



UNIVERSITY OF
CAMBRIDGE

A Network-based End-to-End Trainable Task-oriented Dialogue System

Deep Learning Summer school, 05 Aug 2016

Tsung-Hsien (Shawn) Wen

Dialogue Systems Group

Outline

2

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

Outline

3

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

The two paradigms

4

- ◎ Goal-oriented Dialogue Systems

The two paradigms

4

- ⊙ Goal-oriented Dialogue Systems
 - ⊙ Help the user to accomplish domain tasks

The two paradigms

4

- ⊙ Goal-oriented Dialogue Systems
 - ⊙ Help the user to accomplish domain tasks
 - ⊙ Domain specific, hard to collect data

The two paradigms

4

- ⊙ Goal-oriented Dialogue Systems
 - ⊙ Help the user to accomplish domain tasks
 - ⊙ Domain specific, hard to collect data
- ⊙ Current Systems
 - ⊙ Modular, highly handcrafted, restricted ability

The two paradigms

4

- ⊙ Goal-oriented Dialogue Systems
 - ⊙ Help the user to accomplish domain tasks
 - ⊙ Domain specific, hard to collect data
- ⊙ Current Systems
 - ⊙ Modular, highly handcrafted, restricted ability
- ⊙ Can we train a dialogue system on a small dataset w/ a minimal amount of handcrafting?

The two paradigms

4

- ⊙ Goal-oriented Dialogue Systems
 - ⊙ Help the user to accomplish domain tasks
 - ⊙ Domain specific, hard to collect data
- ⊙ Current Systems
 - ⊙ Modular, highly handcrafted, restricted ability
- ⊙ Can we train a dialogue system on a small dataset w/ a minimal amount of handcrafting?
- ⊙ How can we collect data to train this model?

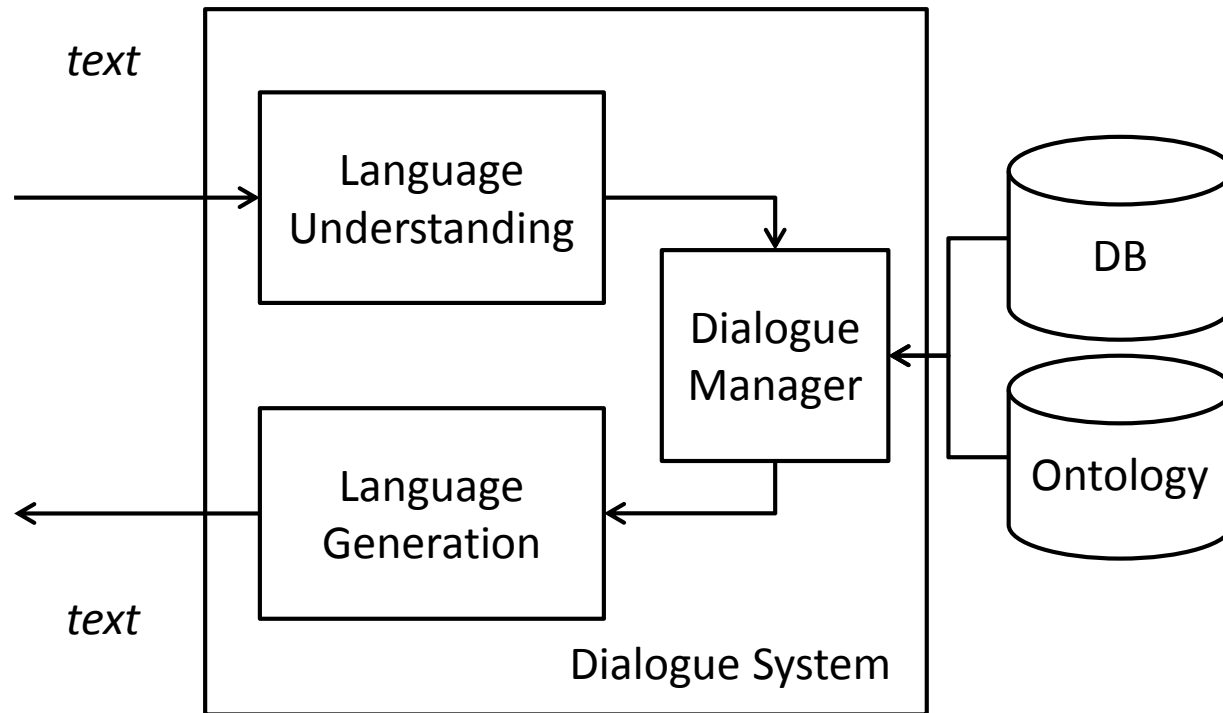
Outline

5

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

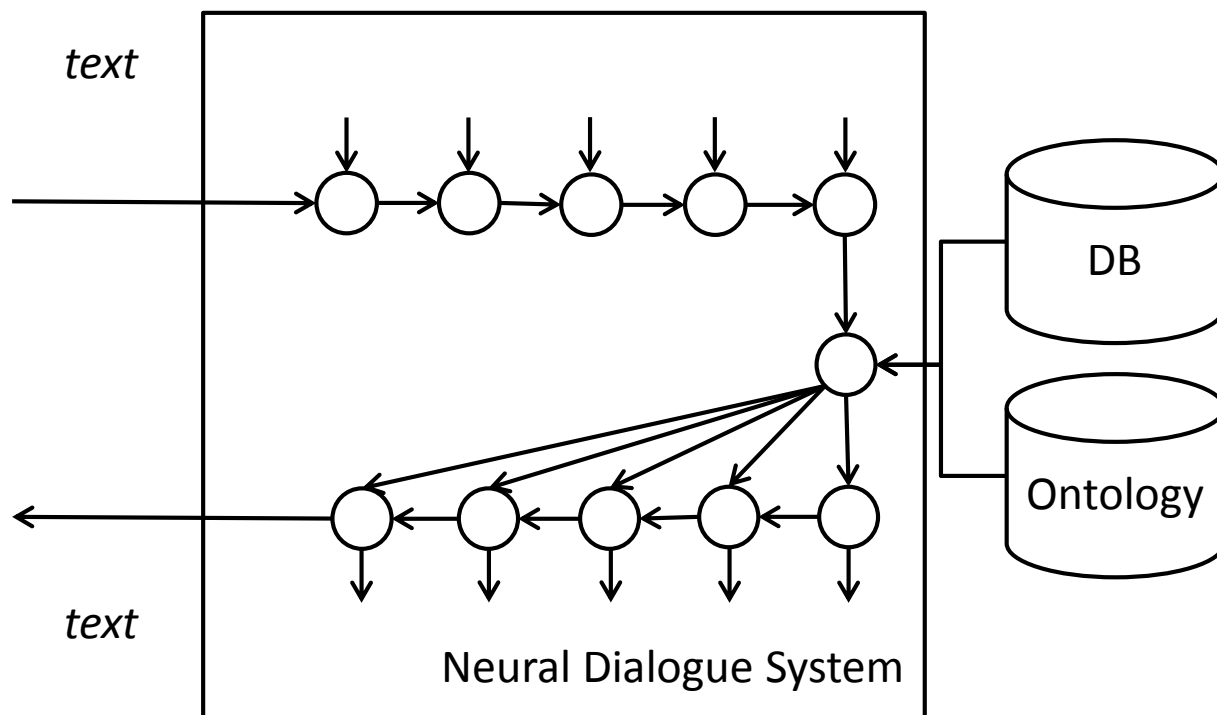
Traditional Dialogue Systems

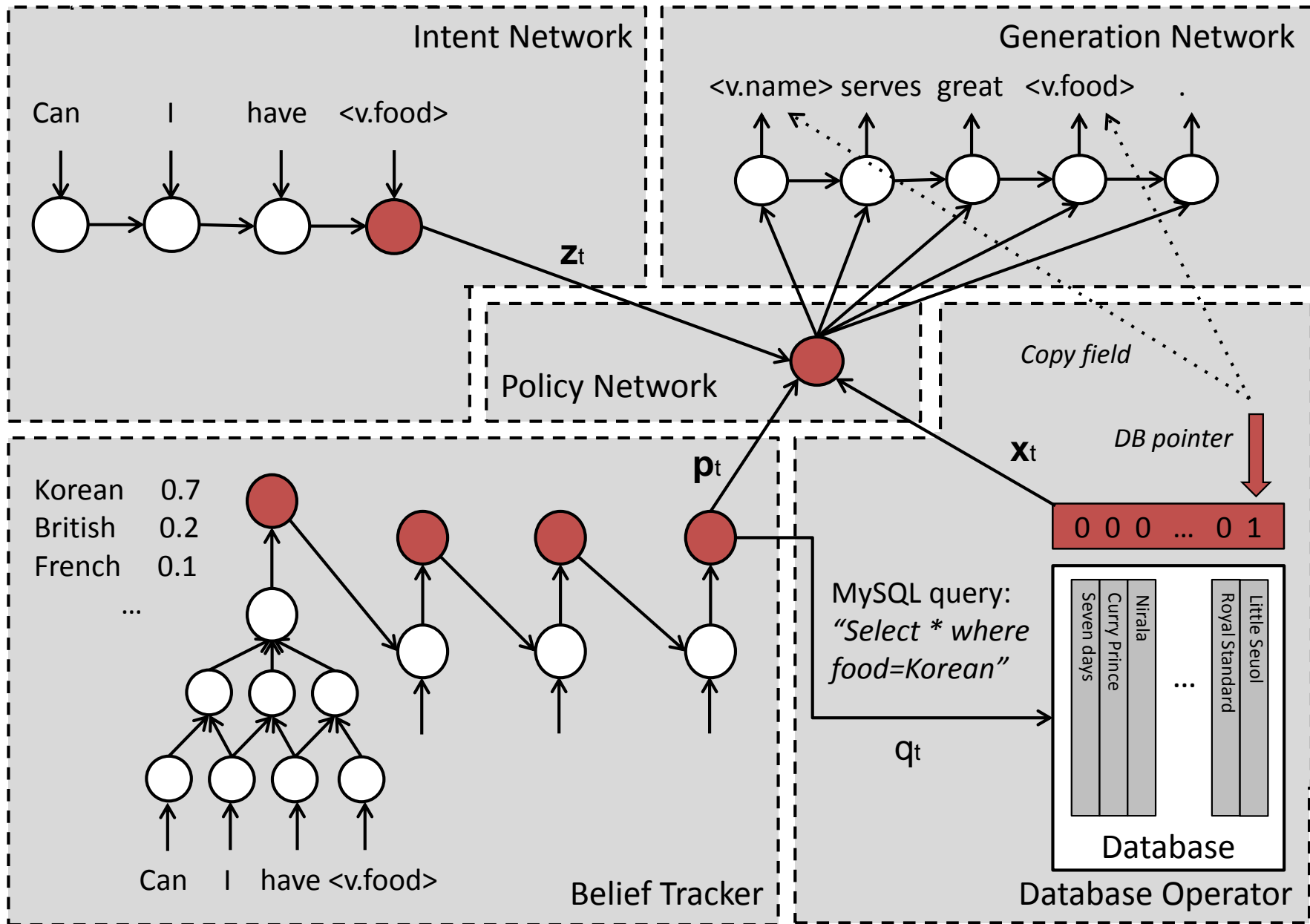
6

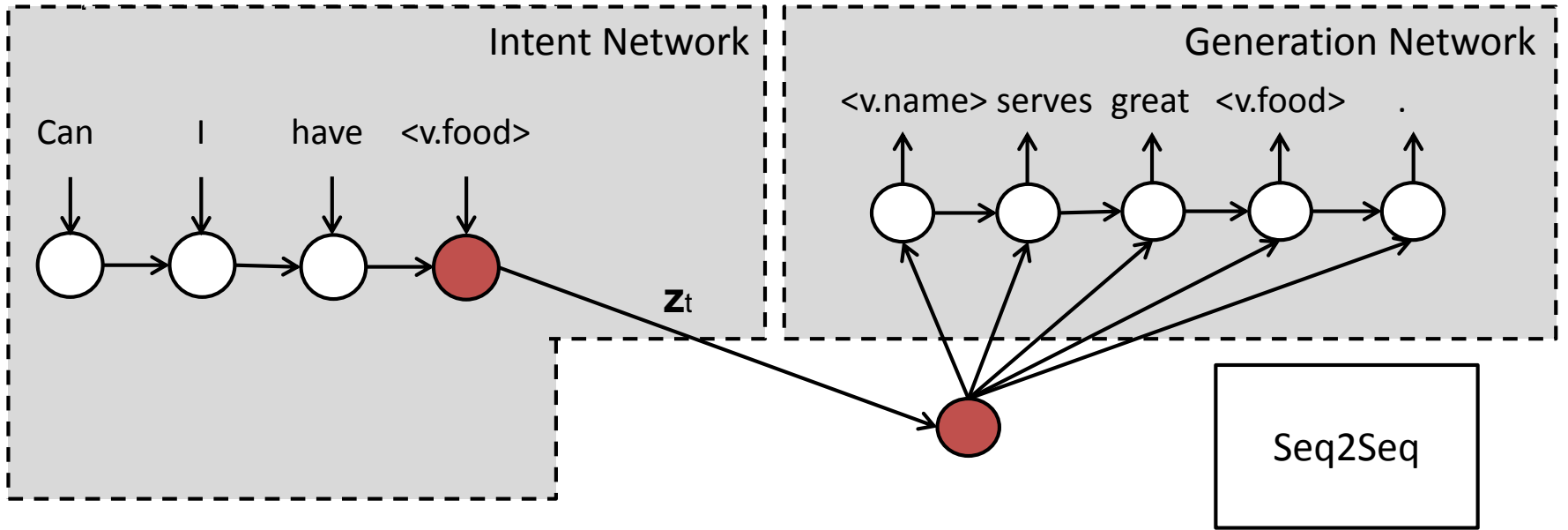


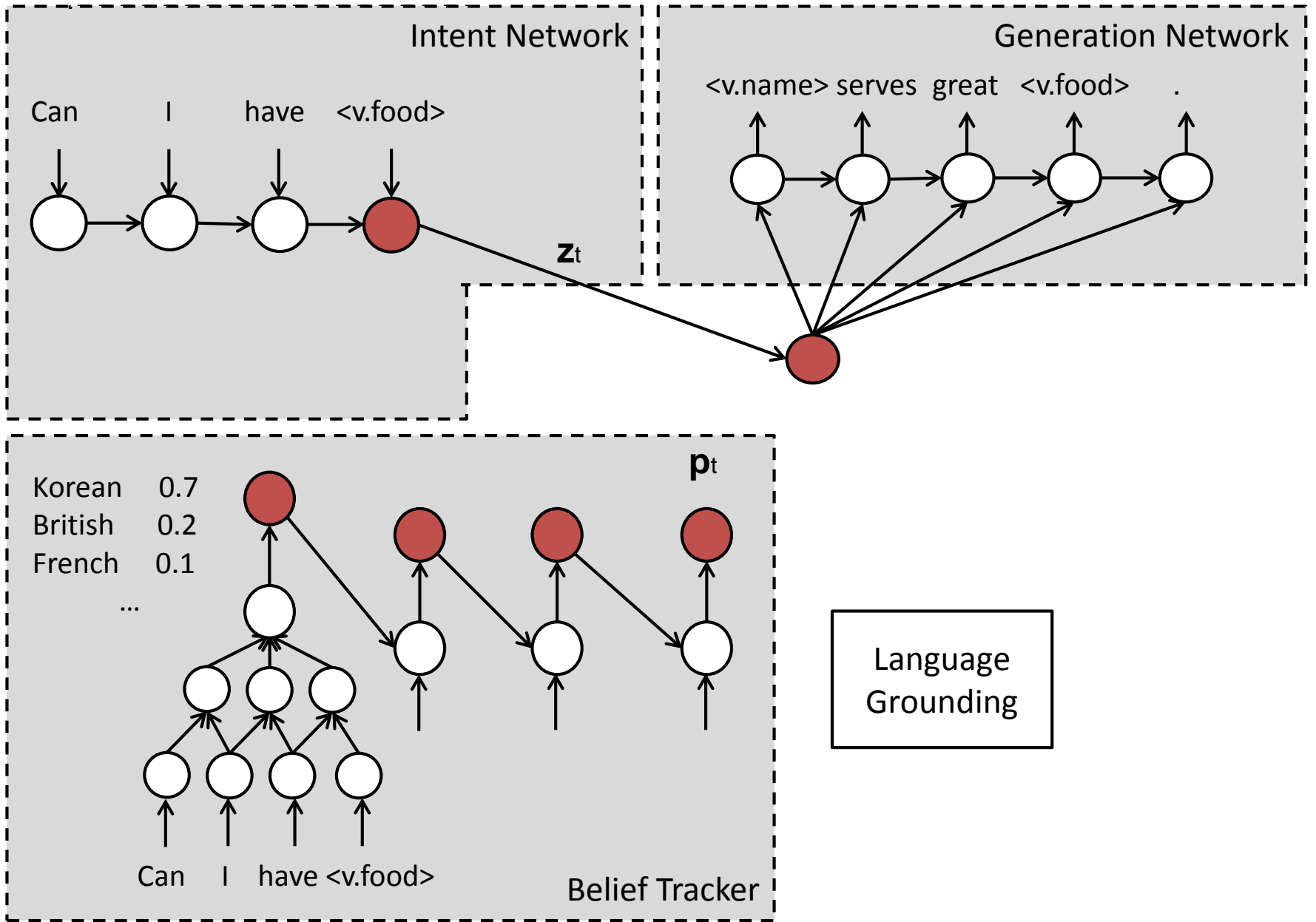
Neural Dialogue Systems

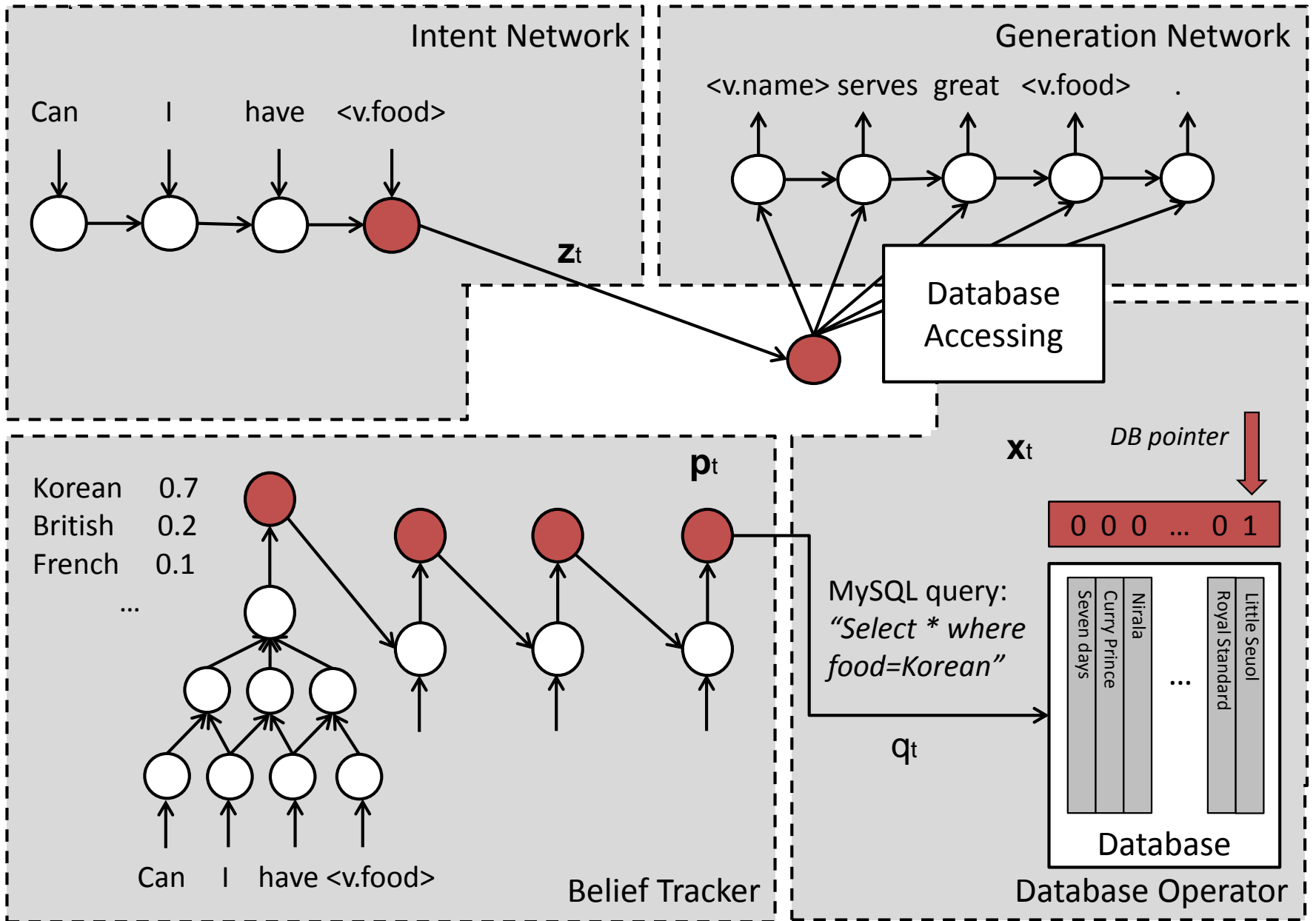
7

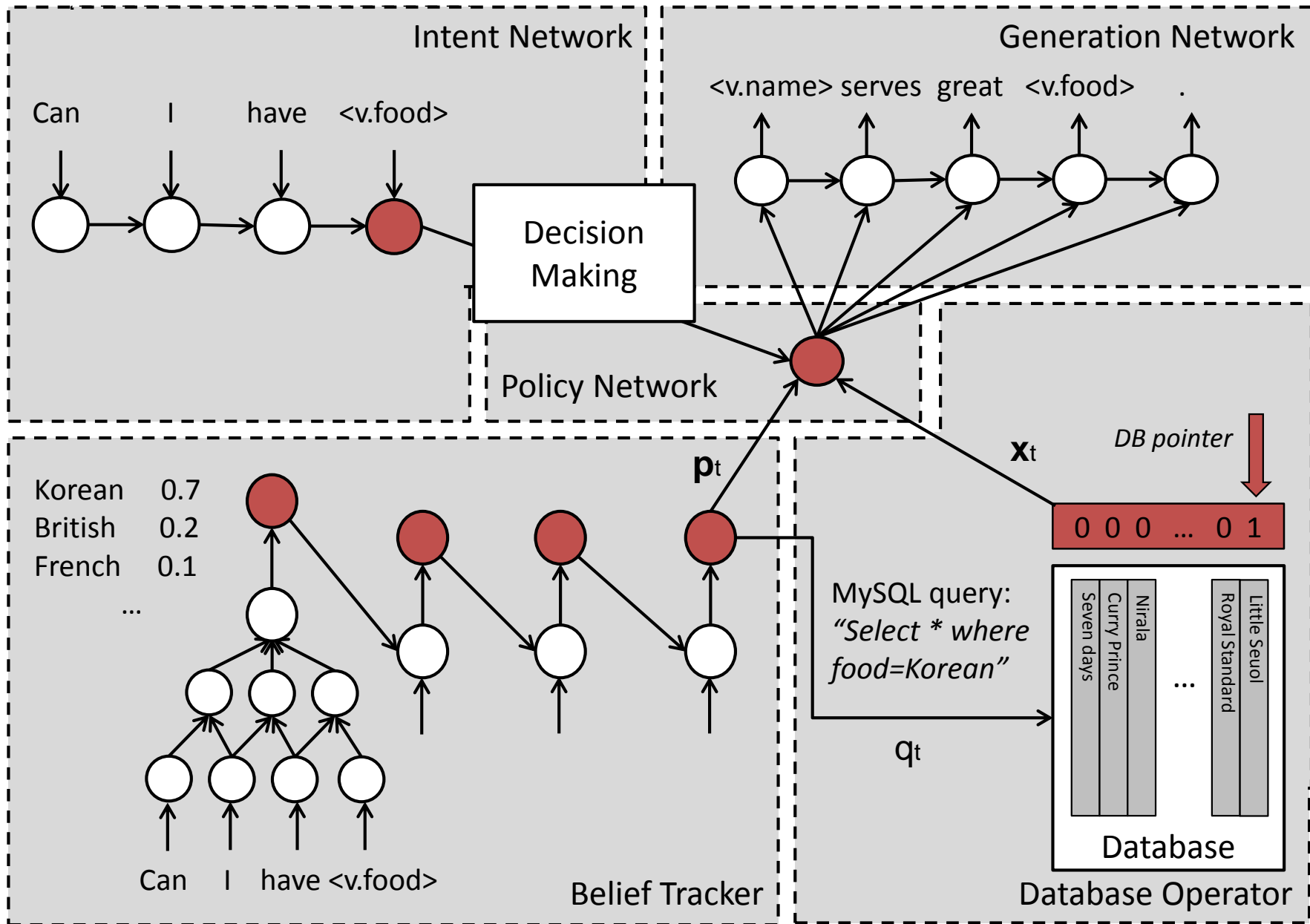


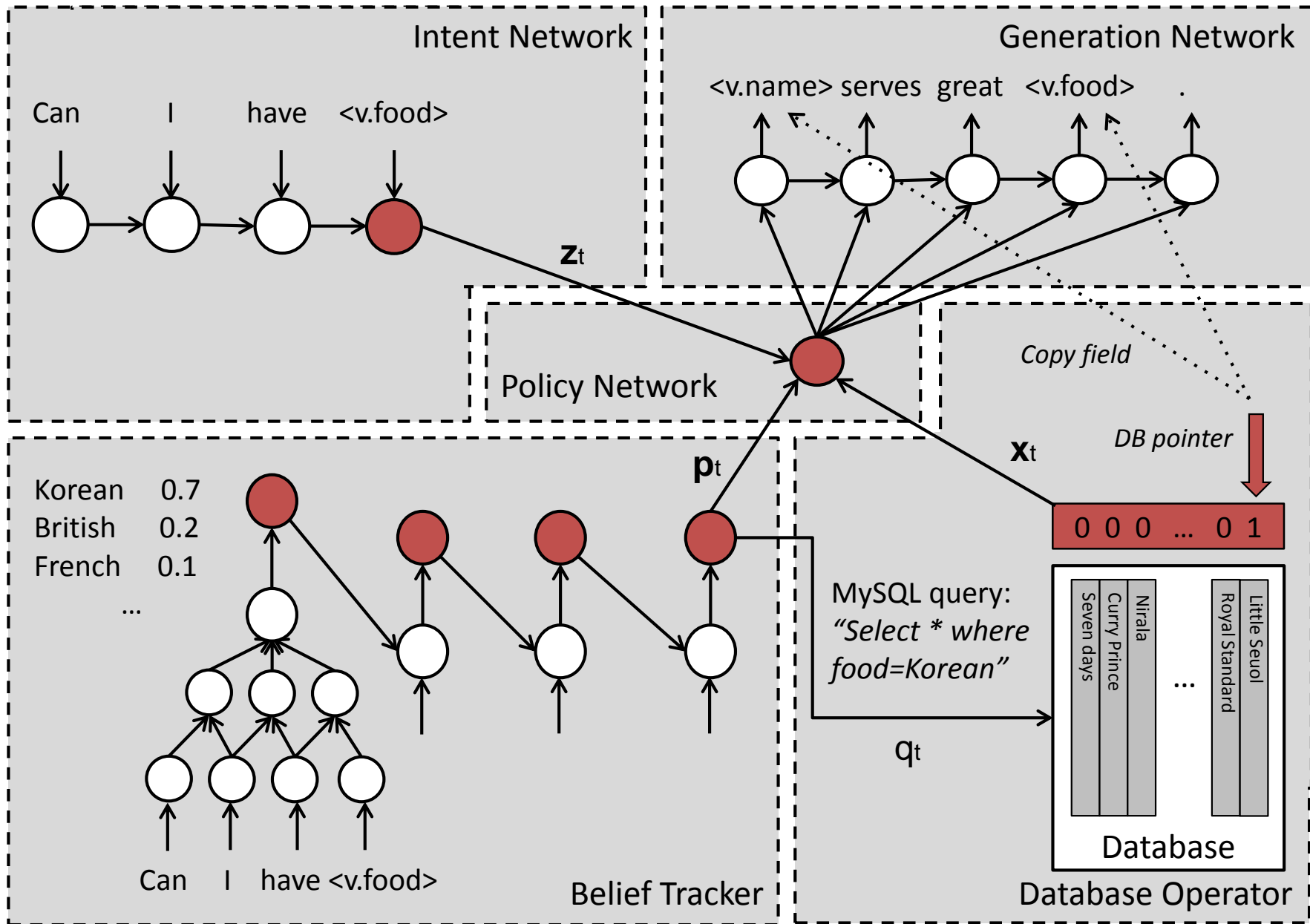












Outline

14

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

Wizard of Oz Data Collection

15

- ⦿ Online parallel version of WOZ on MTurk

Wizard of Oz Data Collection

15

- ⦿ Online parallel version of WOZ on MTurk
 - ⦿ Randomly hire a worker to be user/wizard.

Wizard of Oz Data Collection

15

- ⊙ Online parallel version of WOZ on MTurk
 - ⊙ Randomly hire a worker to be user/wizard.
 - ⊙ Task: Enter an appropriate response for ONE TURN.

Wizard of Oz Data Collection

15

- ⊙ Online parallel version of WOZ on MTurk
 - ⊙ Randomly hire a worker to be user/wizard.
 - ⊙ Task: Enter an appropriate response for ONE TURN.
 - ⊙ Repeat the process until all dialogues are finished.

Wizard of Oz Data Collection

15

- ⊙ Online parallel version of WOZ on MTurk
 - ⊙ Randomly hire a worker to be user/wizard.
 - ⊙ Task: Enter an appropriate response for ONE TURN.
 - ⊙ Repeat the process until all dialogues are finished.

- ⊙ Advantage:

Wizard of Oz Data Collection

15

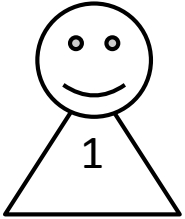
- ⊙ Online parallel version of WOZ on MTurk
 - ⊙ Randomly hire a worker to be user/wizard.
 - ⊙ Task: Enter an appropriate response for ONE TURN.
 - ⊙ Repeat the process until all dialogues are finished.

- ⊙ Advantage:
 - ⊙ Avoid latency, parallelisable, cheap

Wizard of Oz Data Collection

16

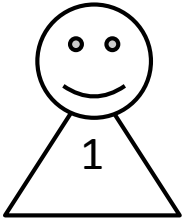
Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



Wizard of Oz Data Collection

16

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**

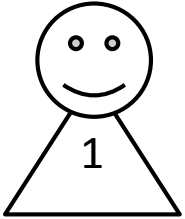


Hi, I want a cheap Korean restaurant.

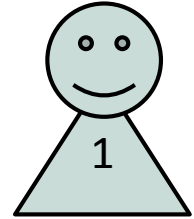
Wizard of Oz Data Collection

16

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



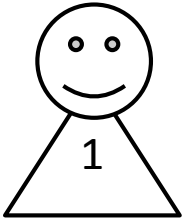
Hi, I want a cheap Korean restaurant.



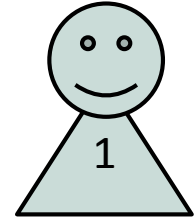
Wizard of Oz Data Collection

16

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

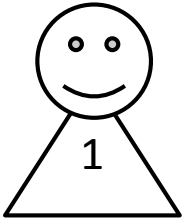


What user wants?	
Food	Korean
Price	Cheap
Area	N/A

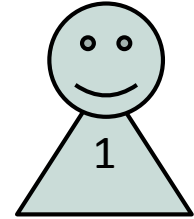
Wizard of Oz Data Collection

16

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*



What user wants?

Food	Korean
Price	Cheap
Area	N/A

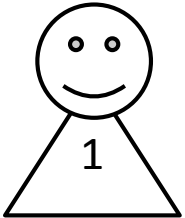
Search Table

Little Seoul	...
Best Korea	...
...	

Wizard of Oz Data Collection

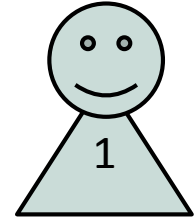
16

Task:
Find a **restaurant**,
cheap, Korean, North
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*



What user wants?

Food	Korean
Price	Cheap
Area	N/A

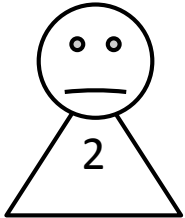
Search Table

Little Seoul	...
Best Korea	...
...	

Wizard of Oz Data Collection

17

Task:
Find a **restaurant**,
cheap, **Korean**, North
Ask **phone number**



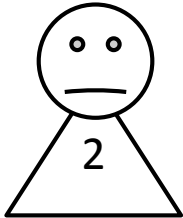
*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

Wizard of Oz Data Collection

17

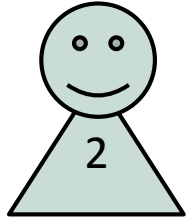
Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

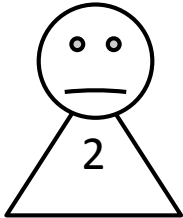
*Somewhere in the **north**.*



Wizard of Oz Data Collection

17

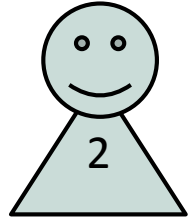
Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

*Somewhere in the **north**.*

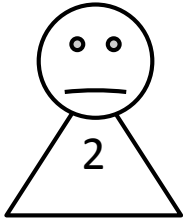


What user wants?	
Food	Korean
Price	Cheap
Area	North

Wizard of Oz Data Collection

17

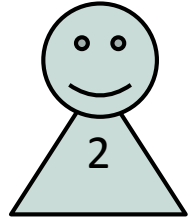
Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

*Somewhere in the **north**.*



What user wants?

Food	Korean
Price	Cheap
Area	North

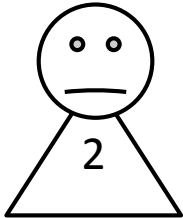
Search Table

Little Seoul	...
--------------	-----

Wizard of Oz Data Collection

17

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**

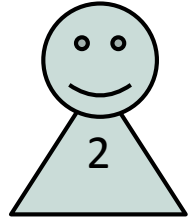


*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

*Somewhere in the **north**.*

***Little Seoul** is nice one in the north.*



What user wants?

Food	Korean
Price	Cheap
Area	North

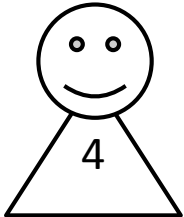
Search Table

Little Seoul	...
--------------	-----

Wizard of Oz Data Collection

18

Task:
Find a **restaurant**,
cheap, **Korean**, **North**
Ask **phone number**



*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

*Somewhere in the **north**.*

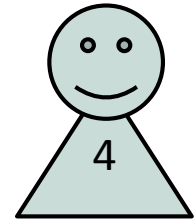
***Little Seoul** is nice one in the north.*

*What is the **phone number**?*

*Its phone number is **01223456789**.*

Thank you very much, good bye.

Thank you for using the system.



What user wants?

Food	Korean
Price	Cheap
Area	North

Search Table

Little Seoul	...
--------------	-----

Wizard of Oz Data Collection

19

*Hi, I want a **cheap Korean restaurant**.*

*What **area** are you looking for ?*

*Somewhere in the **north**.*

***Little Seoul** is nice one in the north.*

*What is the **phone number**?*

*Its phone number is **01223456789**.*

Thank you very much, good bye.

Thank you for using the system.

What user wants?	
Food	Korean
Price	Cheap
Area	North

Data Statistics

20

- ◎ Ontology:

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.
 - ⊙ 3 informable slots: area, price range, food type
 - ⊙ 3 requestable slots: address, phone, postcode

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.
 - ⊙ 3 informable slots: area, price range, food type
 - ⊙ 3 requestable slots: address, phone, postcode

- ⊙ Dataset

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.
 - ⊙ 3 informable slots: area, price range, food type
 - ⊙ 3 requestable slots: address, phone, postcode
- ⊙ Dataset
 - ⊙ 676 dialogues, ~2750 turns

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.
 - ⊙ 3 informable slots: area, price range, food type
 - ⊙ 3 requestable slots: address, phone, postcode
- ⊙ Dataset
 - ⊙ 676 dialogues, ~2750 turns
 - ⊙ 3000 HITS, takes 3 days, costs ~400 USD

Data Statistics

20

- ⊙ Ontology:
 - ⊙ Cambridge restaurant domain, 99 venues.
 - ⊙ 3 informable slots: area, price range, food type
 - ⊙ 3 requestable slots: address, phone, postcode
- ⊙ Dataset
 - ⊙ 676 dialogues, ~2750 turns
 - ⊙ 3000 HITS, takes 3 days, costs ~400 USD
 - ⊙ Data cleaning takes 2-3 days for one person

Outline

21

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

Experiments

22

- ⊙ Experimental details
 - ⊙ Train/valid/test: 3/1/1
 - ⊙ SGD, l2 regularisation, early stopping, gradient clip=1
 - ⊙ Hidden size = 50, Vocab size: ~500

Experiments

22

- ⊙ Experimental details
 - ⊙ Train/valid/test: 3/1/1
 - ⊙ SGD, l2 regularisation, early stopping, gradient clip=1
 - ⊙ Hidden size = 50, Vocab size: ~500
- ⊙ Two stage training:
 - ⊙ Training trackers with label cross entropy
 - ⊙ Training other parts with response cross entropy

Experiments

22

- ⊙ Experimental details
 - ⊙ Train/valid/test: 3/1/1
 - ⊙ SGD, l2 regularisation, early stopping, gradient clip=1
 - ⊙ Hidden size = 50, Vocab size: ~500
- ⊙ Two stage training:
 - ⊙ Training trackers with label cross entropy
 - ⊙ Training other parts with response cross entropy
- ⊙ Decoding
 - ⊙ Beam search w/ beam width 10
 - ⊙ Decode with average word likelihood

Human evaluation

23

System Comparison

Metric	N2N	Modular	Tie
Subj. Success	96.95%	95.12%	-
Avg. # of Turn	3.95	4.54	-
Comparisons(%)			
Naturalness	46.95*	25.61	27.44
Comprehension	45.12*	21.95	32.93
Preference	50.00*	24.39	25.61
Performance	43.90*	25.61	30.49

* $p < 0.005$, # of comparisons: 164

Example dialogues

24

Human : Restaurant in any area that serves Vietnamese food
Machine : thanh binh serves vietnamese food type , is there anything else i can help you with ?
Human : What is the phone number and post code
Machine : the number is 01223 362456 and the area code is c.b 3, 0 a.f . is there anything else i can do for you ?
Human : Good bye
Machine : thank you for using the cambridge restaurant system .

Human : Hi, I'm hungry for some Indonesian food
Machine : i am sorry , there are no indonesian restaurants . would you like a different food ?
Human : Oh no. How about Indian?
Machine : there are 9 restaurants serving indian types of food . do you have a preference on side of town ?
Human : East side please
Machine : sitar tandoori is a indian restaurant in the east side of town . would you like their information ?
Human : Yes, the address and phone number please
Machine : sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955 .
Human : Okay, thank you
Machine : thank you , goodbye .

Example dialogues

24

Human : Restaurant in any area that serves Vietnamese food
Machine : thanh binh serves vietnamese food type . is there anything else i can help you with ?
Human : What is the phone number and post code
Machine : the number is 01223 362456 and the area code is c.b 3, 0 a.f . is there anything else i can do for you ?
Human : Good bye
Machine : thank you for using the cambridge restaurant system .

Human : Hi, I'm hungry for some Indonesian food
Machine : i am sorry , there are no indonesian restaurants . would you like a different food ?
Human : Oh no. How about Indian?
Machine : there are 9 restaurants serving indian types of food . do you have a preference on side of town ?
Human : East side please
Machine : sitar tandoori is a indian restaurant in the east side of town . would you like their information ?
Human : Yes, the address and phone number please
Machine : sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955 .
Human : Okay, thank you
Machine : thank you , goodbye .

Example dialogues

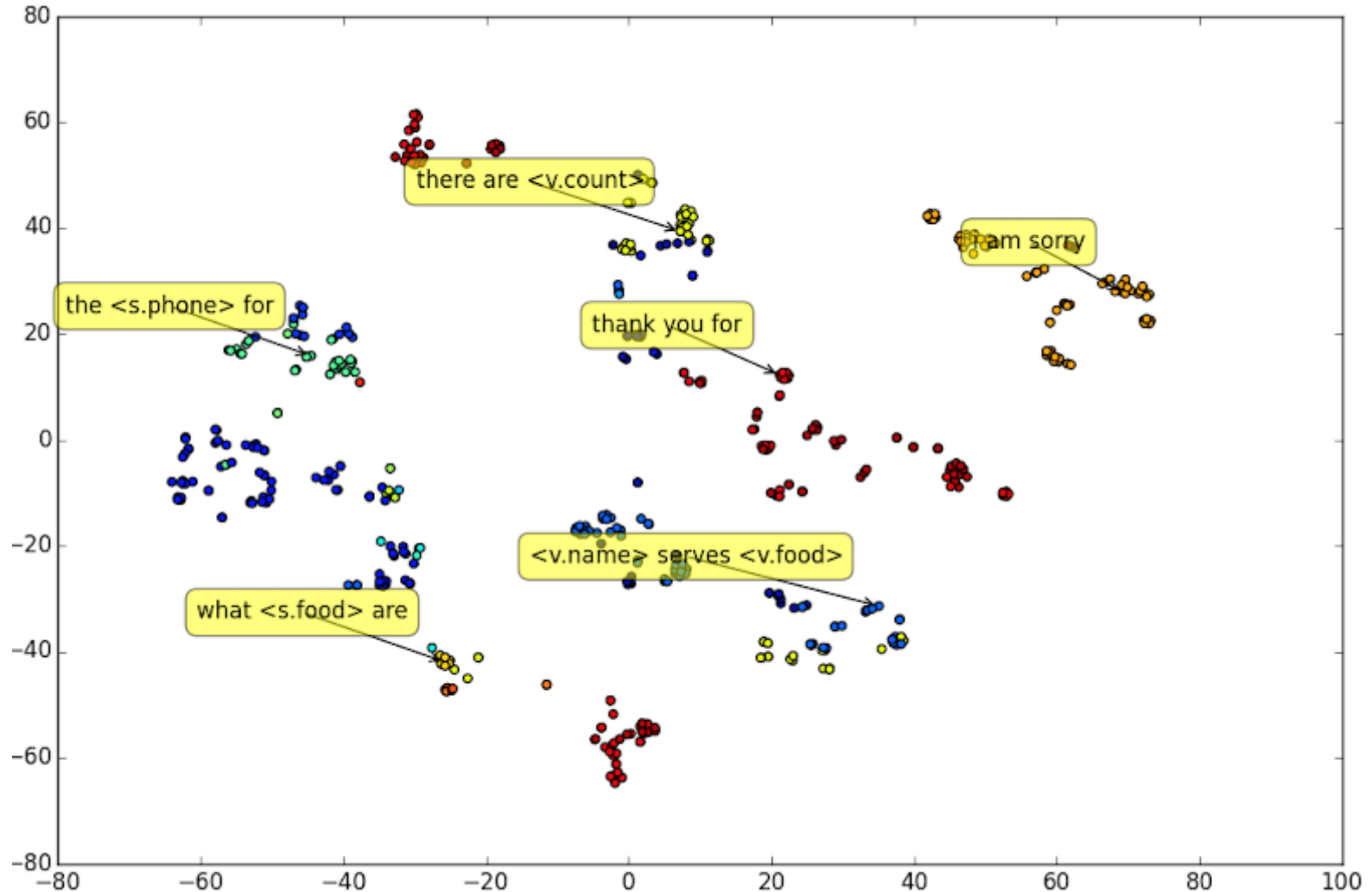
24

Human : Restaurant in any area that serves Vietnamese food
Machine : thanh binh serves vietnamese food type , is there anything else i can help you with ?
Human : What is the phone number and post code
Machine : the number is 01223 362456 and the area code is c.b 3, 0 a.f . is there anything else i can do for you ?
Human : Good bye
Machine : thank you for using the cambridge restaurant system .

Human : Hi, I'm hungry for some Indonesian food
Machine : i am sorry , there are no indonesian restaurants . would you like a different food ?
Human : Oh no. How about Indian?
Machine : there are 9 restaurants serving indian types of food . do you have a preference on side of town ?
Human : East side please
Machine : sitar tandoori is a indian restaurant in the east side of town . would you like their information ?
Human : Yes, the address and phone number please
Machine : sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955 .
Human : Okay, thank you
Machine : thank you , goodbye .

Visualising action embedding

25



Outline

26

- ⊙ Intro
- ⊙ Neural Dialogue System
- ⊙ Wizard-of-Oz Data Collection
- ⊙ Experiments
- ⊙ Conclusion

Conclusion

27

- ⦿ An end-to-end trainable task-oriented dialogue system architecture is proposed.

Conclusion

- ⊙ An end-to-end trainable task-oriented dialogue system architecture is proposed.
- ⊙ A complementary WOZ data collection is also proposed (no latency, parallel, cheap).

Conclusion

- ⊙ An end-to-end trainable task-oriented dialogue system architecture is proposed.
- ⊙ A complementary WOZ data collection is also proposed (no latency, parallel, cheap).
- ⊙ Results show that it can learn from human-human conversations and help users to complete tasks.

The paper

28

- ⊙ Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M.R. Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A Network-based End-to-End Trainable Task-oriented Dialogue System. *arXiv preprint: 1604.04562* 2016.



UNIVERSITY OF
CAMBRIDGE

Thank you! Questions?

*Tsung-Hsien Wen is supported by a studentship funded by Toshiba
Research Europe Ltd, Cambridge Research Laboratory*

Dialogue Systems Group

Response Generation Task

30

Model	Match (%)	Success (%)	BLEU
Seq2Seq [Sutskever et al, 2014]	-	-	0.1718
HRED [Serban et al, 2015]	-	-	0.1861
Our model w/o req. trackers	89.70	30.60	0.1799
Our full model	86.34	75.16	0.2313
Our full model + attention	90.88	80.02	0.2388

Example dialogues

31

Human : hello
Machine : what cuisine would you like ?
Human : afghan
Machine : i am sorry , there are no afghan restaurants . would you like something else ?
Human : korean
Machine : little seoul serves korean food , it is located at 108 regent street city centre .
Human : is it expensive?
Machine : little seoul serves korean food and is in the expensive price range .
Human : what area is that
Machine : little seoul is located in the centre area of town . would you like their number ?
Human : yes
Machine : the telephone number is 01223 308681 . is there anything else i can help you with today ?
Human : no goodbye
Machine : thank you for using the cambridge restaurant system , goodbye .

Example dialogues

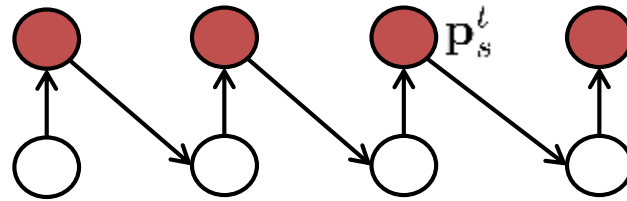
31

Human : hello
Machine : what cuisine would you like ?
Human : afghan
Machine : i am sorry , there are no afghan restaurants . would you like something else ?
Human : korean
Machine : little seoul serves korean food , it is located at 108 regent street city centre .
Human : is it expensive?
Machine : little seoul serves korean food and is in the expensive price range .
Human : what area is that
Machine : little seoul is located in the centre area of town . would you like their number ?
Human : yes
Machine : the telephone number is 01223 308681 . is there anything else i can help you with today ?
Human : no goodbye
Machine : thank you for using the cambridge restaurant system , goodbye .

Jordan RNN-CNN belief trackers

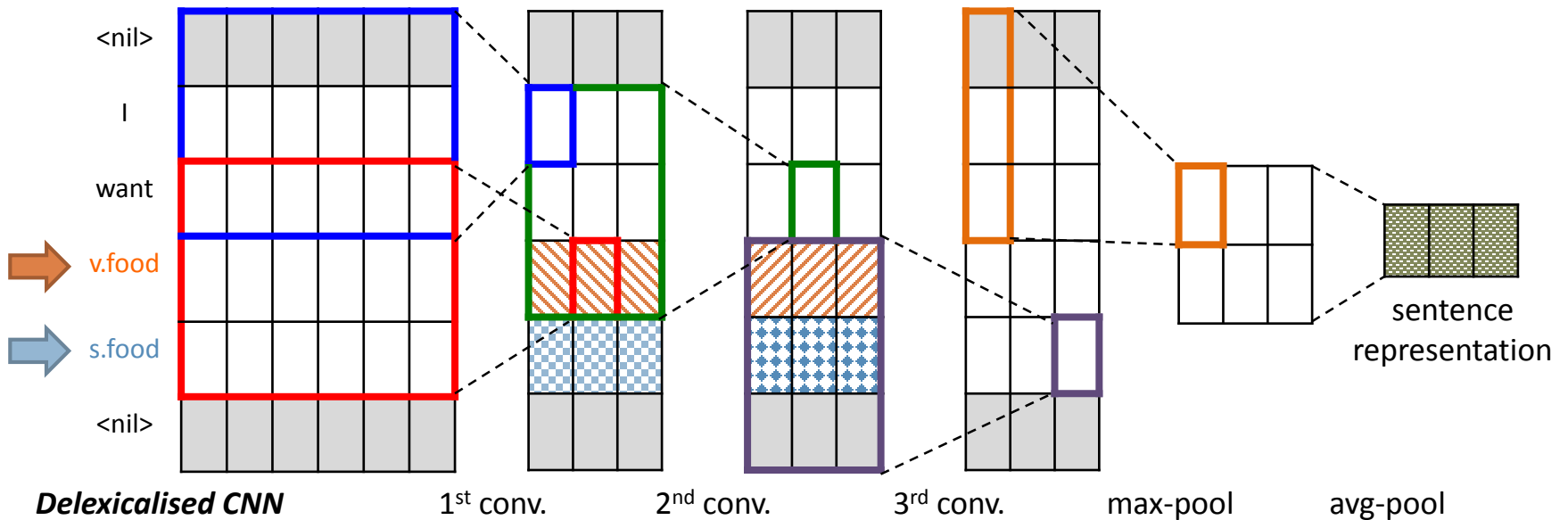
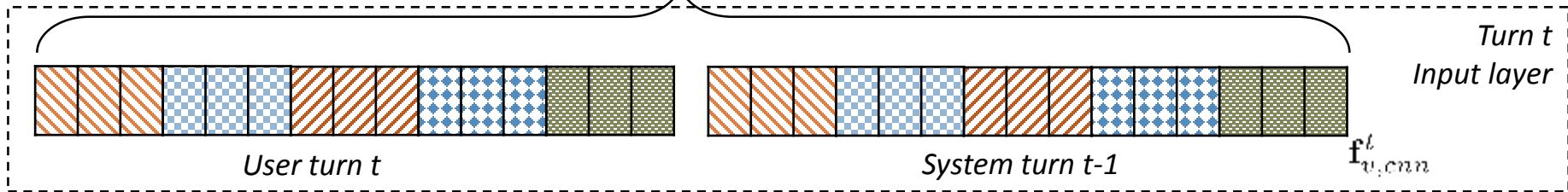
32

Jordan RNN



Output layer

Hidden layer



Jordan RNN-CNN belief trackers

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer

<nil>	█	█	█	█	█	█
I						
want						
Korean						
food						
<nil>	█	█	█	█	█	█

Delexicalised CNN

Jordan RNN-CNN belief trackers

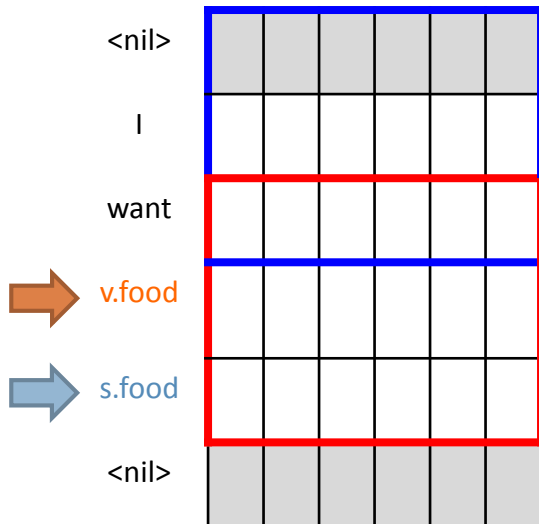
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Delexicalised CNN

Jordan RNN-CNN belief trackers

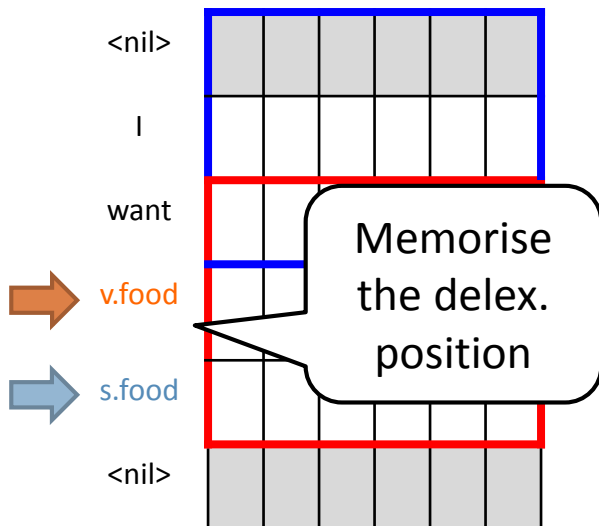
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Delexicalised CNN

Jordan RNN-CNN belief trackers

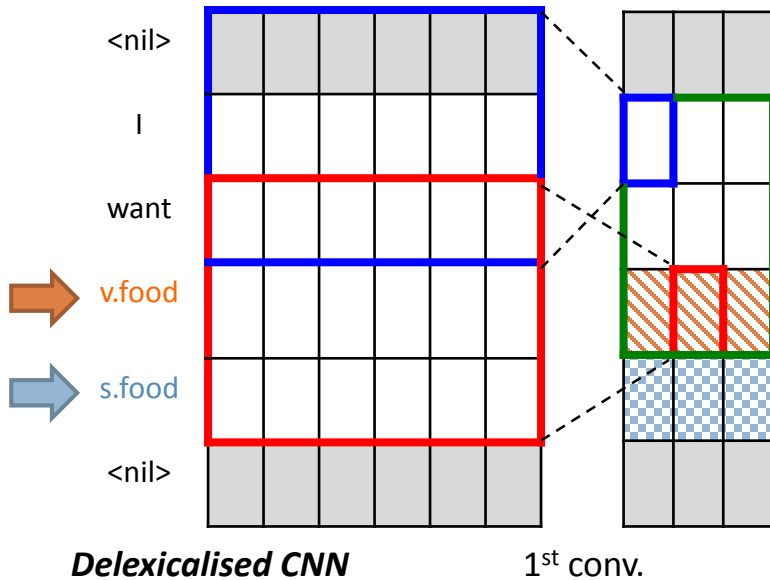
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Delexicalised CNN

1st conv.

Jordan RNN-CNN belief trackers

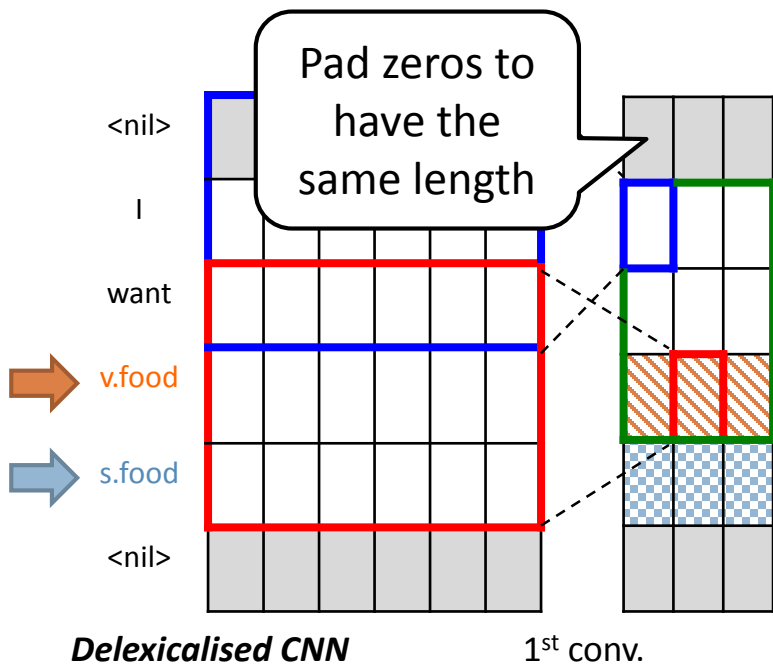
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

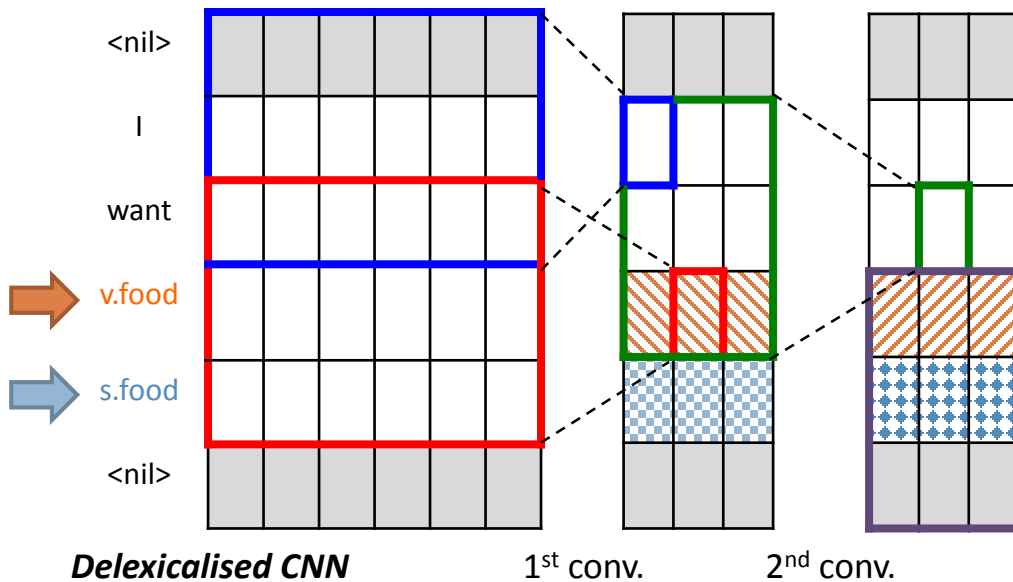
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

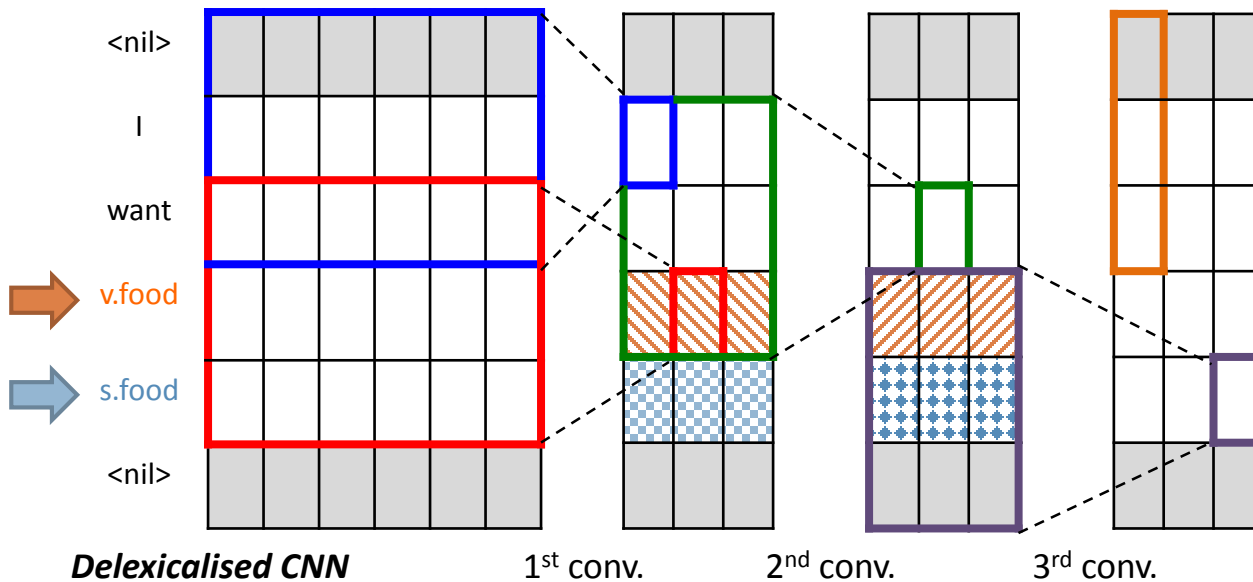
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

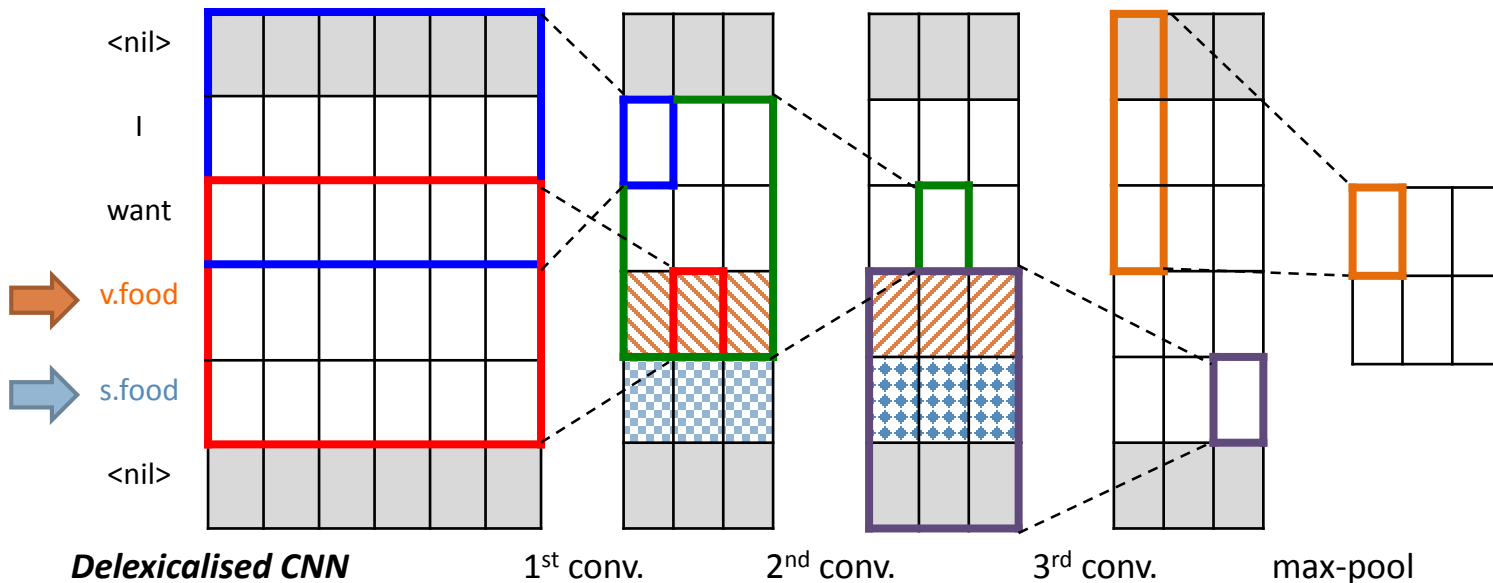
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

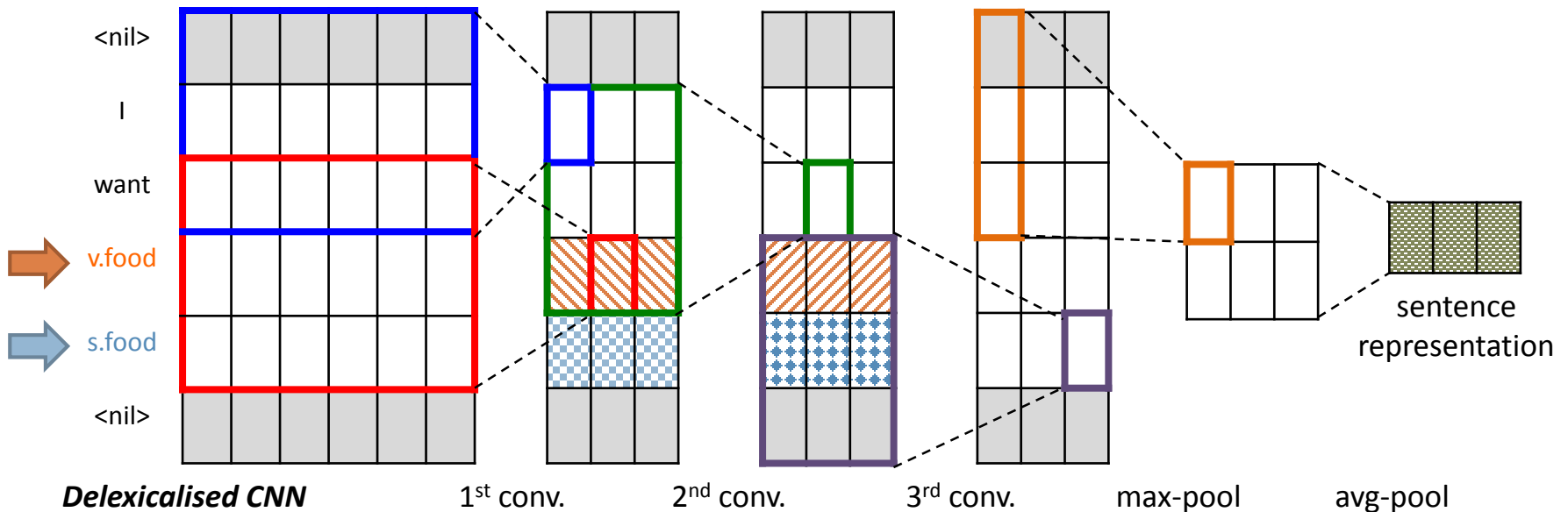
33

British French Korean ... Chinese

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

33

British French Korean ... Chinese

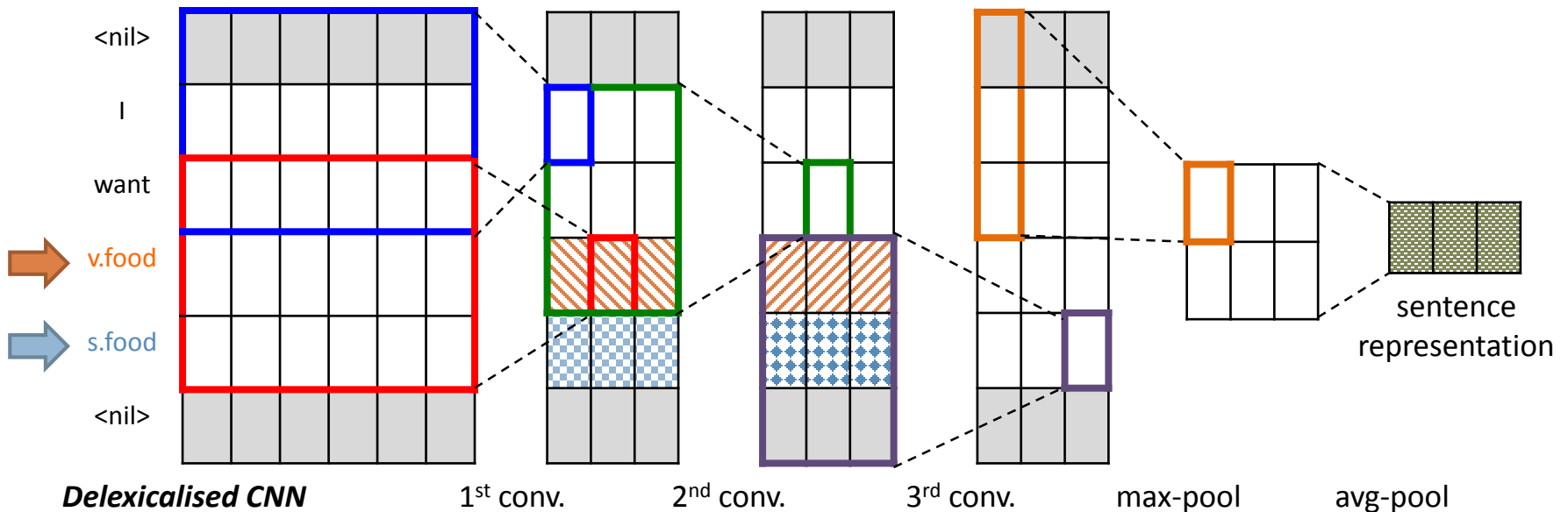
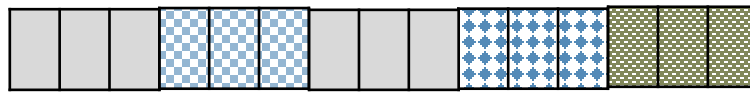
1.3

Output layer

Hidden layer

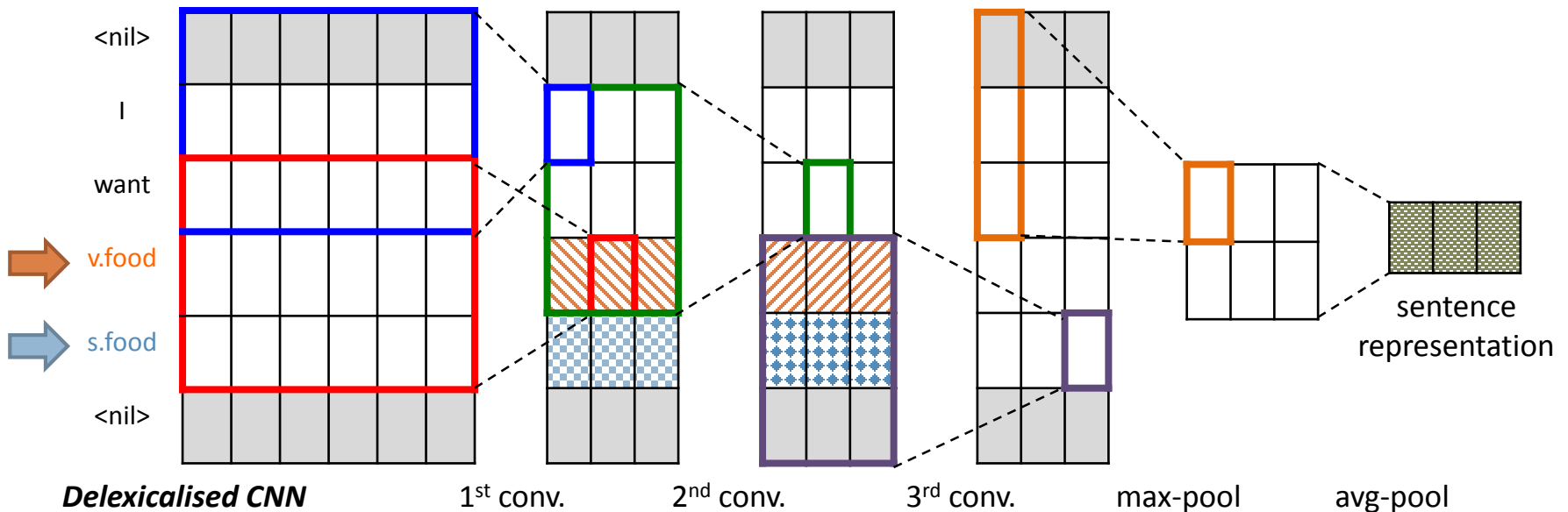
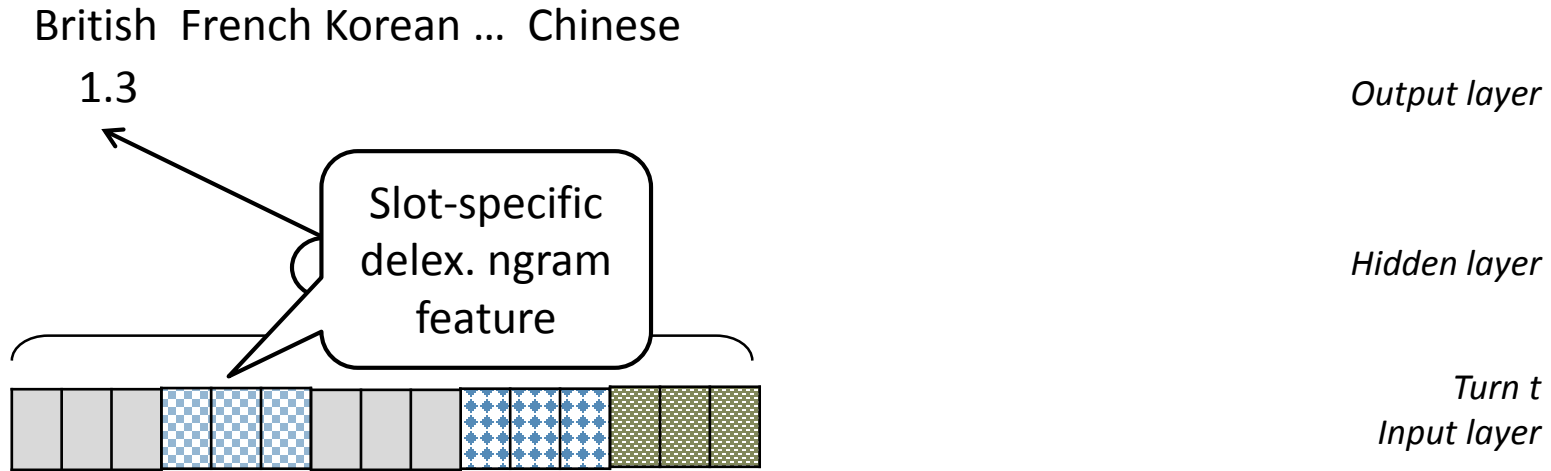
Turn t

Input layer



Jordan RNN-CNN belief trackers

33



Jordan RNN-CNN belief trackers

33

British French Korean ... Chinese

1.3

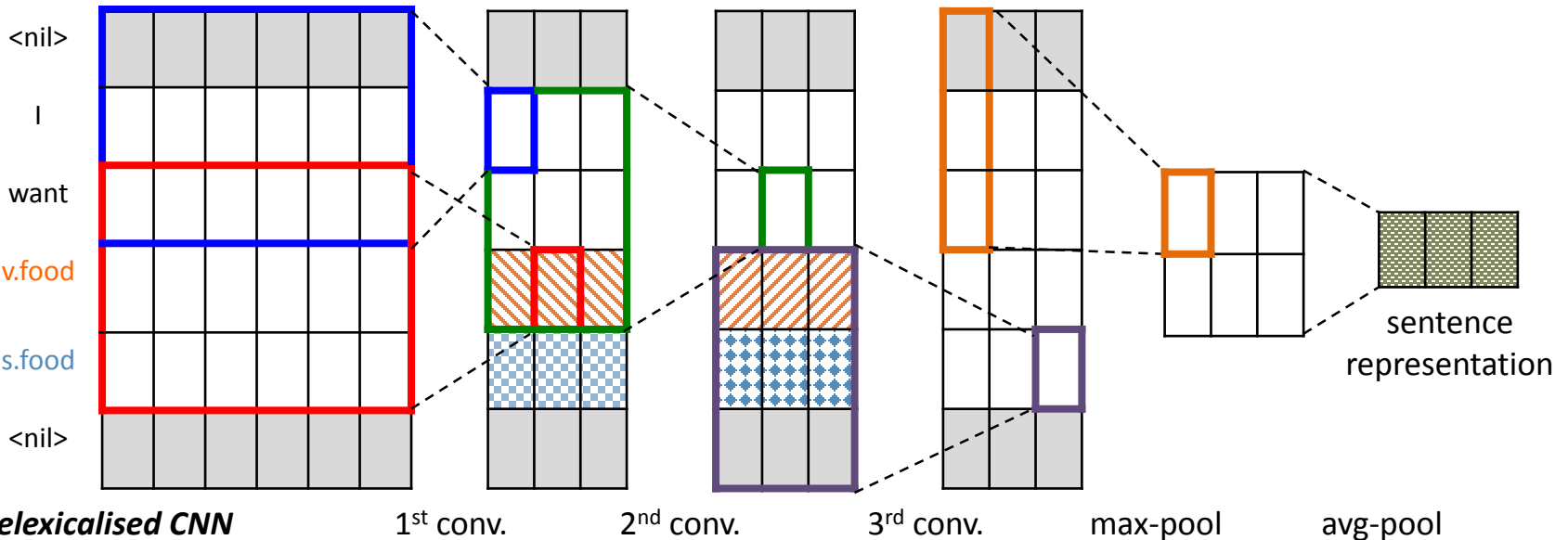
Value-specific
delex. ngram
placeholder

Slot-specific
delex. ngram
feature

Output layer

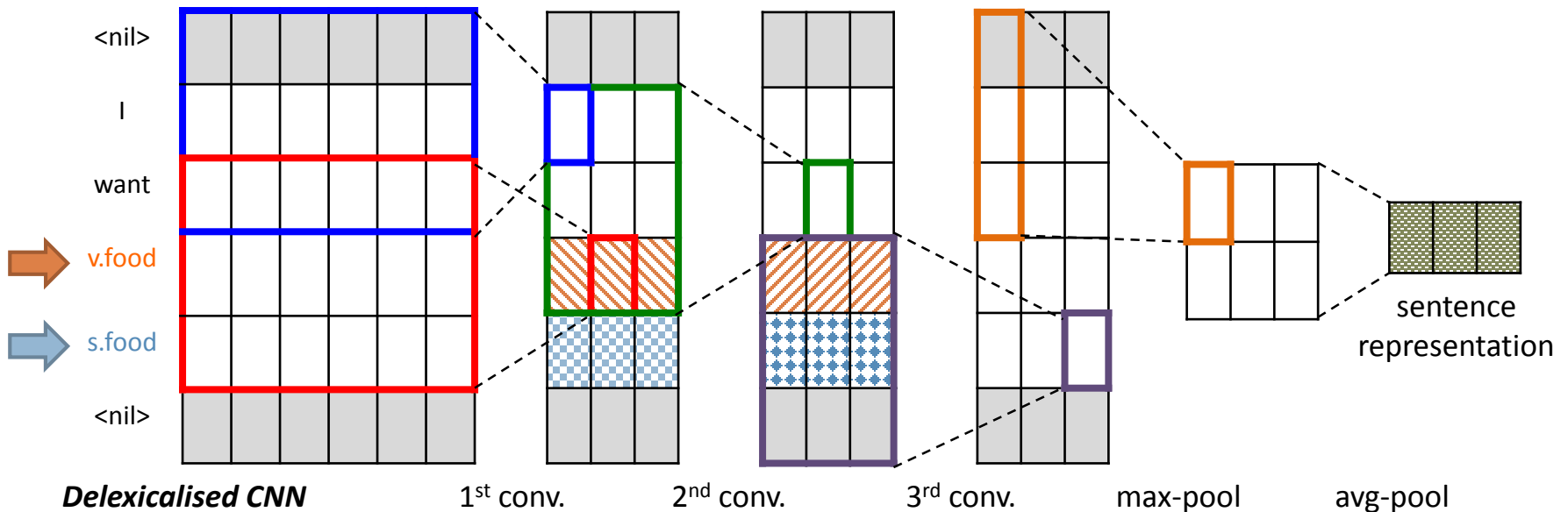
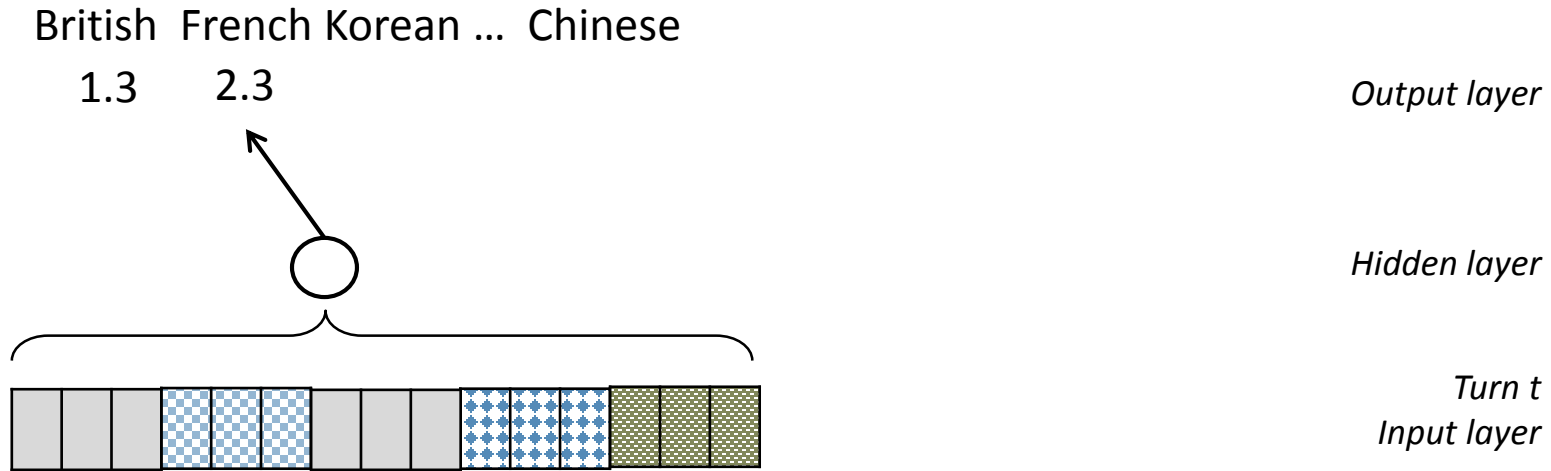
Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

33



Jordan RNN-CNN belief trackers

33

British French Korean ... Chinese

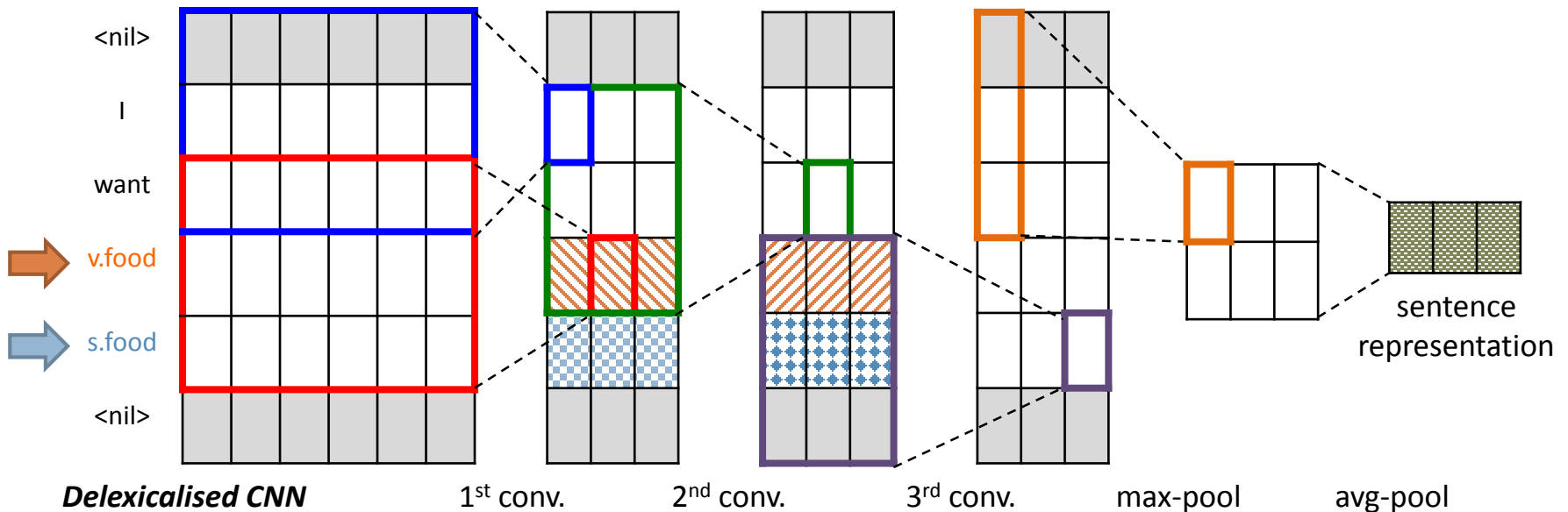
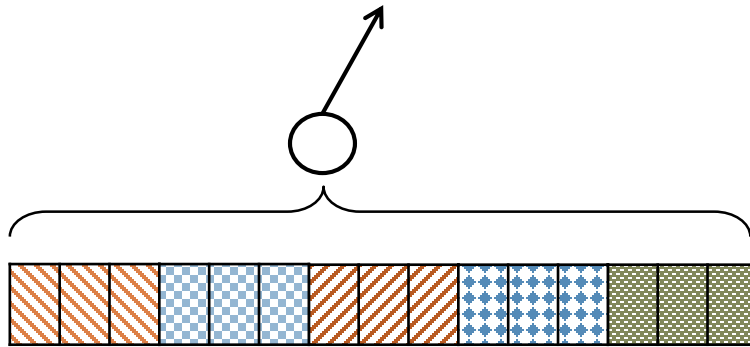
1.3 2.3 9.7

Output layer

Hidden layer

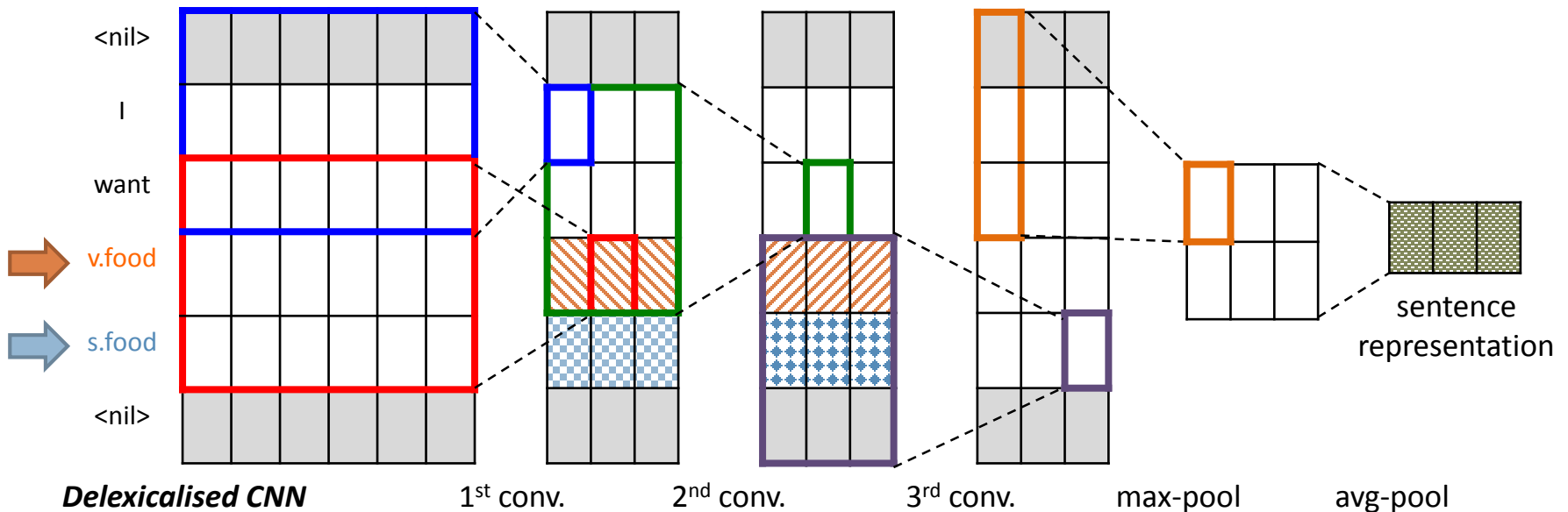
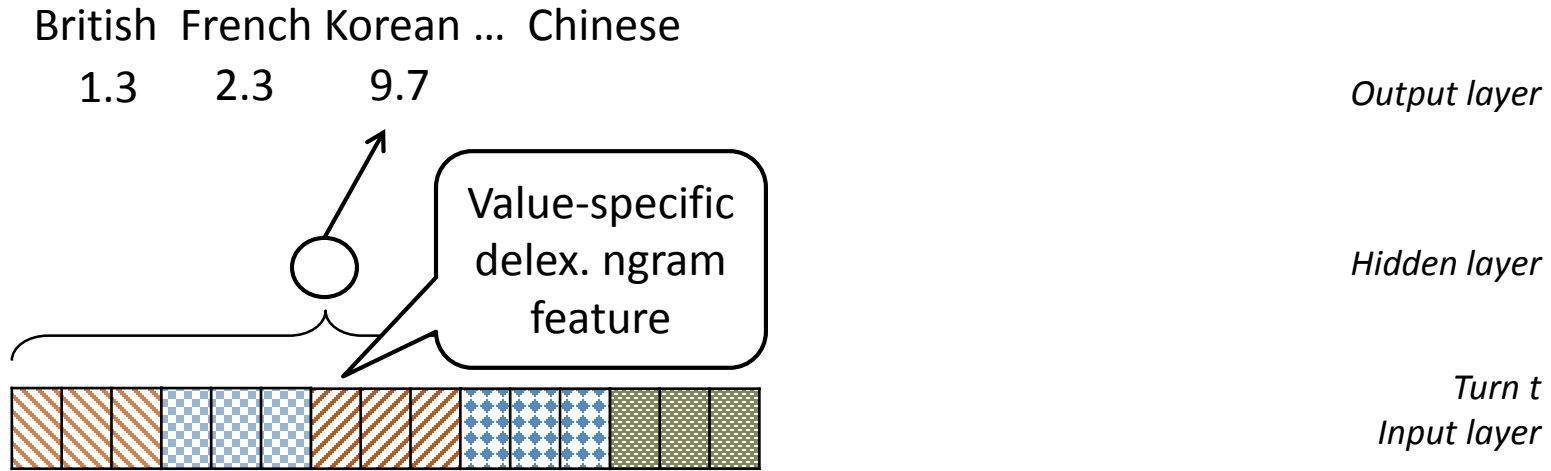
Turn t

Input layer



Jordan RNN-CNN belief trackers

33



Jordan RNN-CNN belief trackers

33

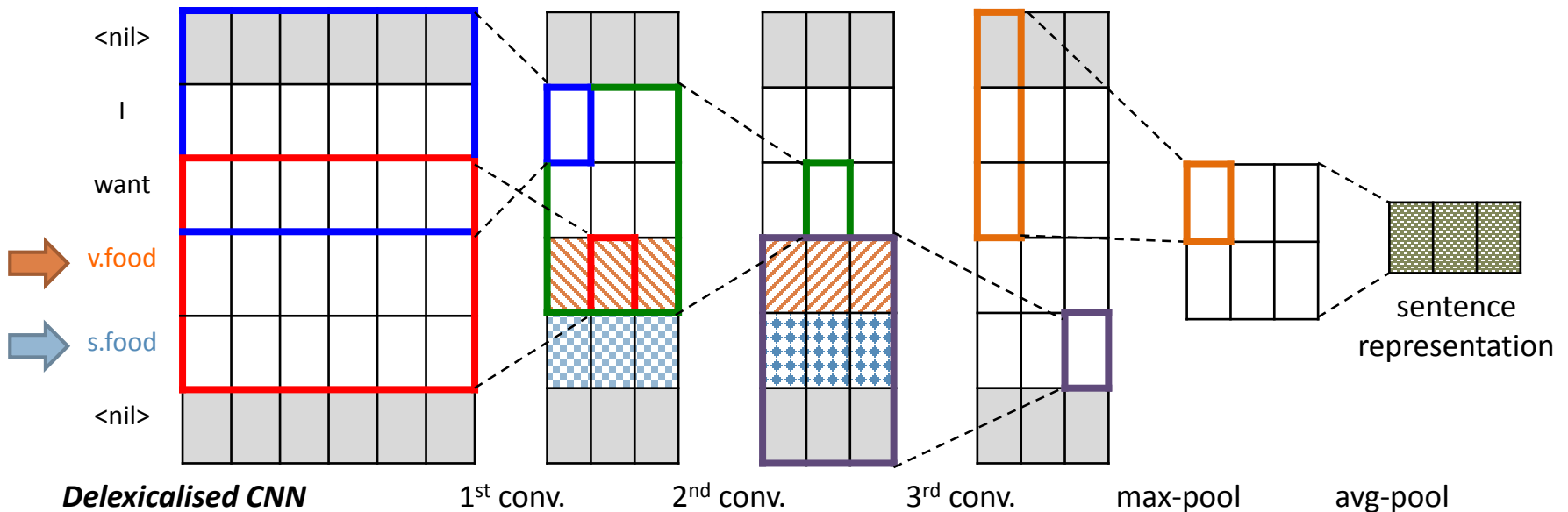
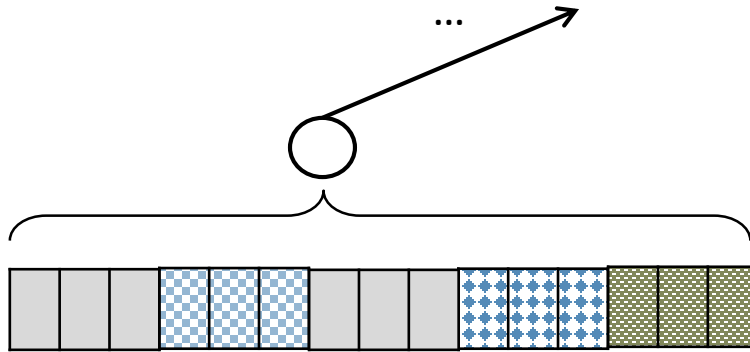
British French Korean ... Chinese

1.3 2.3 9.7 ... 1.2

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

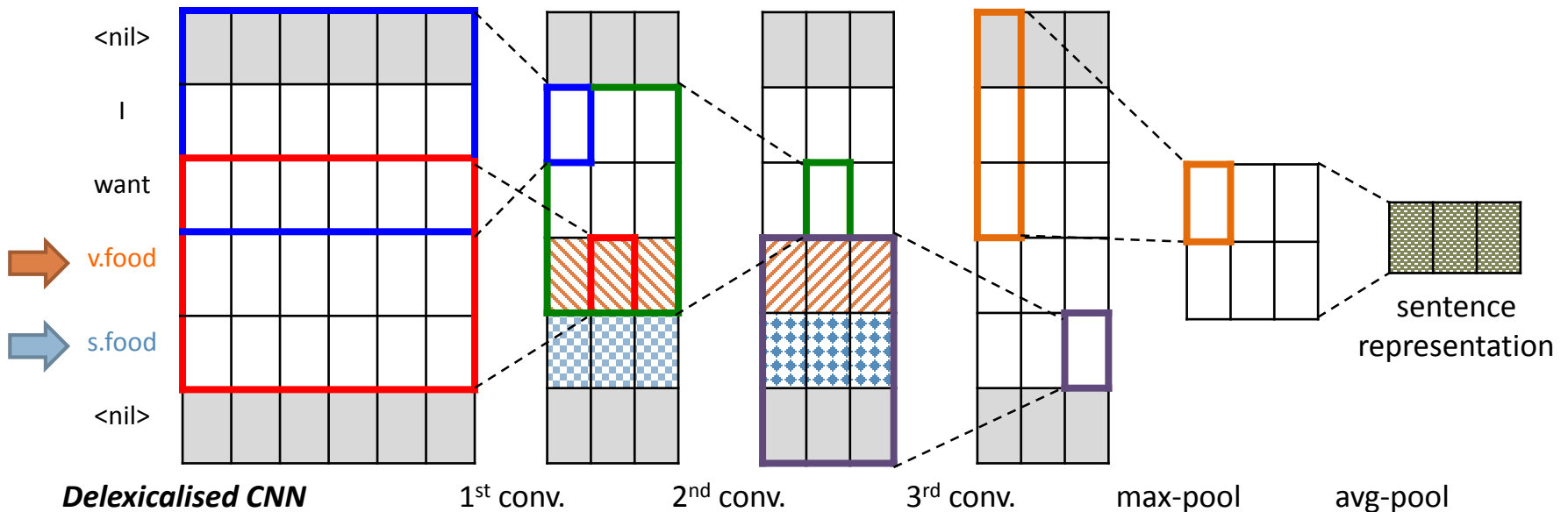
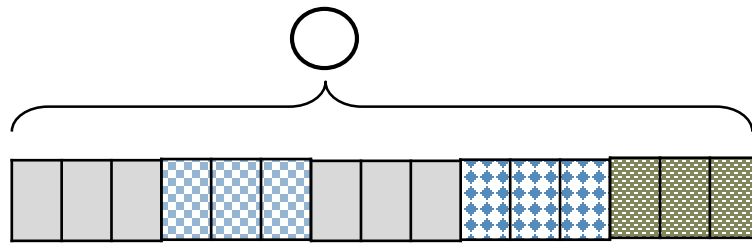
33

British French Korean ... Chinese
.01 .02 .85 .01
...

Output layer

Hidden layer

Turn t
Input layer



Jordan RNN-CNN belief trackers

33

British French Korean ... Chinese

.01 .02 .85 .01

[Henderson et al, 2014]

Output layer

...

Hidden layer

Turn t
Input layer

