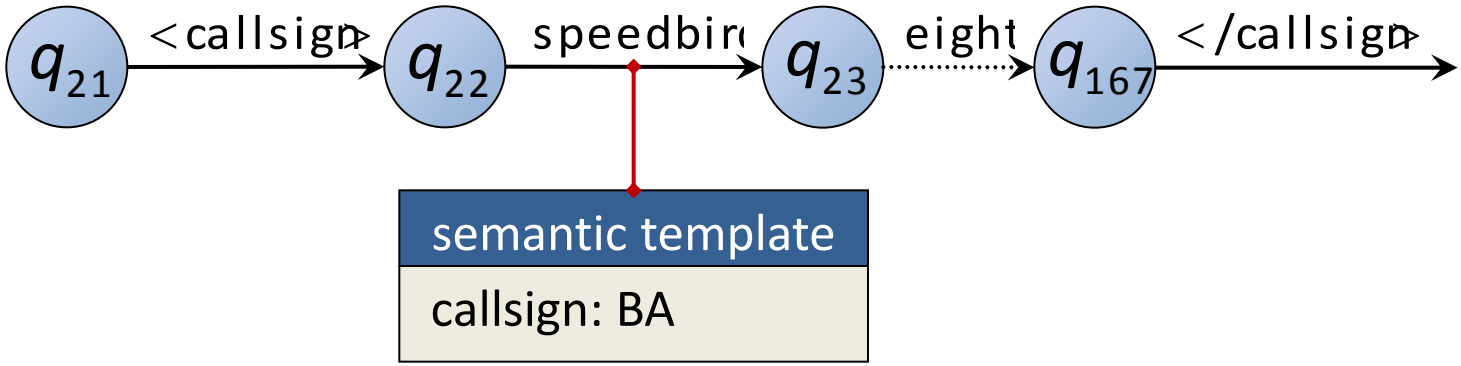
- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

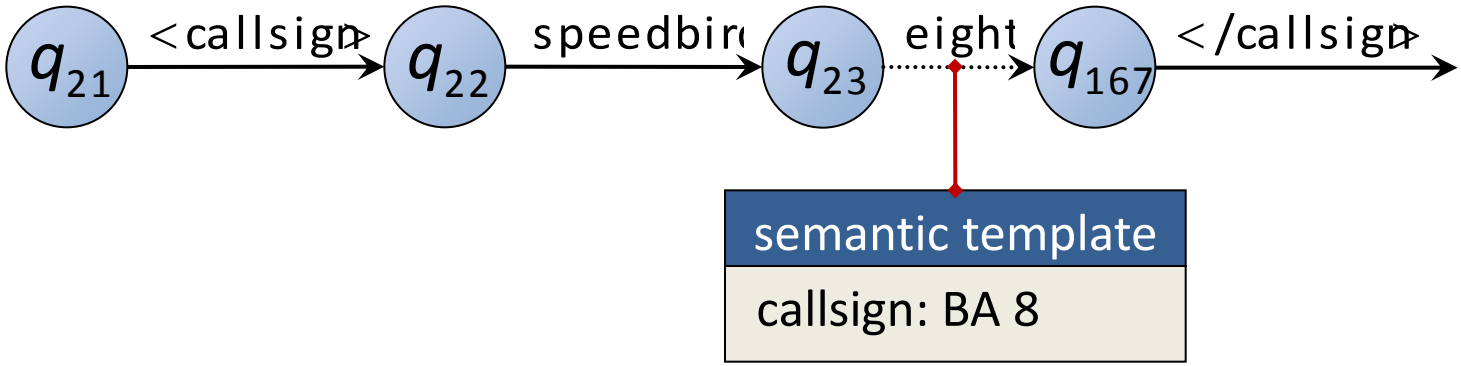
- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

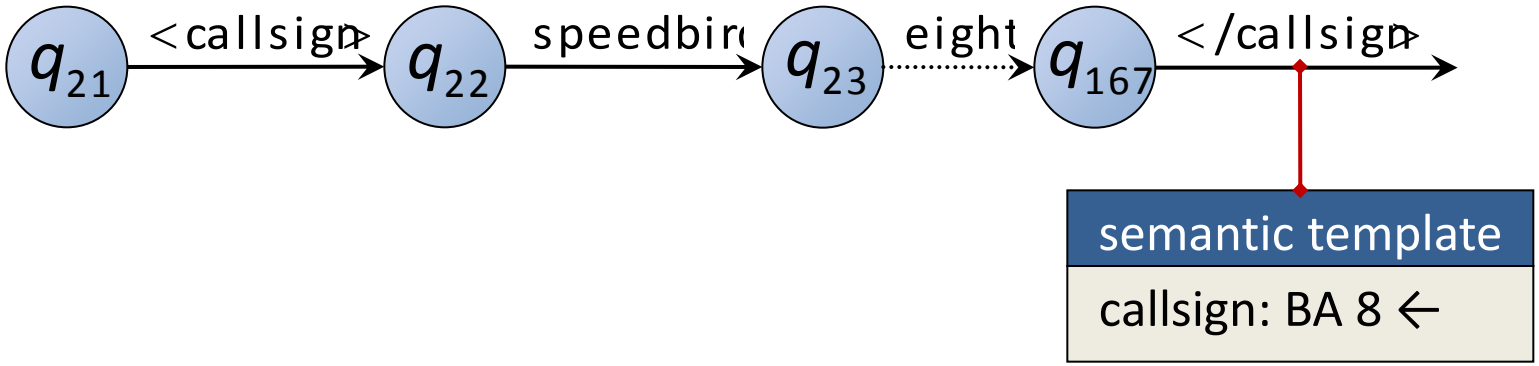
- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- This allows us to directly extract semantic frames by simply adding a pointer to a semantic template structure, which is filled on the fly

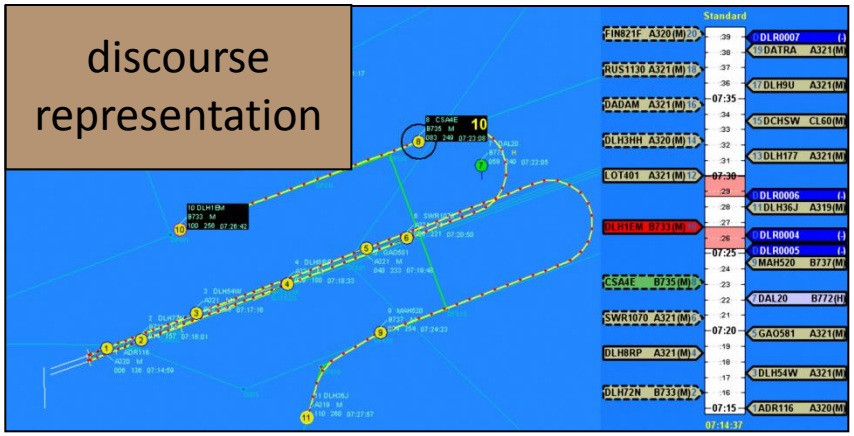


Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

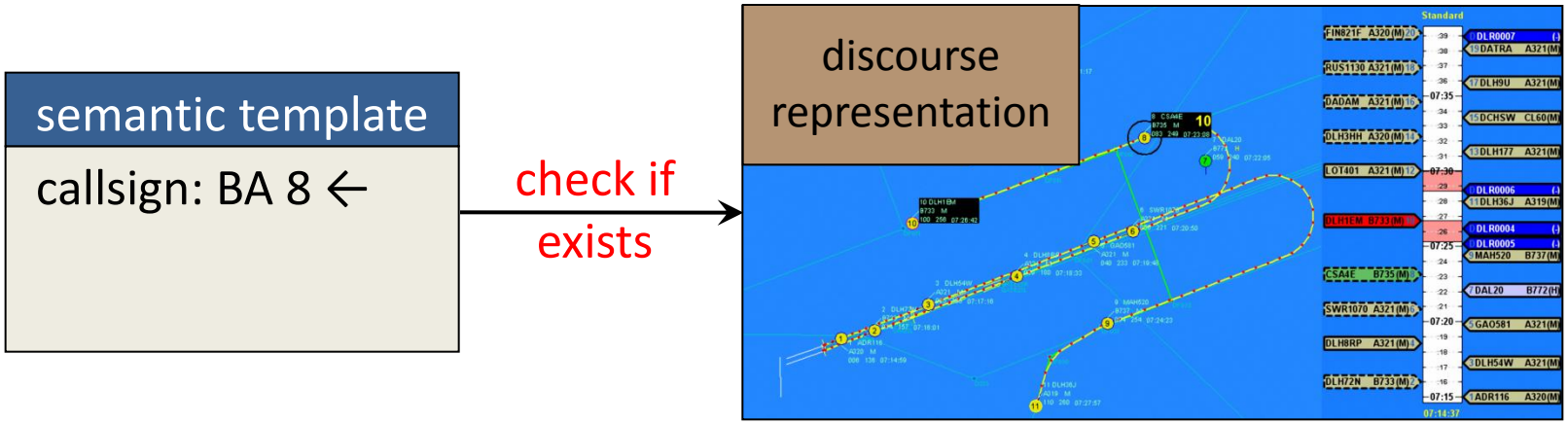
semantic template
callsign: BA 8 ←



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

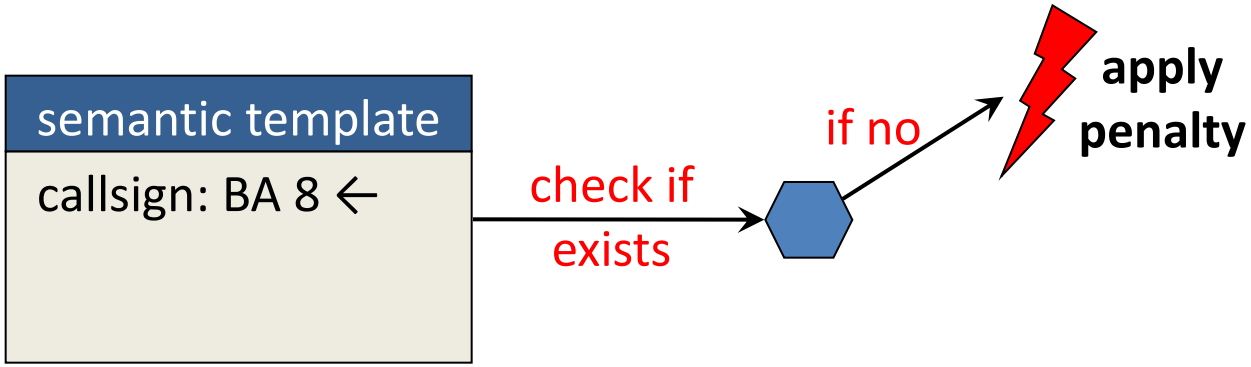
- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

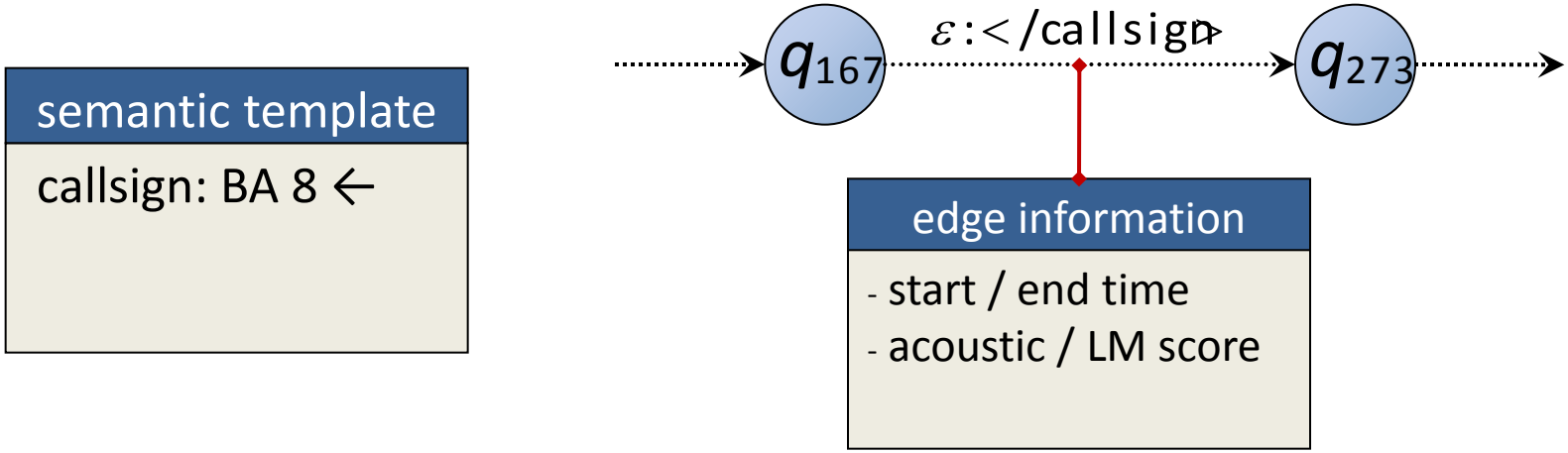
- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

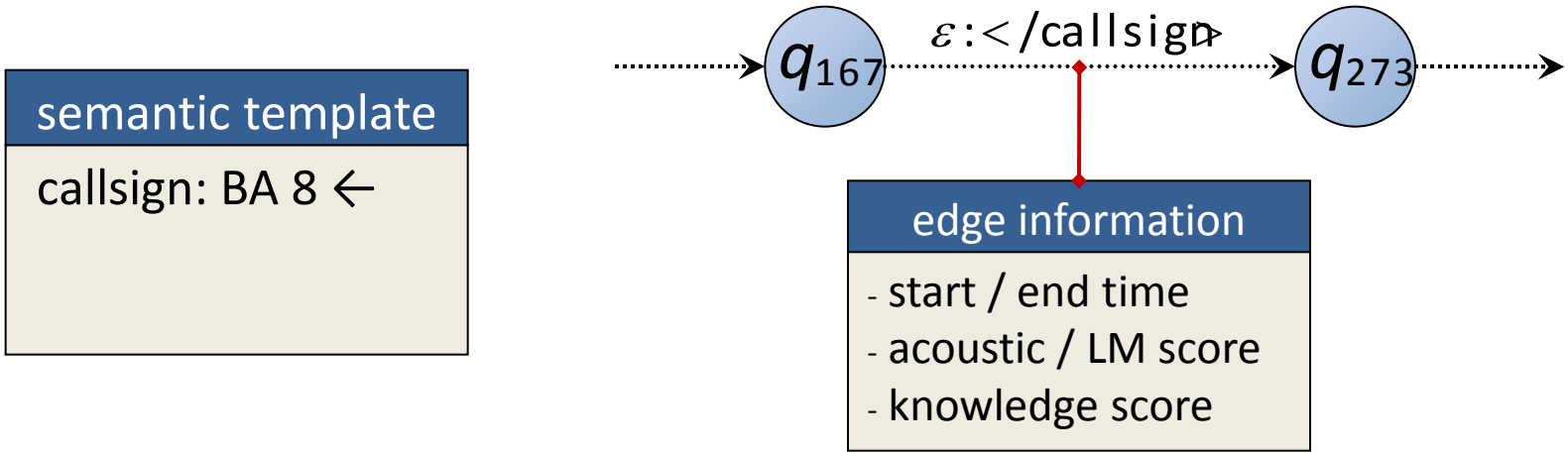
- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

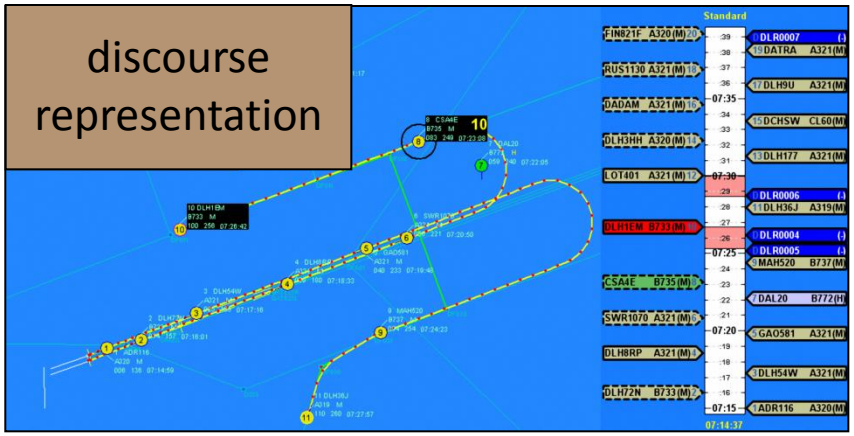


Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8 ←

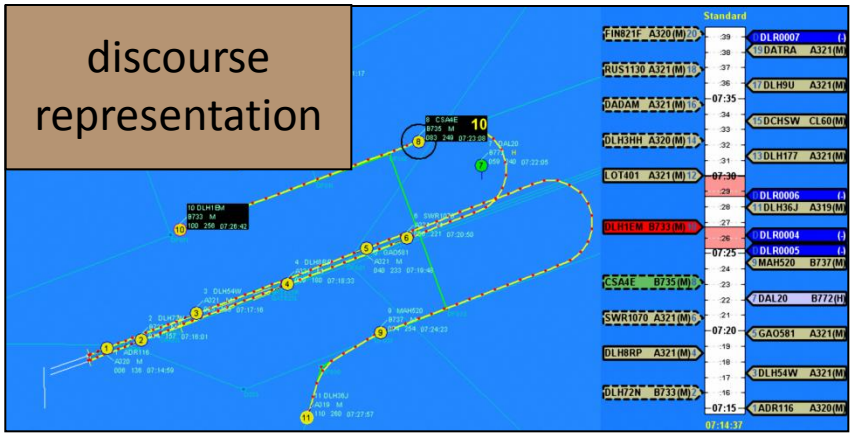


Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8 cmd: REDUCE val: SPD=220 ←



Knowledge-Based Lattice Rescoring

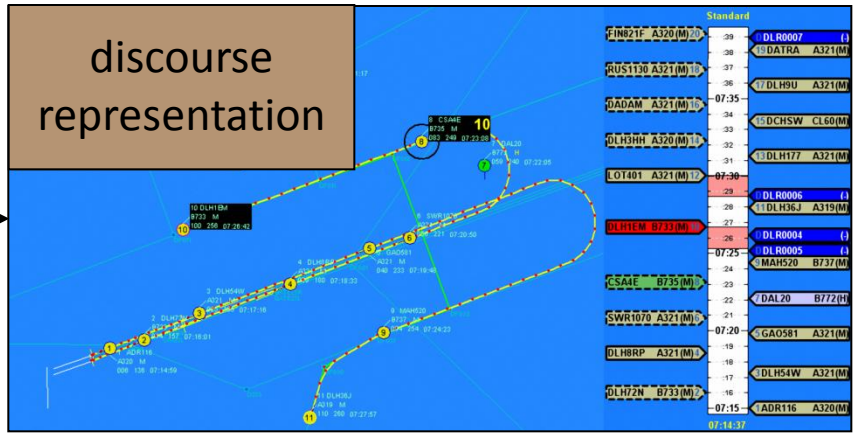
- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template

callsign: BA 8
cmd: REDUCE
val: SPD=220 ←

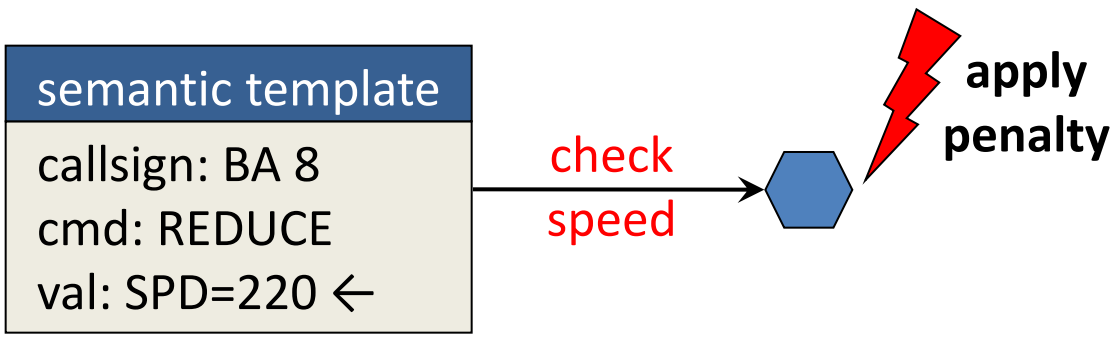
check
speed



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

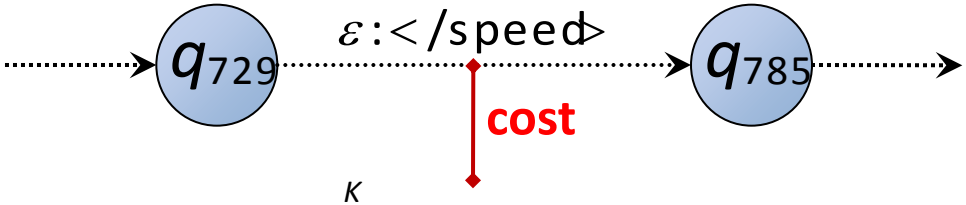


Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8
cmd: REDUCE
val: SPD=220 ←



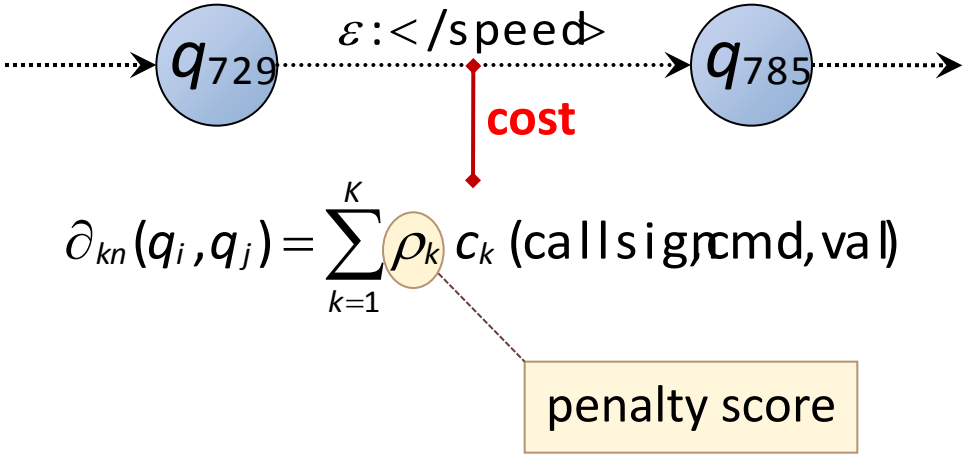
$$\partial_{kn}(q_i, q_j) = \sum_{k=1}^{K} \rho_k \, c_k \, (\text{callsign}, \text{cmd}, \text{val})$$

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8
cmd: REDUCE
val: SPD=220 ←

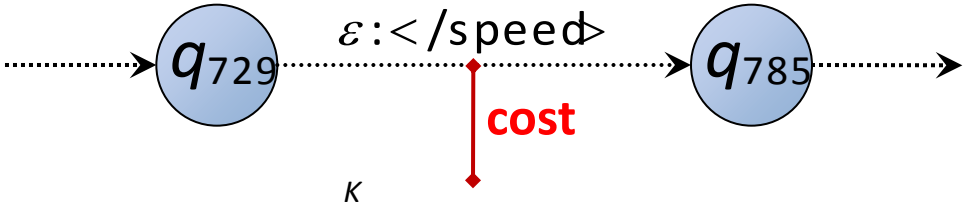


Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8
cmd: REDUCE
val: SPD=220 ←



$$\partial_{kn}(q_i, q_j) = \sum_{k=1}^{K} \rho_k \, c_k(\text{callsign}, \text{cmd}, \text{val})$$

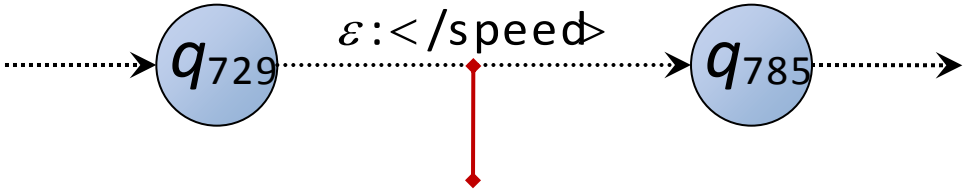
constraint penalty function $c_k(\cdot) \rightarrow [0,1]$

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Principal Idea:** penalize invalid callsigns and unlikely command values based on discourse representation system

semantic template
callsign: BA 8
cmd: REDUCE
val: SPD=220 ←



total cost:

$$\partial(\cdot) = \omega_{ac}\partial_{ac}(\cdot) + \omega_{lm}\partial_{lm}(\cdot) + \omega_{kn}\partial_{kn}(\cdot)$$

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

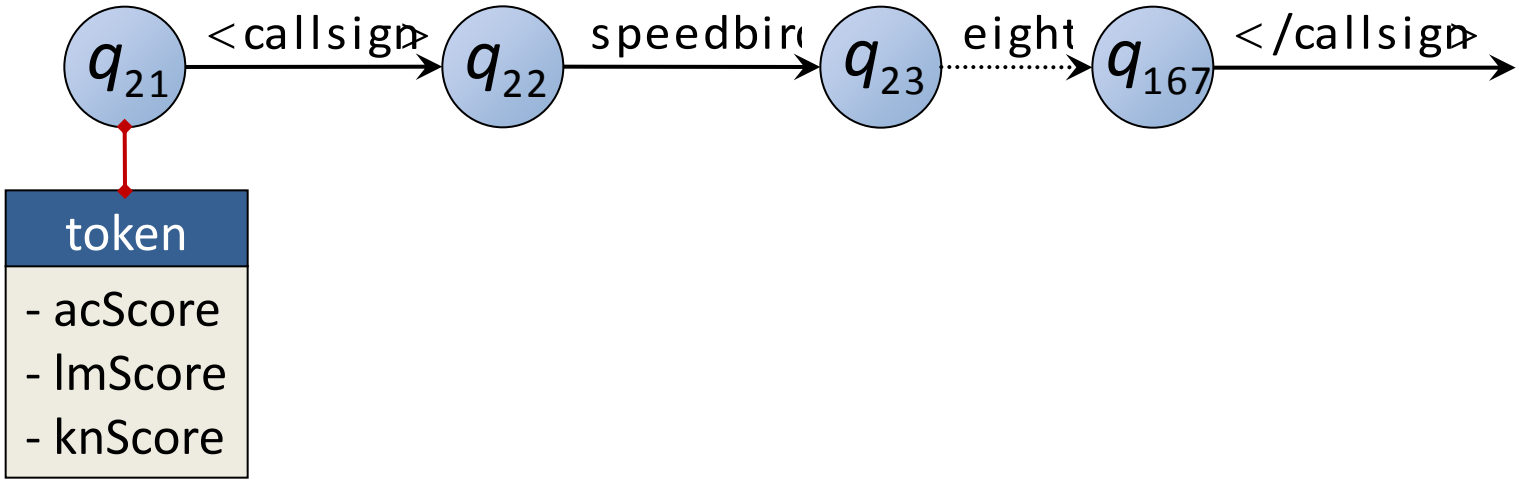
- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties

1 Propagate scores along the nodes

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

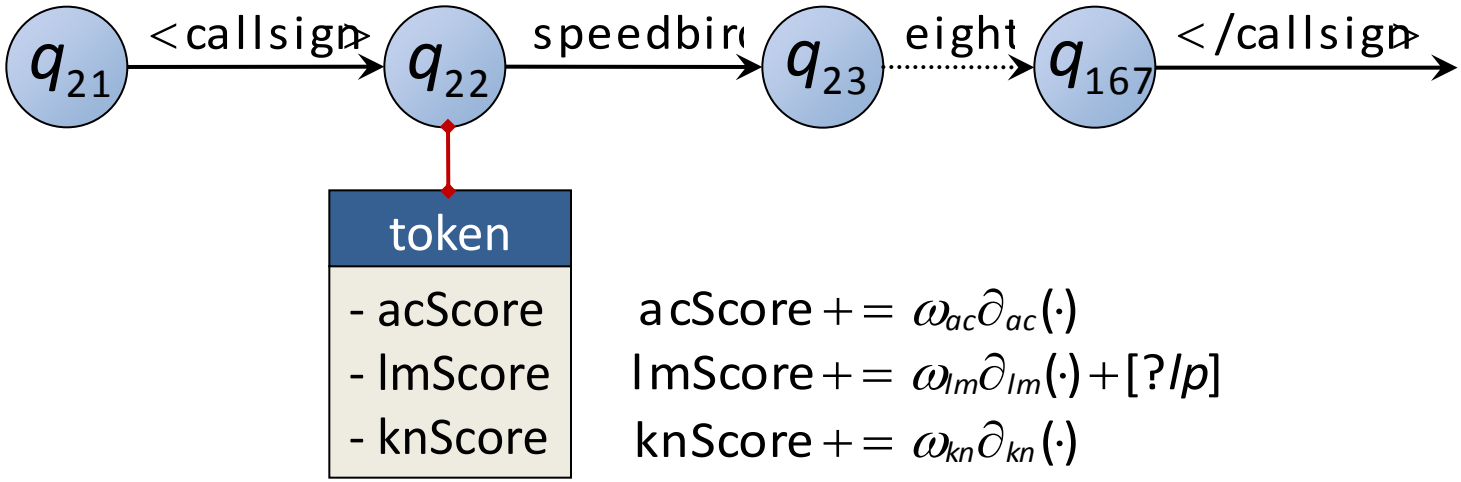
- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

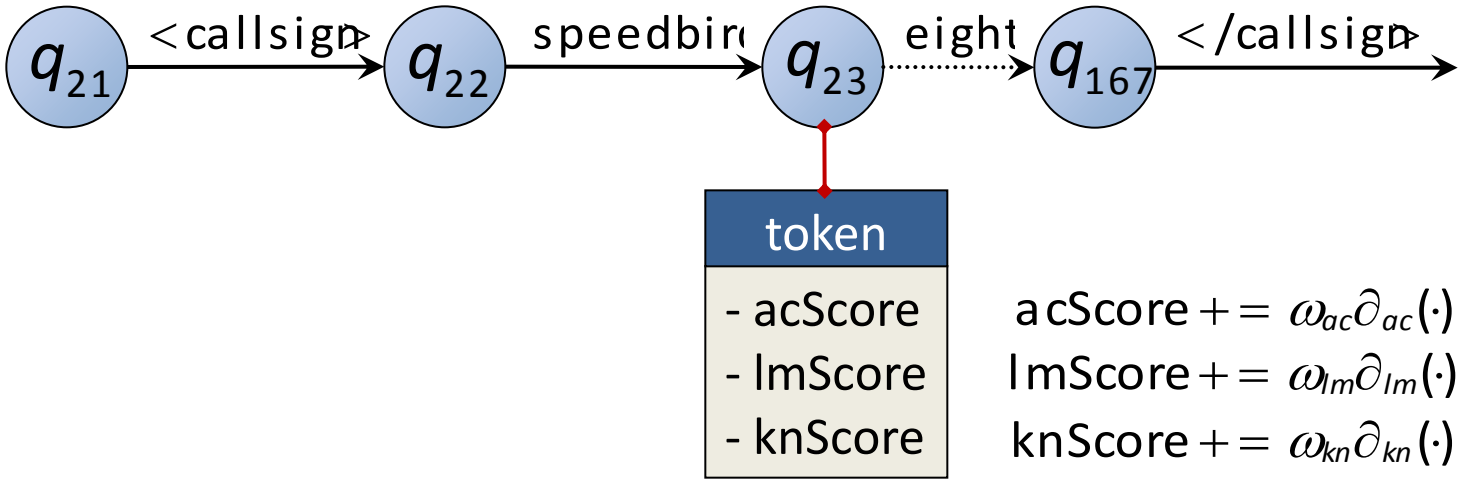
- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties



Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties

1 Propagate scores along the nodes

Knowledge-Based Lattice Rescoring

- **WFST Decoder:** can be used to generate phone-to-word transducer lattice during Viterbi search

- **Rescoring:** in analogy to rescoring with different language model scales and word insertion penalties

1. Propagate scores along the nodes

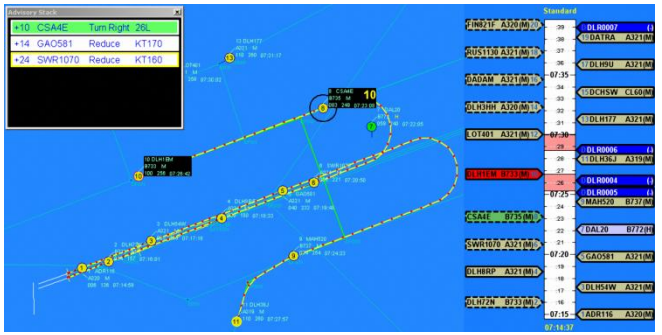
2. Trace back from best final state

Section IV

Rescoring Experiments

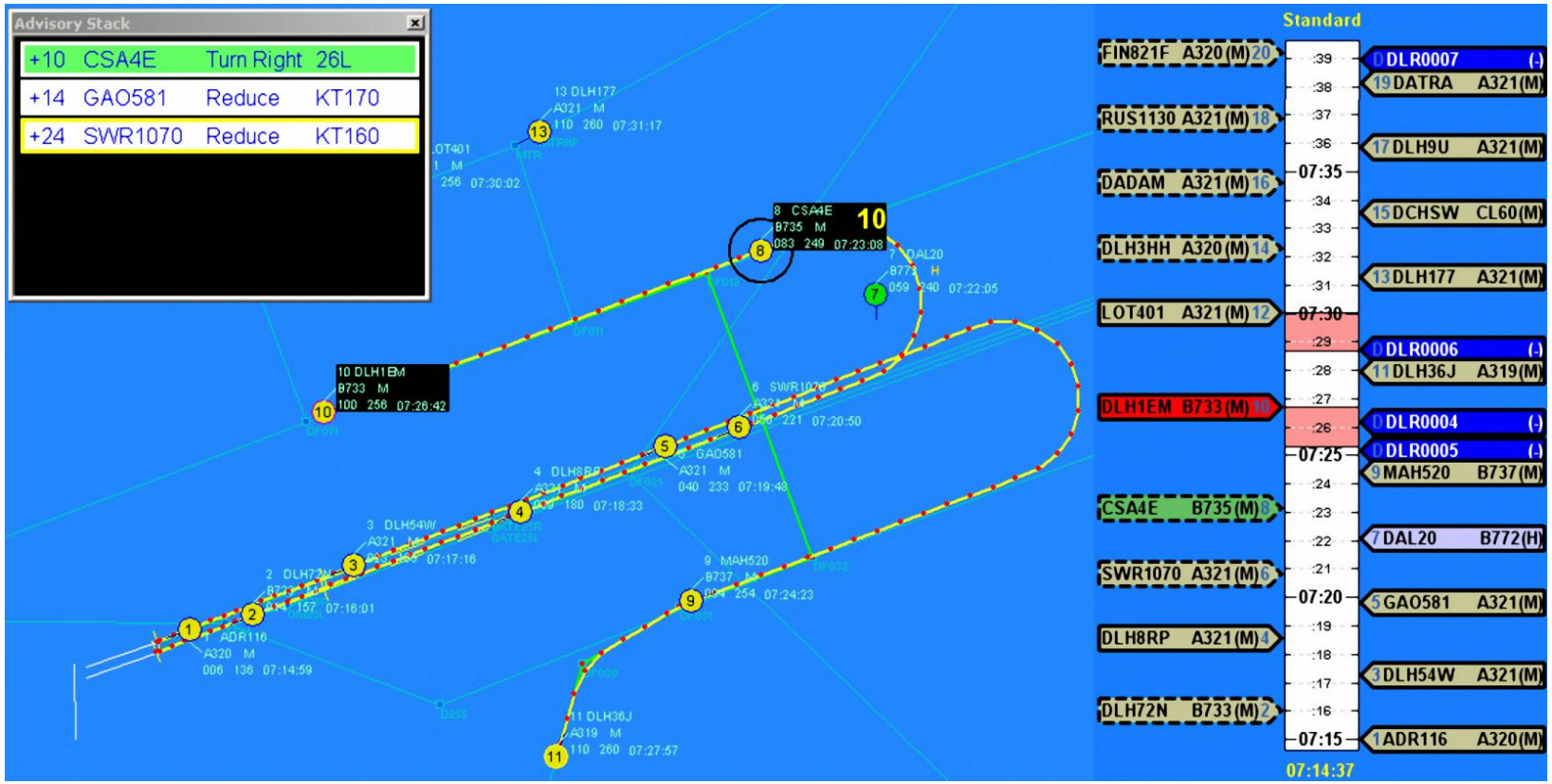
Rescoring Experiments

- **Corpus:**
  - recorded using the 4D-CARMA software from DLR
  - includes aircraft state vectors (5 second intervals)
  - total of 1,107 ATC commands
  - 9.5 words per sentence
  - approx. 100 minutes of speech



Rescoring Experiments

- **GUI used by the participants**



Rescoring Experiments

- **Recognition Grammar:**
  - branching factor of 7.68
  - 244,220 unique callsigns
  - sentence perplexity of 3.9E+09

Rescoring Experiments

- **Recognition Grammar:**
  - branching factor of 7.68
  - 244,220 unique callsigns
  - sentence perplexity of 3.9E+09

- **Used Constraints:**
  - callsign
  - speed
  - altitude

Rescoring Experiments

- **Recognition Grammar:**
  - branching factor of 7.68
  - 244,220 unique callsigns
  - sentence perplexity of 3.9E+09

- **Used Constraints:**
  - callsign
  - speed
  - altitude

Rescoring	WER	SER	MRR
None (baseline)	2.81	22.58	0.849
Callsign	0.55	4.61	0.966
Callsign, Spd, Alt	0.52	4.52	0.967
Oracle	0.31	2.07	0.979

Section V

Conclusions

Conclusions

- We have shown how dynamic context knowledge can profitably be used for rescoring ASR hypotheses

Conclusions

- We have shown how dynamic context knowledge can profitably be used for rescoring ASR hypotheses

- This is of particular interest in scenarios where explicit context information is available, such as
  - ATC: radar-derived aircraft state vectors
  - video games
  - virtual reality


INTERSPEECH 2012
Spoken Language Processing and Biomedicine
Thank you very much
for your attention!
SAARLAND
UNIVERSITY
GERMANY
DLR
GERMAN
AEROSPACE CENTER

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

- **Can easily be extended to class-based N-grams**

```
[callsign] [cmd_turn] [direction] heading [heading] .
```

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

- **Can easily be extended to class-based N-grams**

```
[callsign] [cmd_turn] [direction] heading [heading] .
```

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

- **Can easily be extended to class-based N-grams**

```
[callsign] [cmd_turn] [direction] heading [heading] .
```

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

- **Can easily be extended to class-based N-grams**

```
[callsign] [cmd_turn] [direction] heading [heading] .
```

The Air Traffic Control Task

- **XML-tags for easy extraction of semantic frames:**

```
<callsign> delta four three niner </callsign> <cmd_turn>
turn </cmd_turn> <direction> right </direction> heading
<heading> two two zero </heading>
```

- **Can easily be extended to class-based N-grams**

```
[callsign] [cmd_turn] [direction] heading [heading] .
```