

User Modeling in Language Learning with Macaronic Texts

Adithya Renduchintala

Rebecca Knowles

Philipp Koehn

Jason Eisner



JOHNS HOPKINS
UNIVERSITY

The German BookSeller's Peace Prize was awarded to Navid Kermani in June last year. Navid Kermani was geboren in Germany. He started his career as a reporter for a popular newspaper. Later he studied philosophy, drama and oriental studies...

*The German BookSeller's Peace Prize was awarded to Navid Kermani in June last year. **Navid Kermani was geboren in Germany.** He started his career as a reporter for a popular newspaper. Later he studied philosophy, drama and oriental studies.*

The German BookSeller's Peace Prize was awarded to Navid Kermani in June last year. Navid Kermani was geboren in Germany. He started his career as a reporter for a popular newspaper. Later he studied philosophy, drama and oriental studies.

Lernen by Immersion

- We know this works to acquire L1 (native)

Lernen by Immersion

- We know this works to acquire L1
- We also know this to work to acquire L2 (second language)
 - “It is widely agreed that much second language vocabulary learning occurs incidentally while the learner is engaged in extensive reading.” (Huckin & Coady, 1999)

Lernen by Immersion

If you want to learn German
start reading German!

Lernen by Immersion

If you want to learn German
start reading German!

Some obvious problems...

Lernen by Immersion



Lernen by Immersion



Lernen by Immersion

Can we leverage **mixing L1 and L2** to learn new L2 vocabulary?

Lernen by Immersion

Navid Kermani was geboren in Germany.

Lernen *by Immersion*

Navid Kermani was geboren *in Germany.*

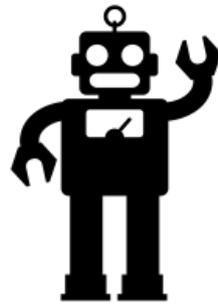
Macaronic Text

- Macaronic:
 - Of or containing a mixture of vernacular words with Latin words or with vernacular words given Latinate endings: macaronic verse.
 - Of or involving a mixture of two or more languages.
- Like code-switching but more deliberate and often for humor.

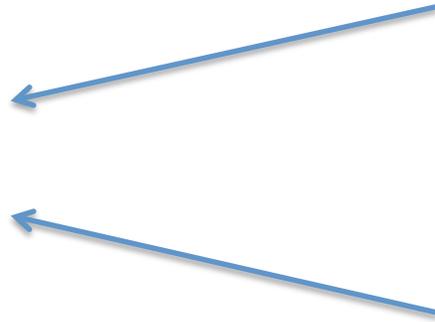
Research Goal

We want to investigate **macaronic immersion** as a tool for **language learning**.

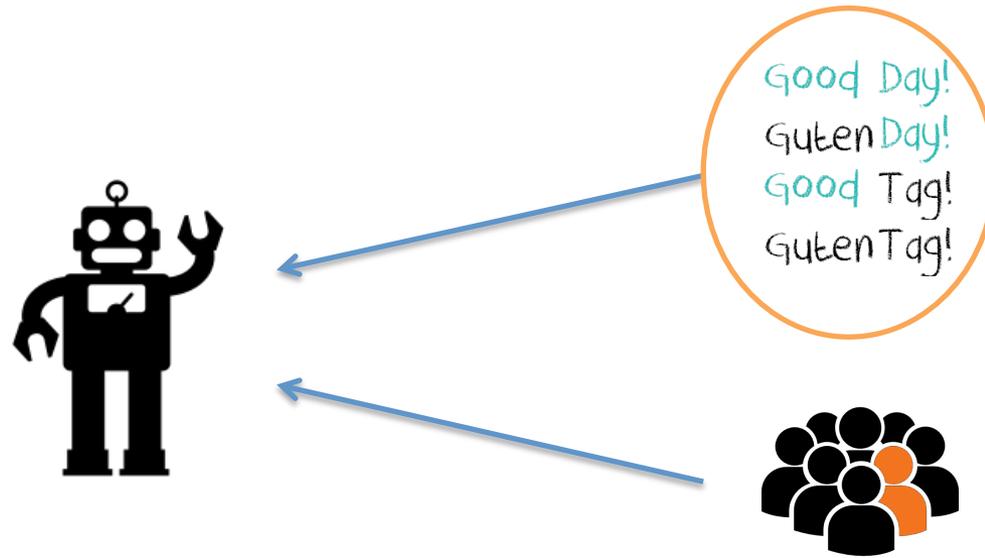
System: Components



Good Day!
Guten Day!
Good Tag!
Guten Tag!



System: Components



Generate a spectrum of macaronic content.

System: Components

- [Demonstration of Macaronic Interface](#)

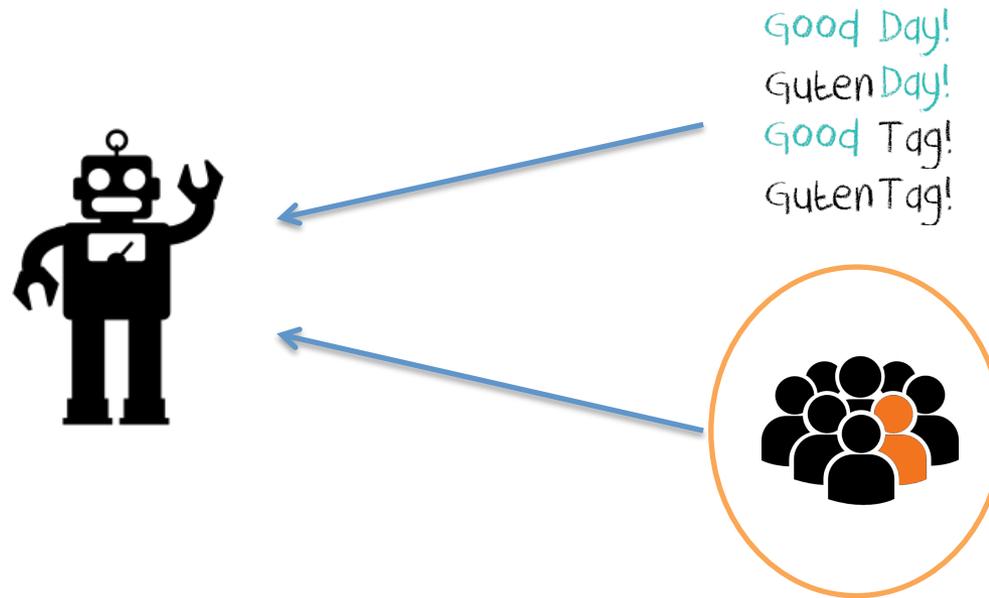
System: Components



Completely in
English

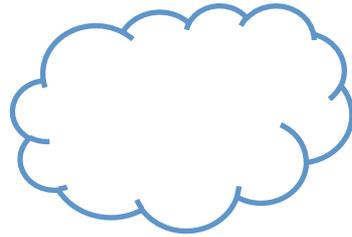
Komplett in
Deutsch!

System: Components



Model the learner and present content to their level.

Modeling Learner Comprehension



The *police* verhaftete *the* *bank* *robber*

Modeling Learner Comprehension



The *police* verhaftete *the* *bank* *robber*

Modeling Learner Comprehension



The *police* verhaftete *the* *bank* *robber*

Modeling Learner Comprehension

English Guess – Foreign Word Factor
'EF' Factor

$$\psi = \exp(\theta \cdot \phi(e, f))$$



The police

verhaftete

the

bank

robber

Modeling Learner Comprehension

English Guess – Foreign Word Factor
'EF' Factor

$$\psi^{ef} = \exp(\theta^{ef} \cdot \phi^{ef}(e, f))$$



The police

verhaftete

the

bank

robber

Modeling Learner Comprehension



The police

verhaftete

the

bank

robber

English Guess – Foreign Word Factor
'EF' Factor

$$\psi^{ef} = \exp(\theta^{ef} \cdot \phi^{ef}(e, f))$$

Orthographic Similarity (e, f)
Pronunciation Similarity (e, f)

...

Modeling Learner Comprehension



The police

verhaftete

the

bank

robber

English Guess – Foreign Word Factor
'EF' Factor

$$\psi^{ef} = \exp(\theta^{ef} \cdot \phi^{ef}(e, f))$$

Weights

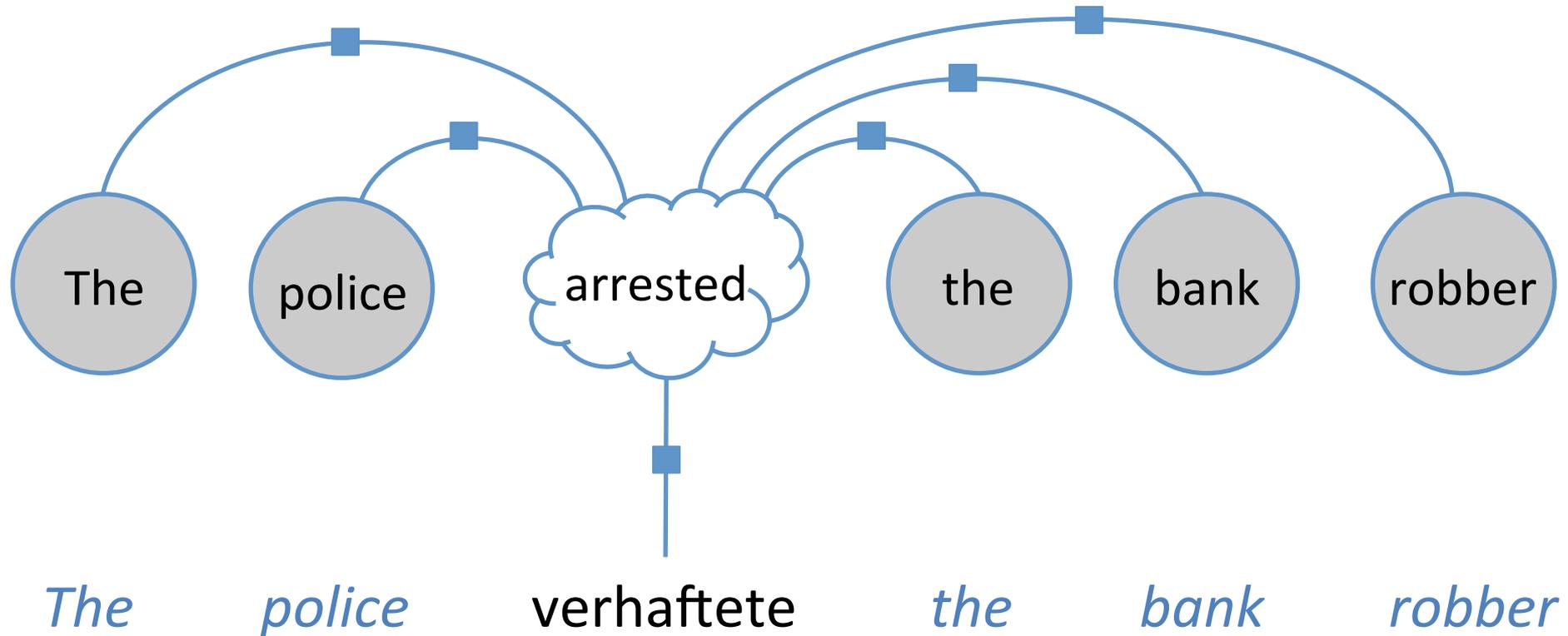
Orthographic Similarity (e, f)
Pronunciation Similarity (e, f)
...

Modeling Learner Comprehension

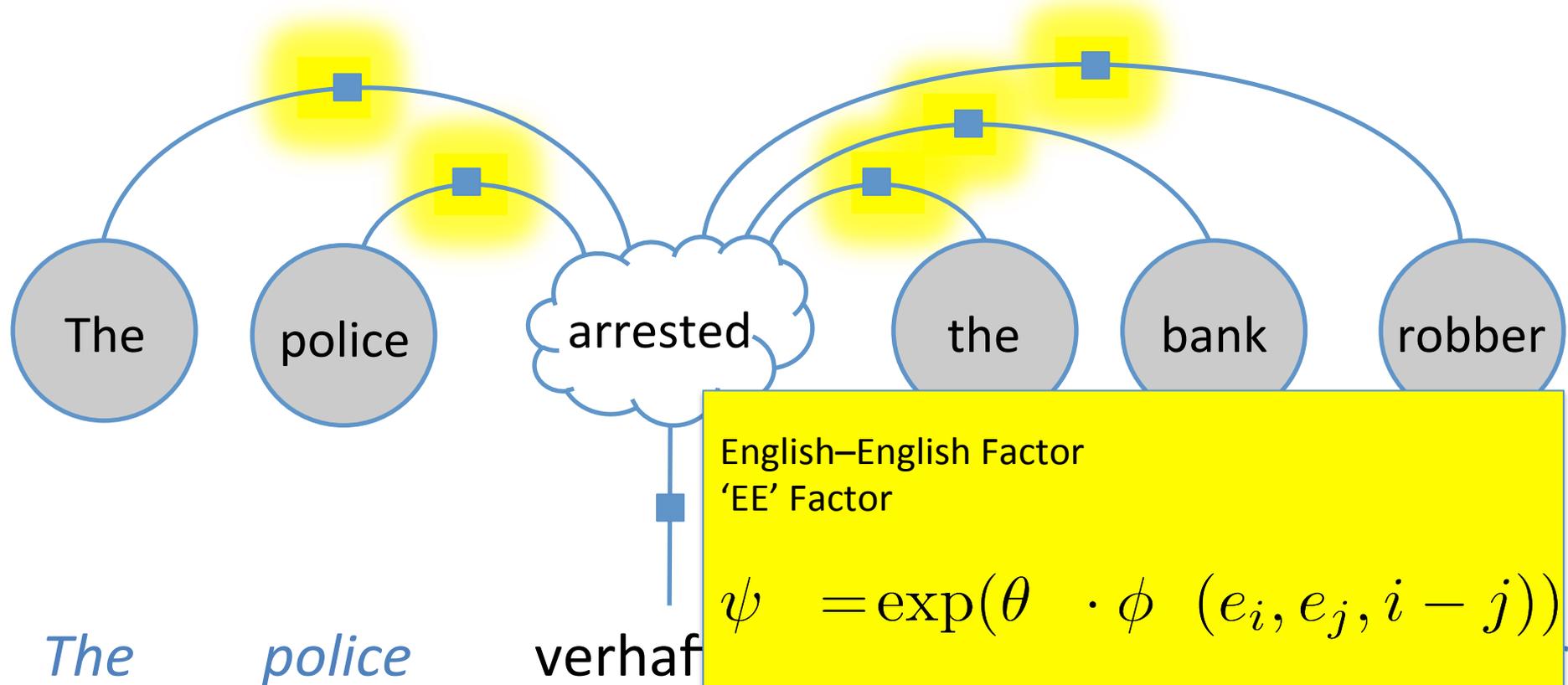


The *police* verhaftete *the* *bank* *robber*

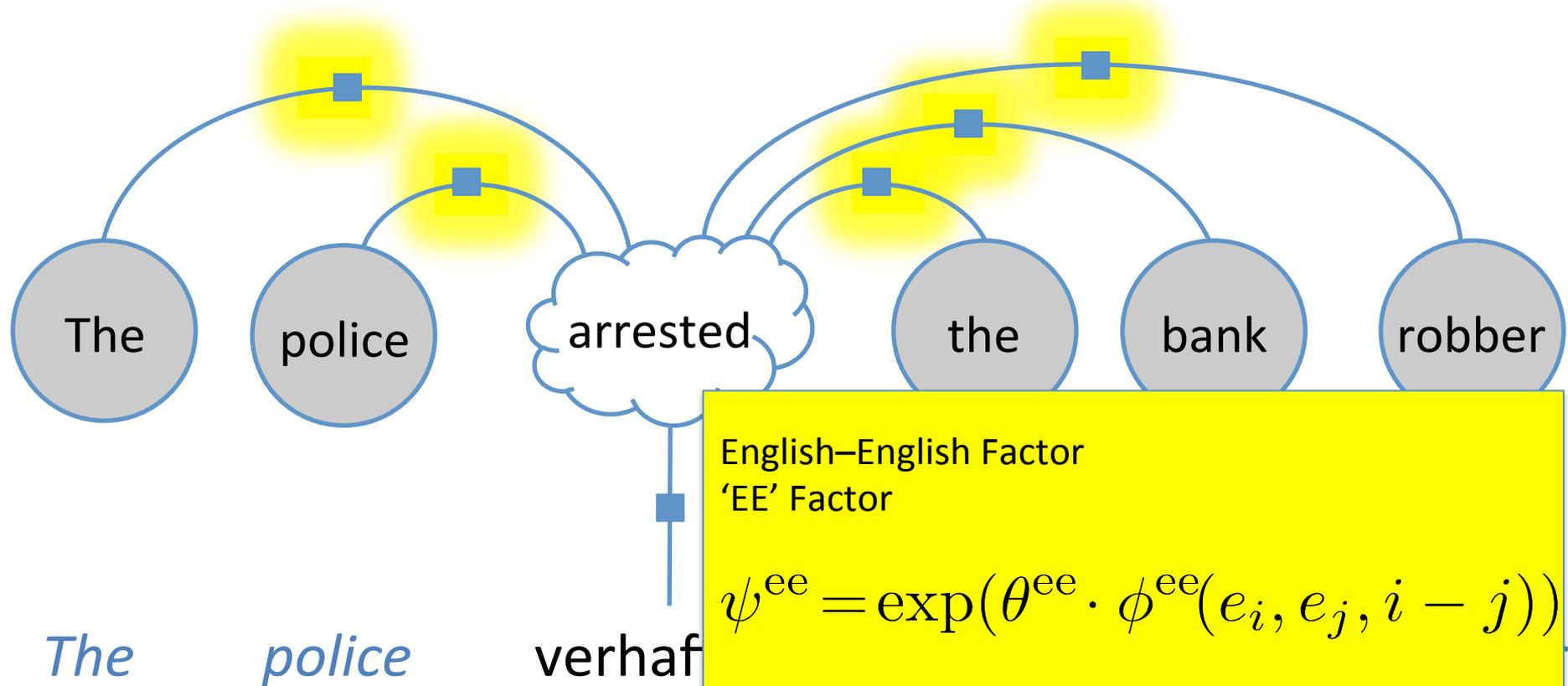
Modeling Learner Comprehension



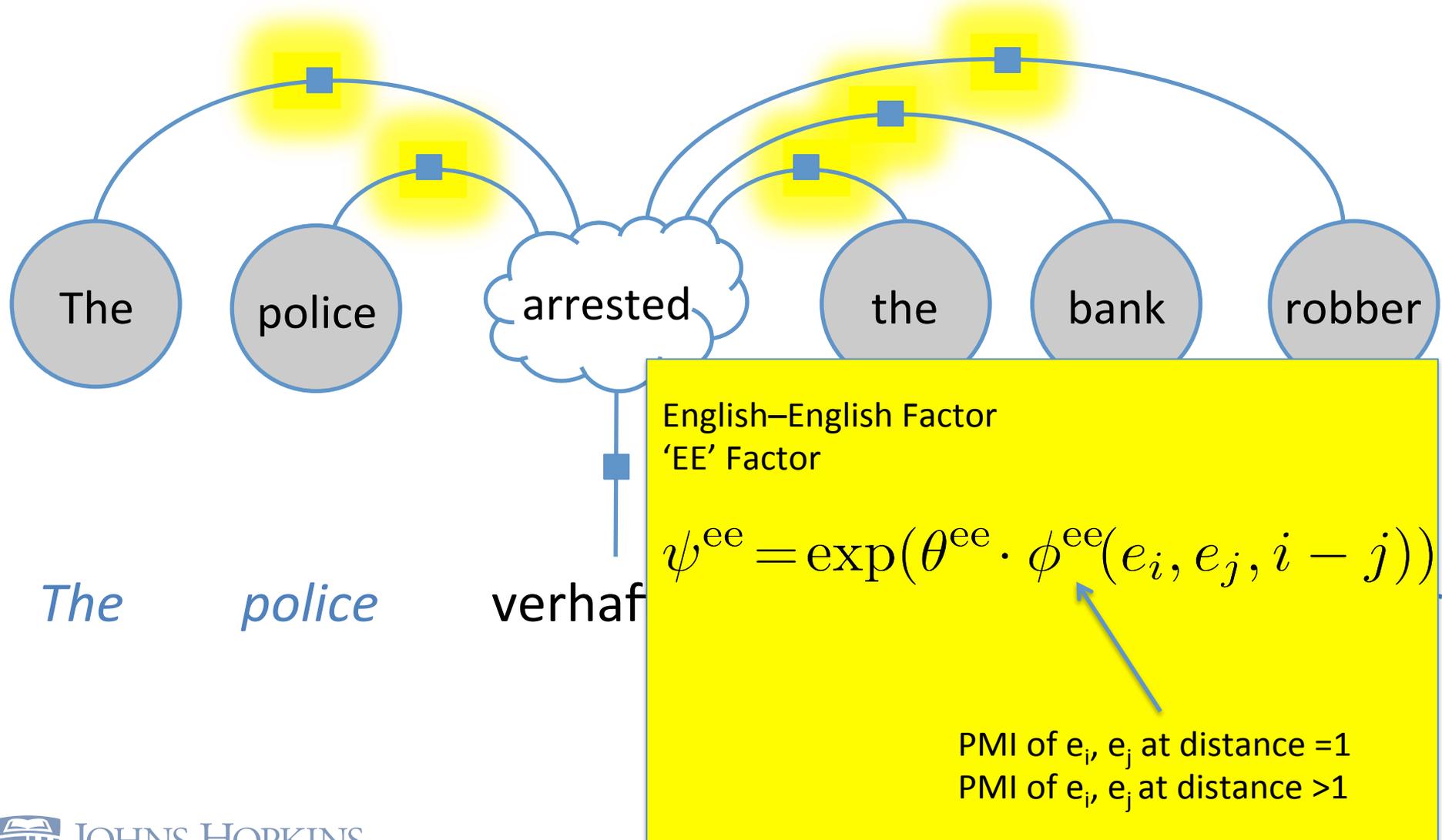
Modeling Learner Comprehension



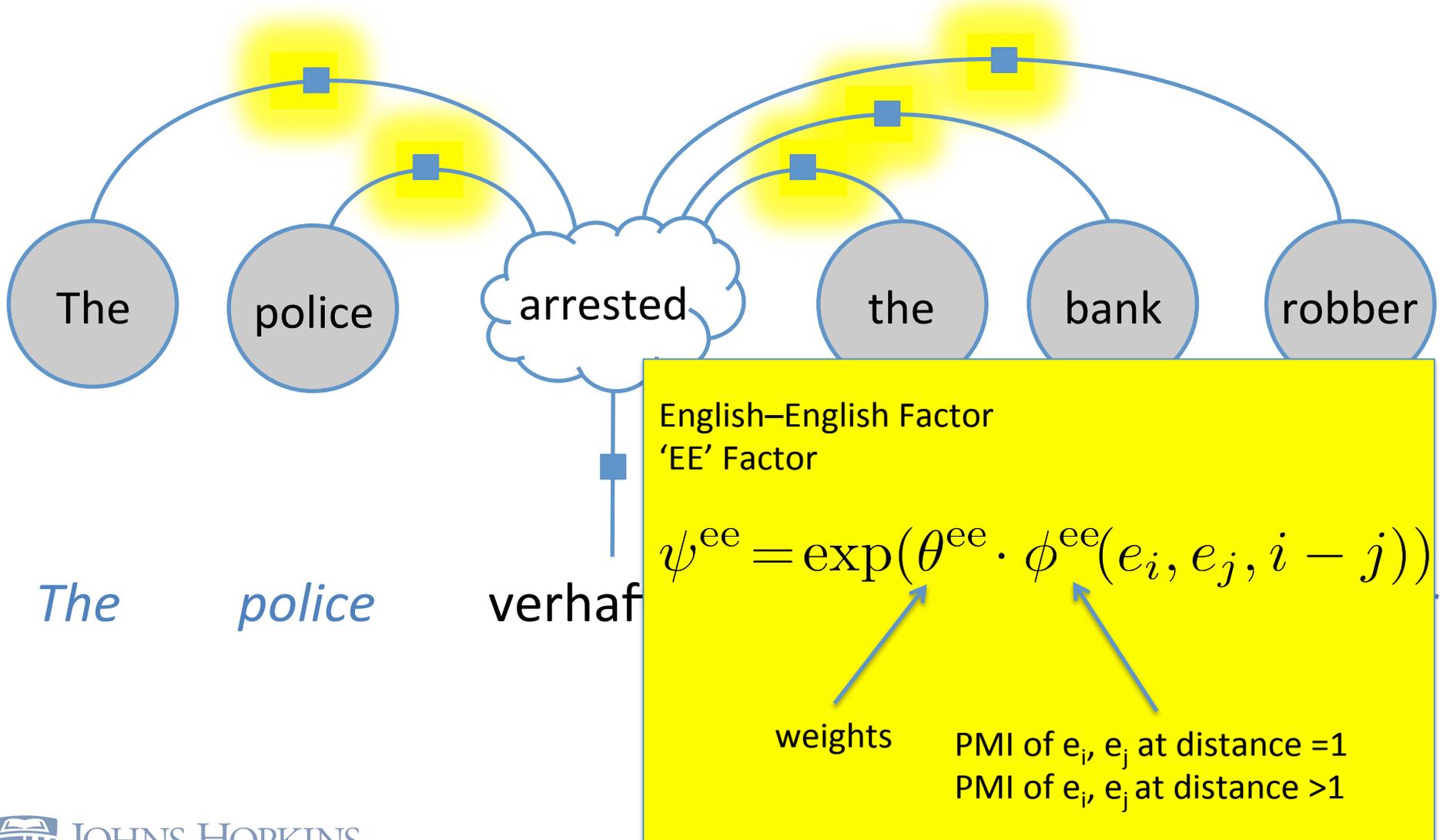
Modeling Learner Comprehension



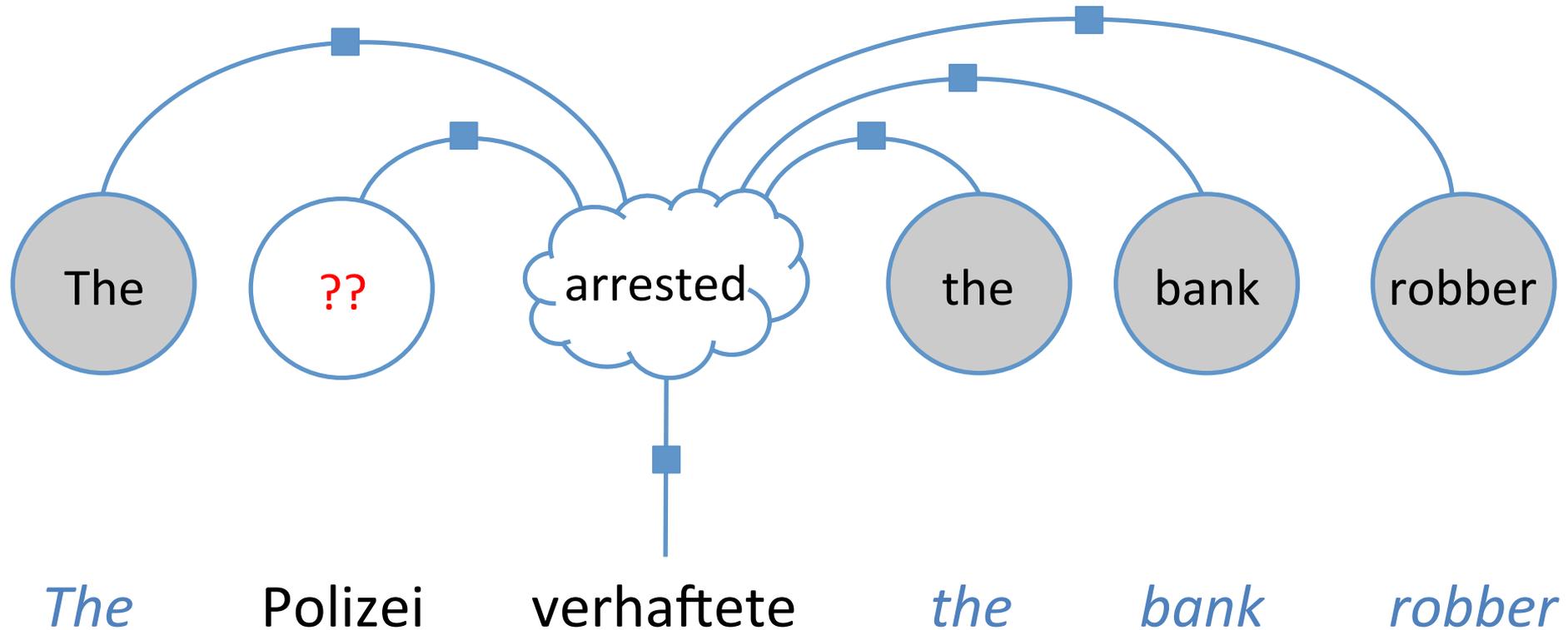
Modeling Learner Comprehension



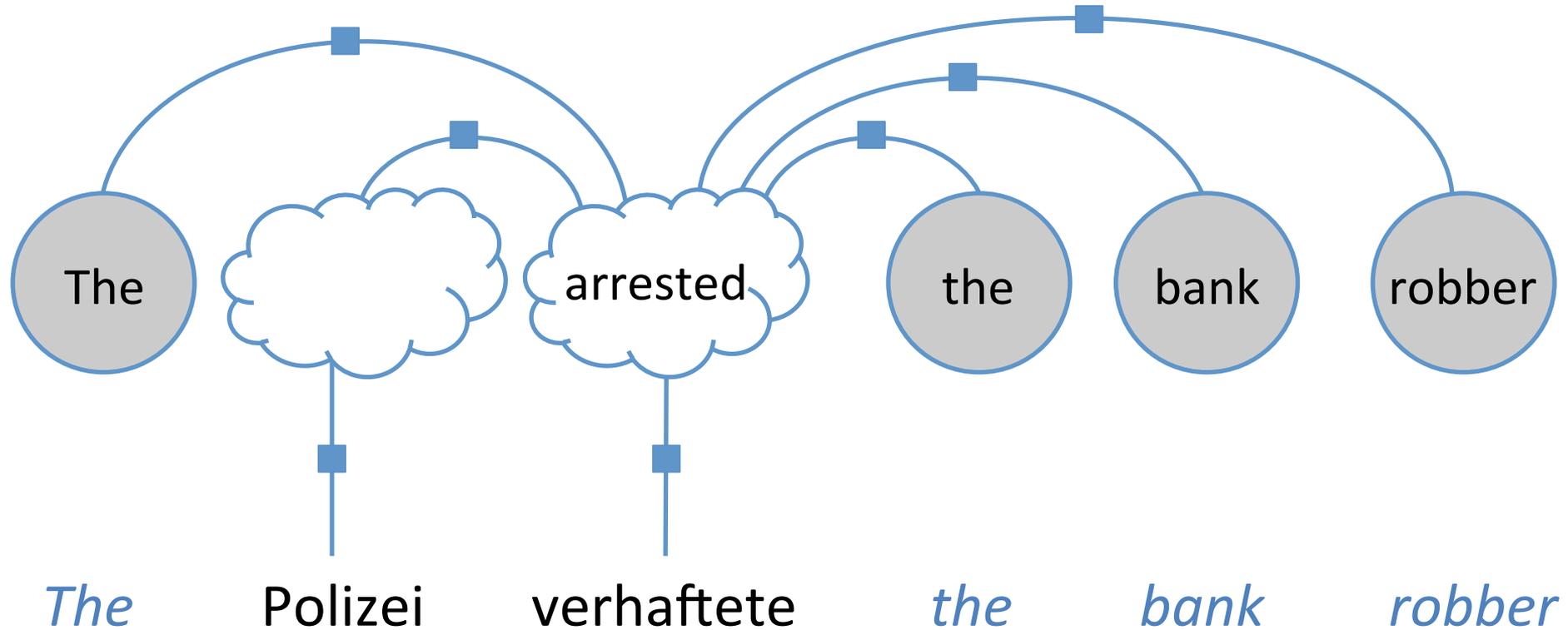
Modeling Learner Comprehension



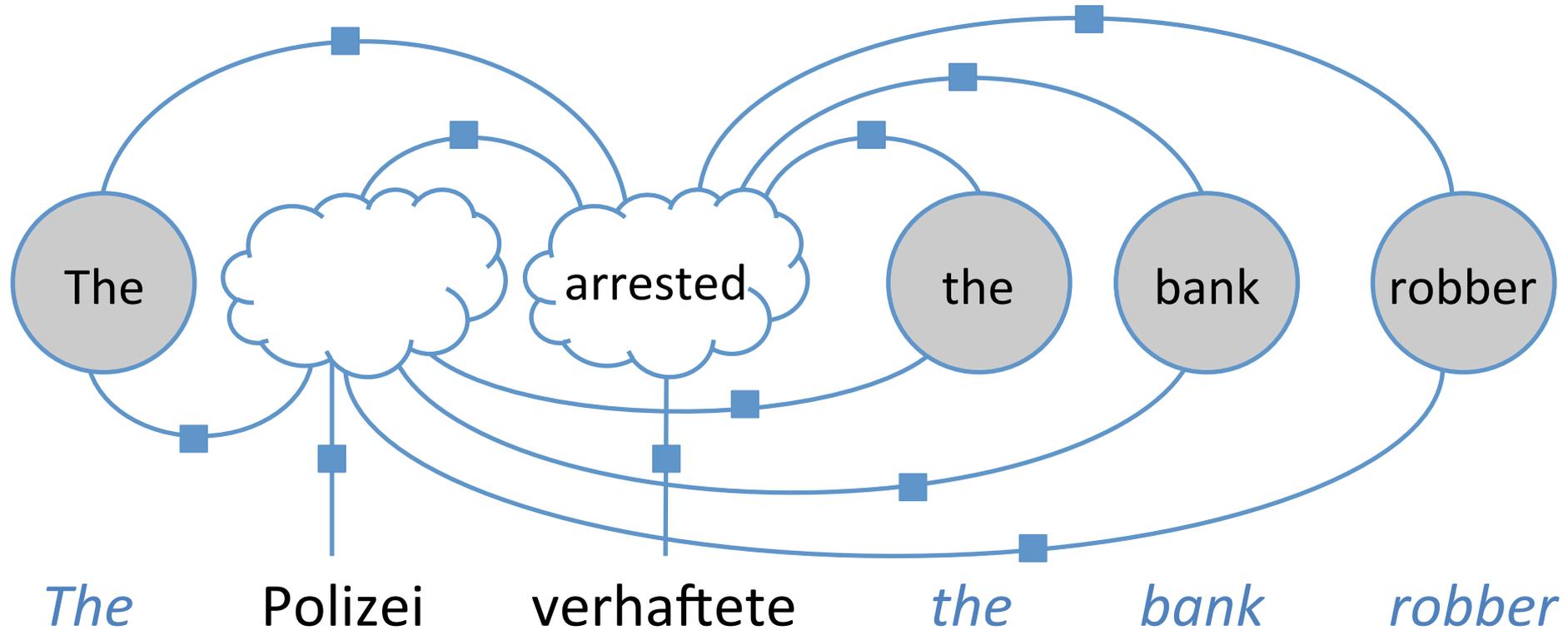
Modeling Learner Comprehension



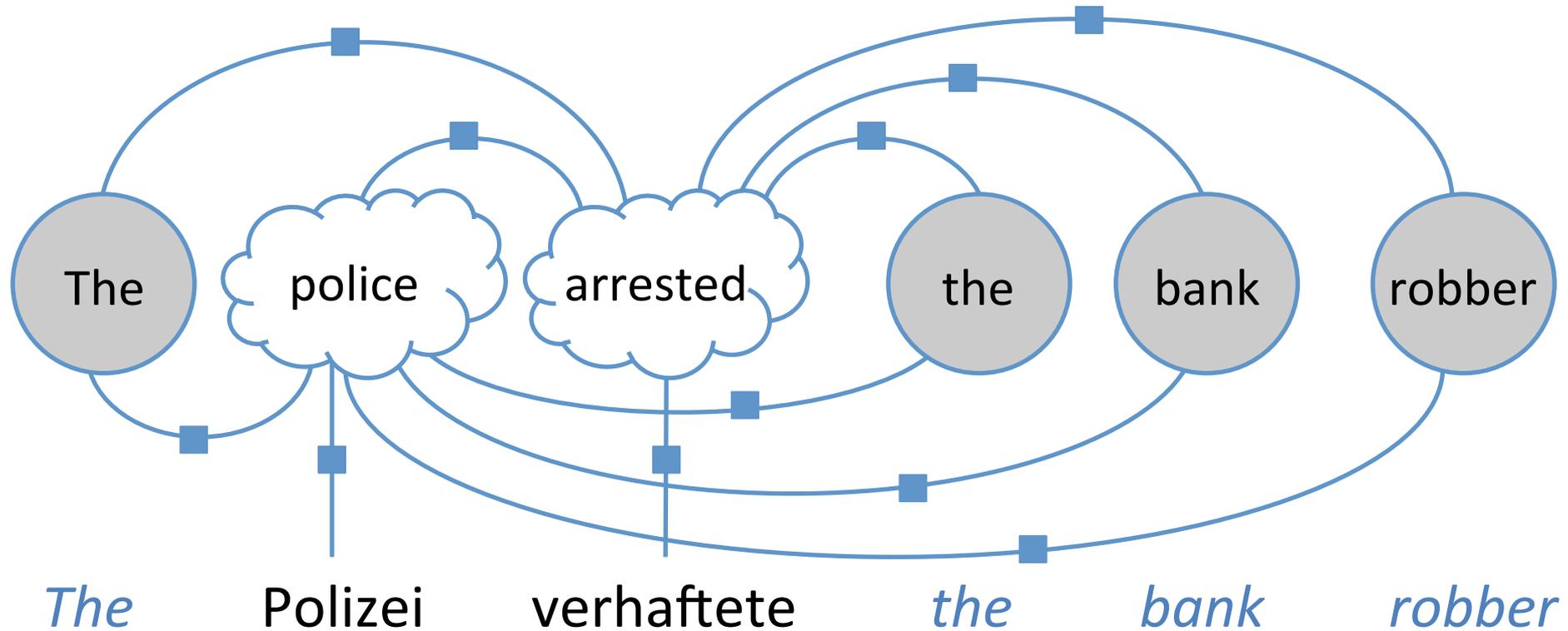
Modeling Learner Comprehension



Modeling Learner Comprehension



Modeling Learner Comprehension



Modeling Learner Comprehension

Unobserved English words Observed English words

Observed Foreign words

$$P_{\theta}(\mathbf{E} = \mathbf{e} \mid \mathbf{E}_{\text{Obs}} = \mathbf{e}_{\text{Obs}}^*, \mathbf{f}, \text{history})$$
$$\propto \prod_{i \notin \text{Obs}} (\psi^{\text{ef}}(e_i, f_i)) \cdot \prod_{j \neq i} \psi^{\text{ee}}(e_i, e_j, i - j)$$

EF Factor
Similarity of e and f

EE Factor
Contextual Influence

The diagram illustrates the components of the learner comprehension model. It shows the probability of unobserved English words given observed English words and foreign words. The model is composed of two main parts: an EF factor representing the similarity between unobserved English words and foreign words, and an EE factor representing the contextual influence between observed English words.

Modeling Learner Comprehension

- We just built a model to jointly translate the German words in context.
- ~~• To get “best” prediction accuracy:
 - Add better features, dictionaries, MT system, ...~~
- To match a naïve human’s guesses:
 - Use only features available to naïve humans.
 - Train to match “actual” human guesses

Modeling Learner Comprehension

- So how do we get the training data?
- [Demo of Data Collection.](#)

Modeling Learner Comprehension

- Note on history features

History+



$$P_{\theta}(\mathbf{E} = \mathbf{e} \mid \mathbf{E}_{\text{Obs}} = \mathbf{e}_{\text{Obs}}^*, \mathbf{f}, \text{history})$$

$$\propto \prod_{i \notin \text{Obs}} (\psi^{\text{ef}}(e_i, f_i)) \cdot \prod_{j \neq i} \psi^{\text{ee}}(e_i, e_j, i - j)$$

Modeling Learner Comprehension

- Note on history features

History+ History-



$$P_{\theta}(\mathbf{E} = \mathbf{e} \mid \mathbf{E}_{\text{Obs}} = \mathbf{e}_{\text{Obs}}^*, \mathbf{f}, \text{history})$$
$$\propto \prod_{i \notin \text{Obs}} (\psi^{\text{ef}}(e_i, f_i)) \cdot \prod_{j \neq i} \psi^{\text{ee}}(e_i, e_j, i - j)$$

Modeling Learner Comprehension

- Loopy Belief Propagation for Inference
 - 3 iterations in Loopy cases
 - Single iteration of message passing with ≤ 2 hidden variables
 - Tree-Like message passing schedule (Dryer & Eisner 2009)
- Optimization using SGD
 - L2 Regularization
 - 3 Epochs
 - learning rate 0.1
 - regularization 0.2
 - Parallelized using Hogwild! Algorithm (Recht et al 2011)

Preliminary Results

- 6K, 2K, 2K train, dev and test instances
- English Vocabulary Size 5K types
- German Vocabulary Size 639 types

Top K	Recall at K
1	16.14
25	35.56
50	40.30

Alternate Evaluation?

The policeman verhaftete the bank robber.

Alternate Evaluation?

Reference



arrested

The policeman verhaftete the bank robber.

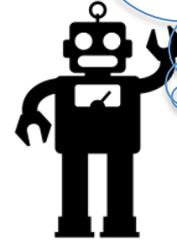
Alternate Evaluation?

Model's
Predictions



chased -1.210
arrested -1.6034
shot -3.3206
verified -5.552

...

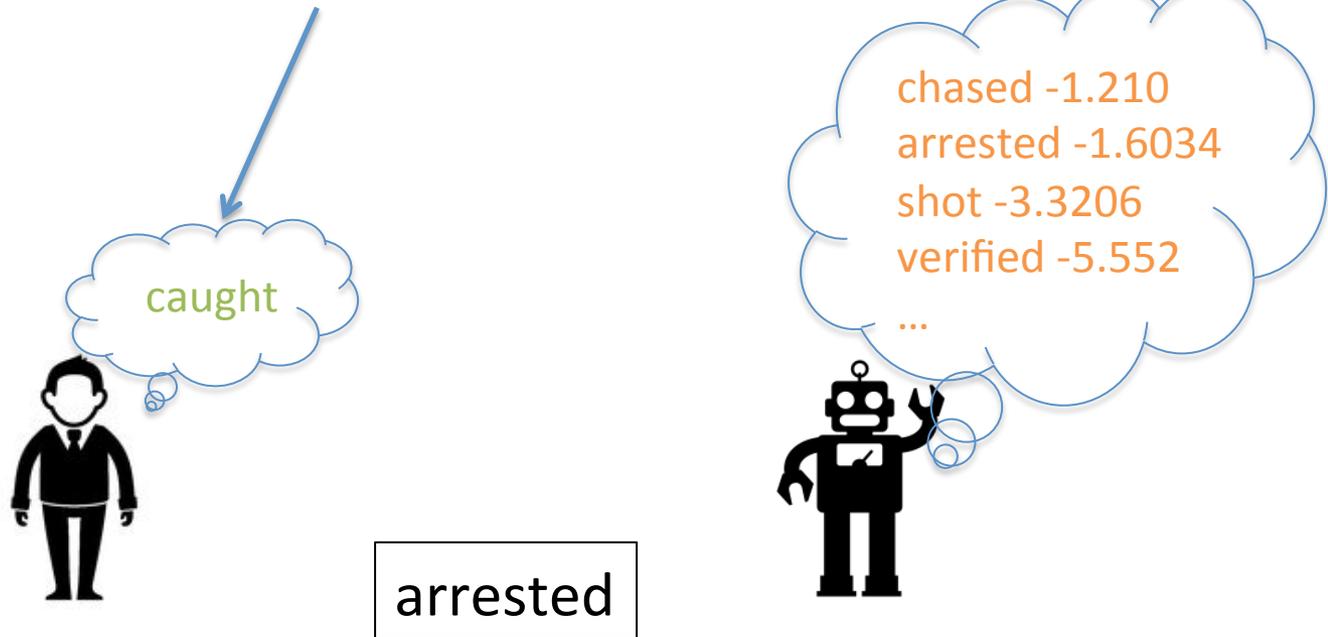


arrested

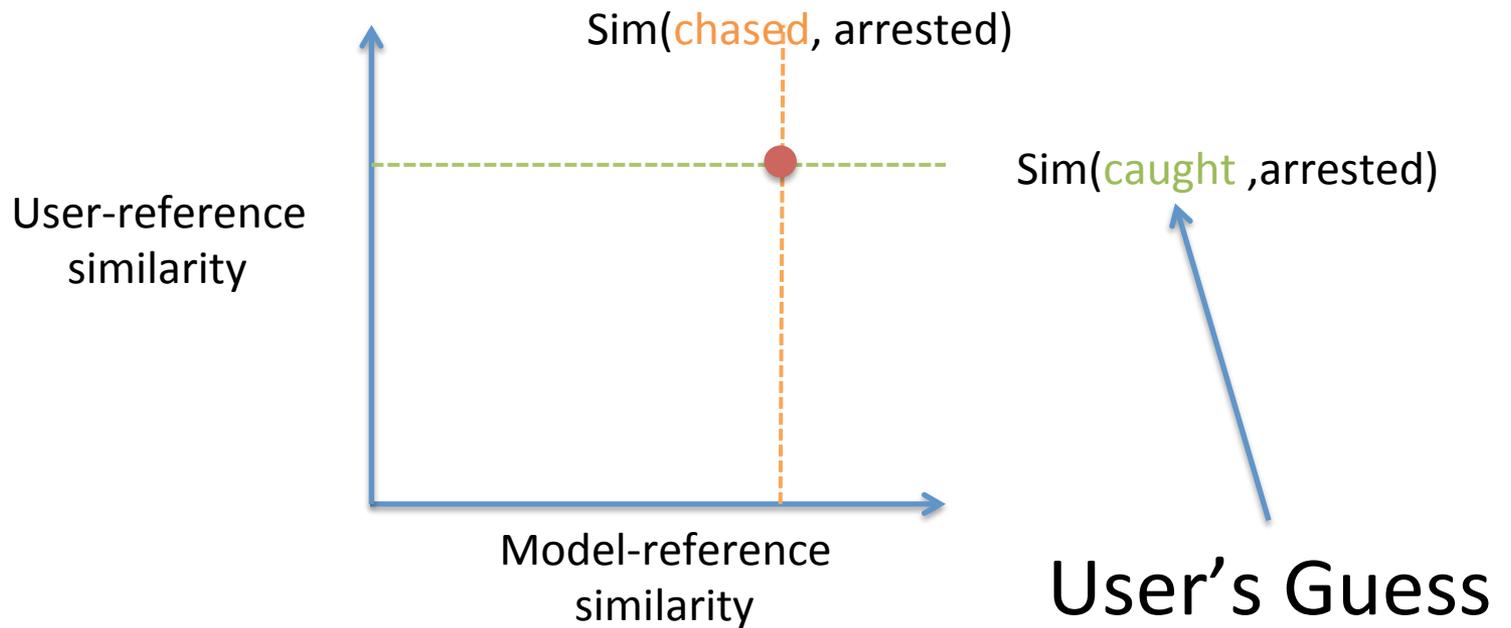
The policeman verhaftete the bank robber.

Alternate Evaluation?

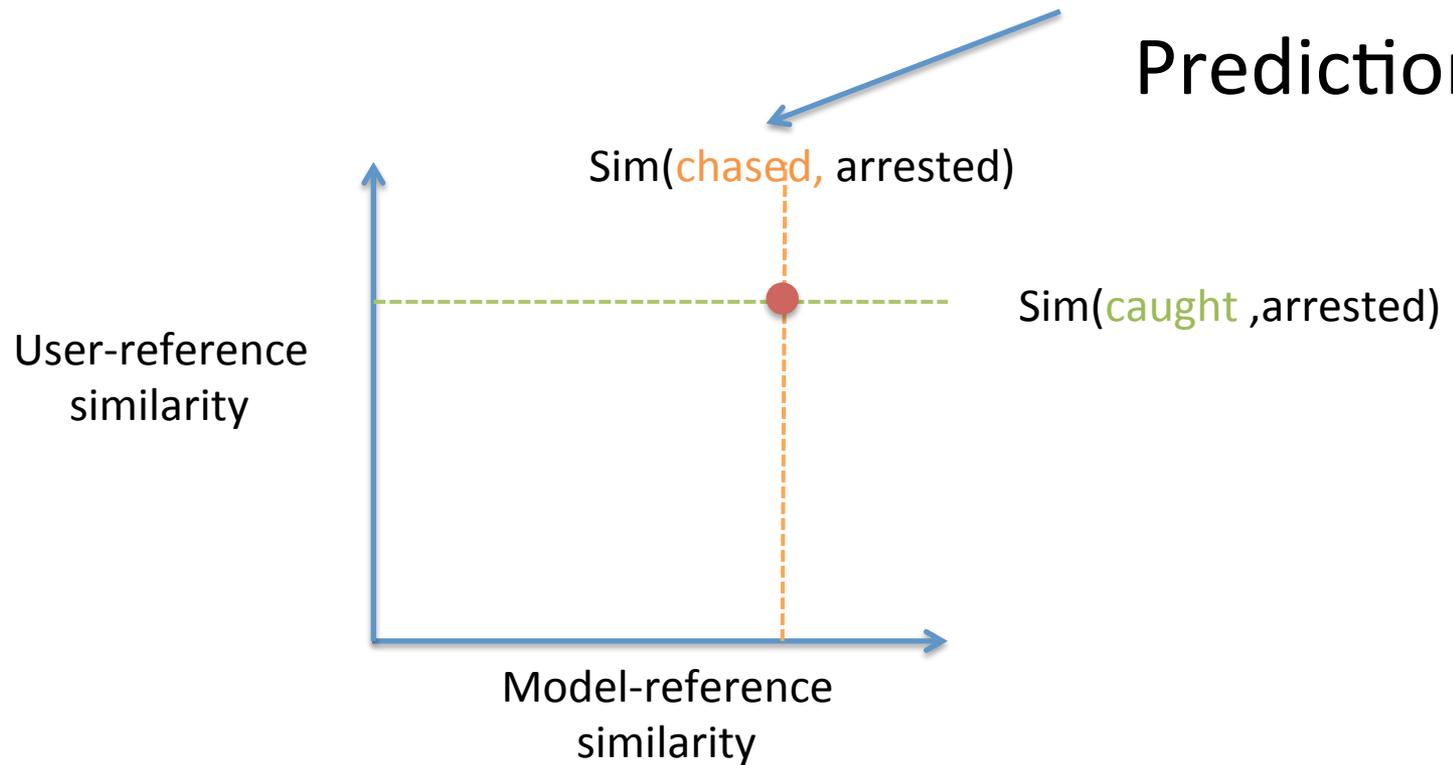
User's Guess

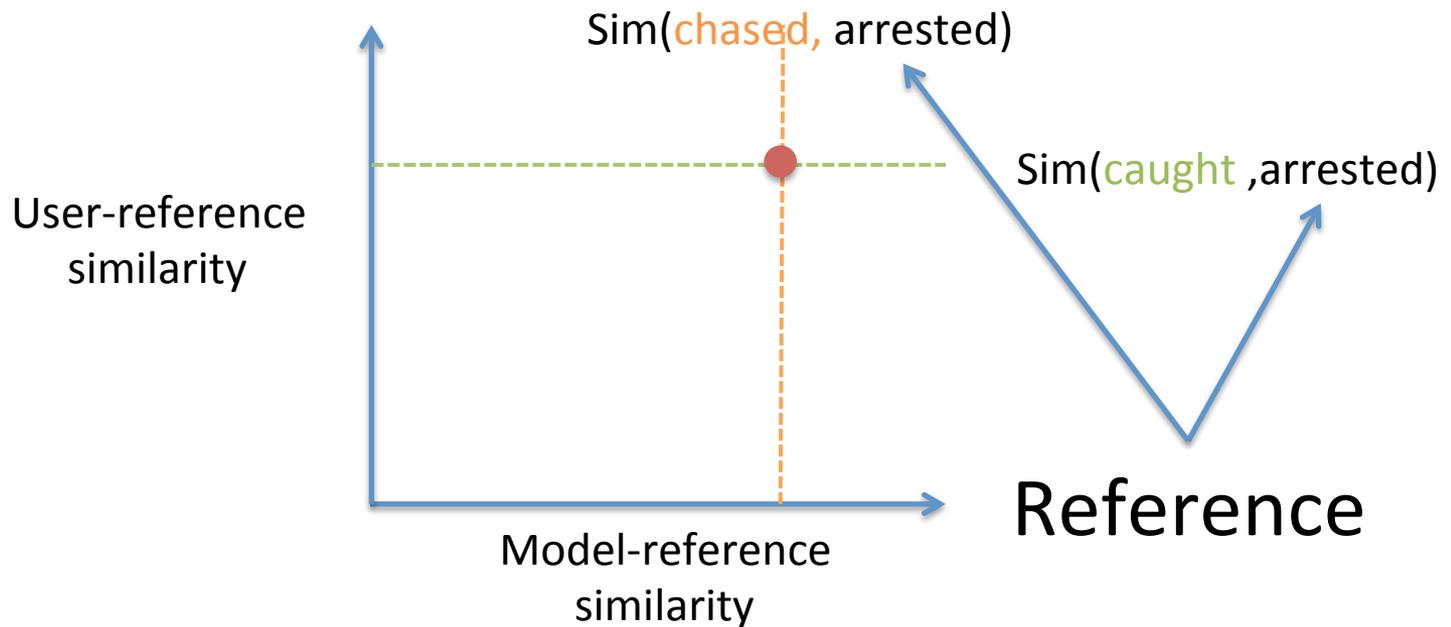


The policeman verhaftete the bank robber.



Model's Prediction

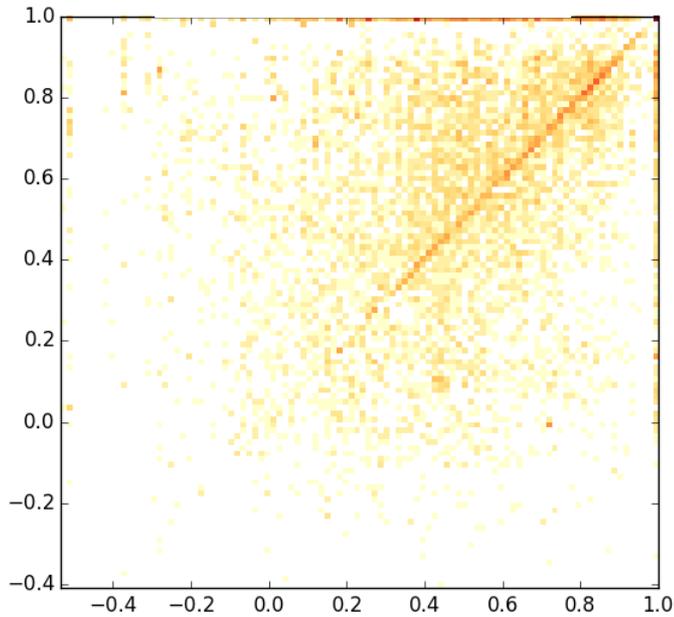




Used cosine similarity for “Sim” function with pre-trained GLoVe word embeddings

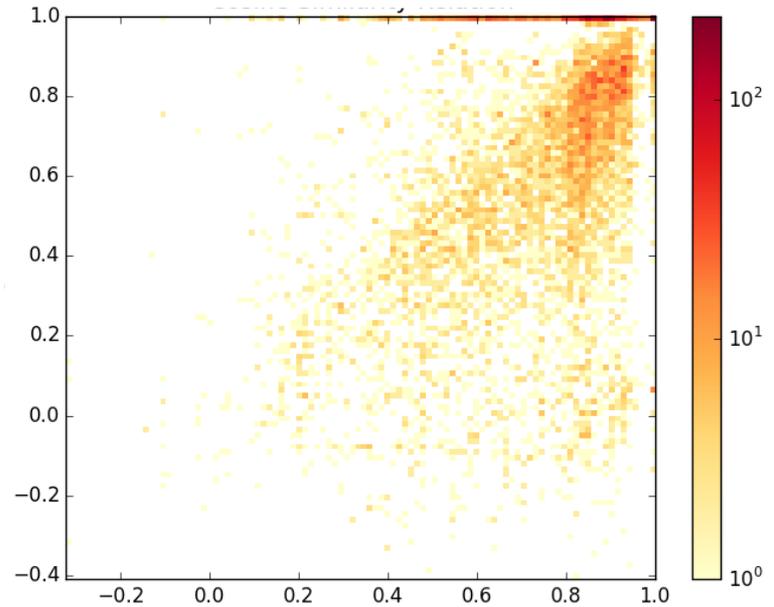
Preliminary Results

User-ref
similarity



Model-ref
similarity

Quality Corr=0.379



Expected
Model-ref
similarity

Quality Corr=0.525

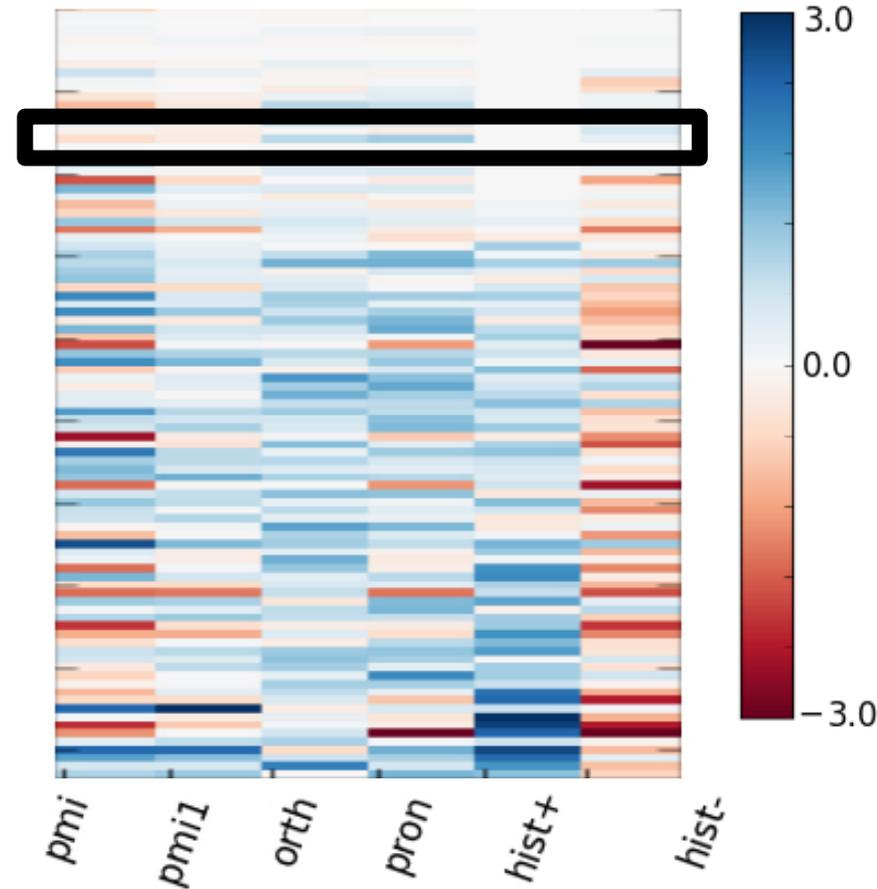
User Learning Styles

- Also trained a user-adapted model.
- 79 different users in our data pool.
- Learned 6 basic feature weights with 79 x 6 user adapted feature weights

Hal Daume III. Frustratingly easy domain adaptation. In Proceedings of ACL, pages 256–263, June 2007

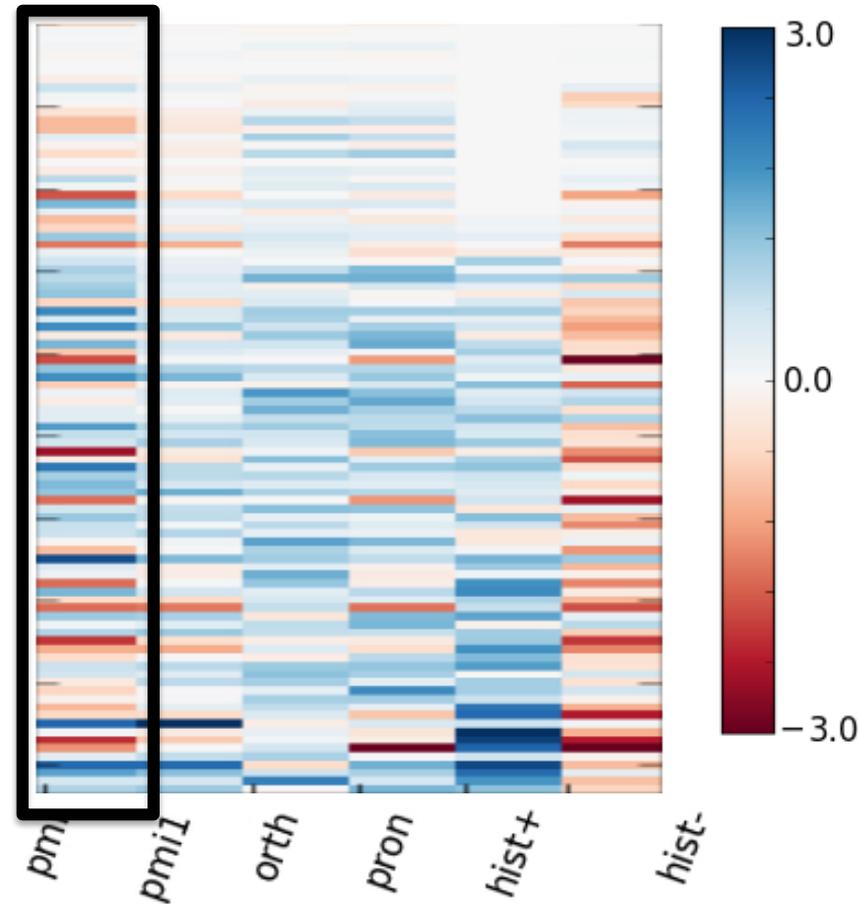
User Learning Styles

Each row represents the feature weights for a specific user.



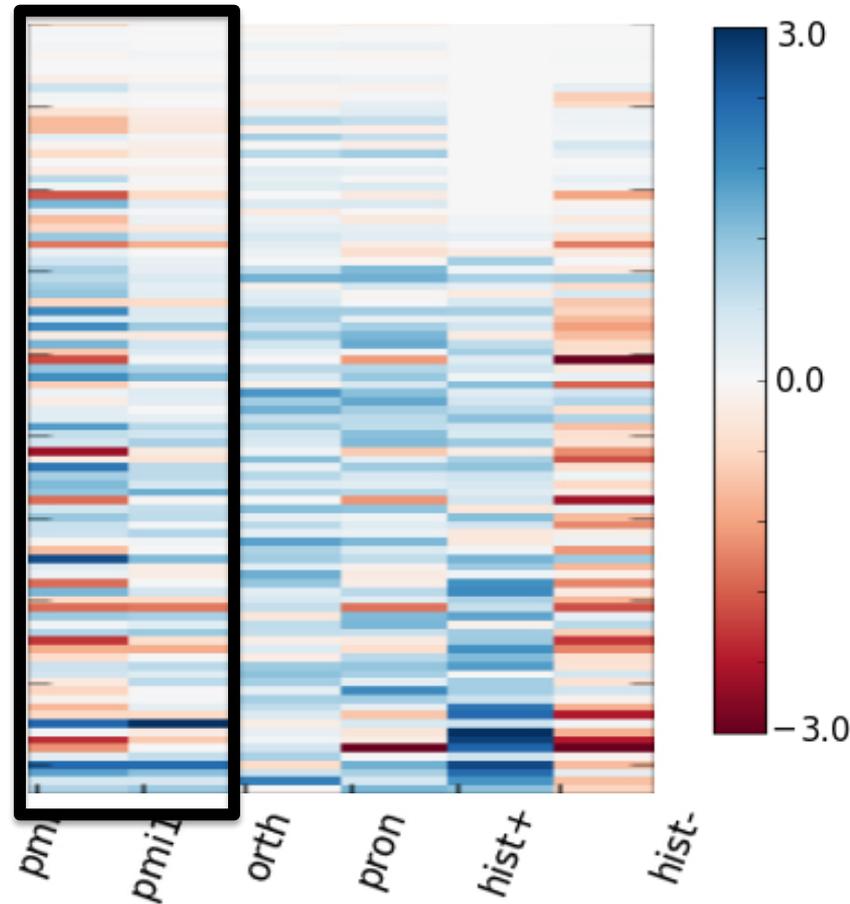
User Learning Styles

Columns are feature weights.



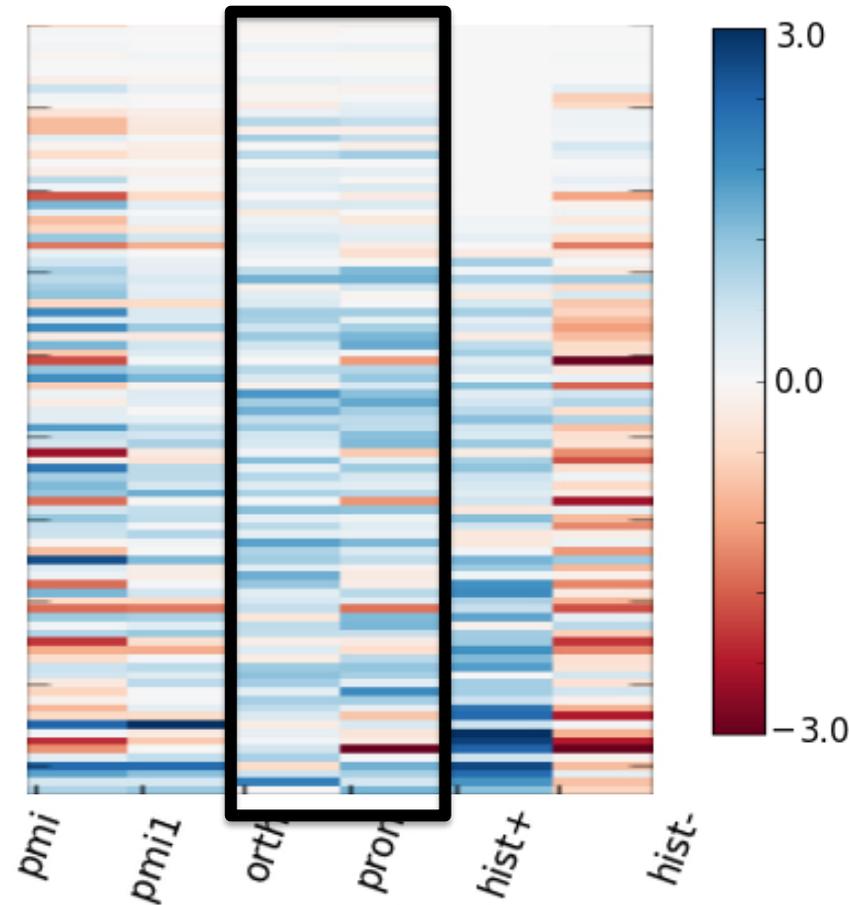
User Learning Styles

PMI @1 and PMI >1
Feature weights



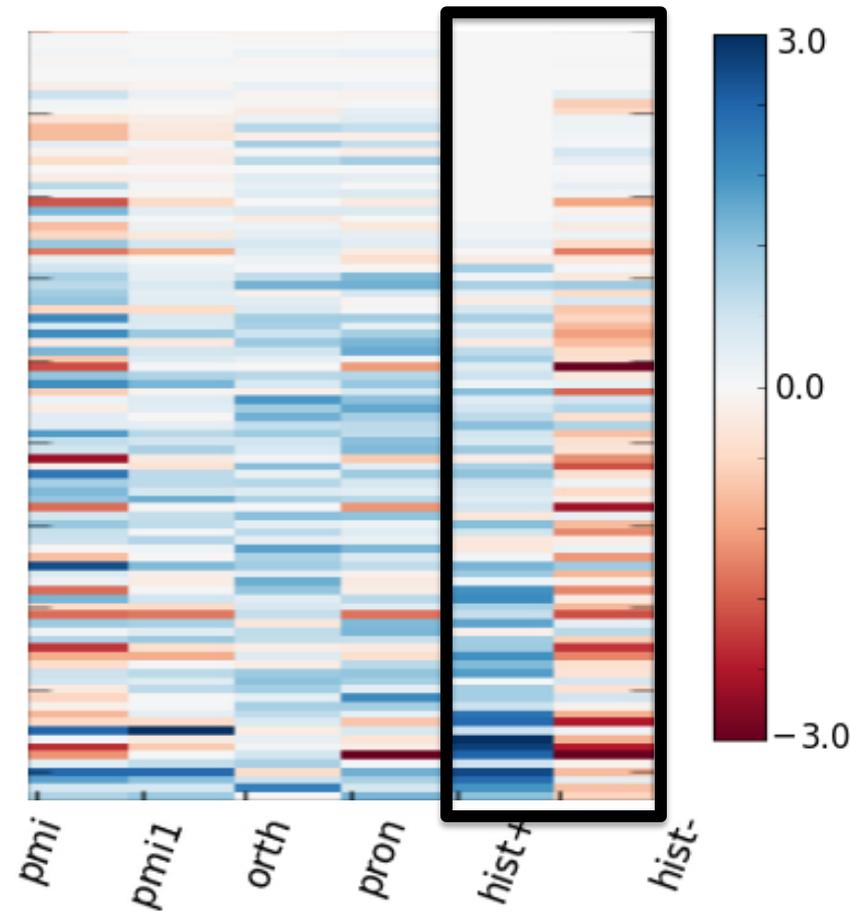
User Learning Styles

Similarity Feature weights



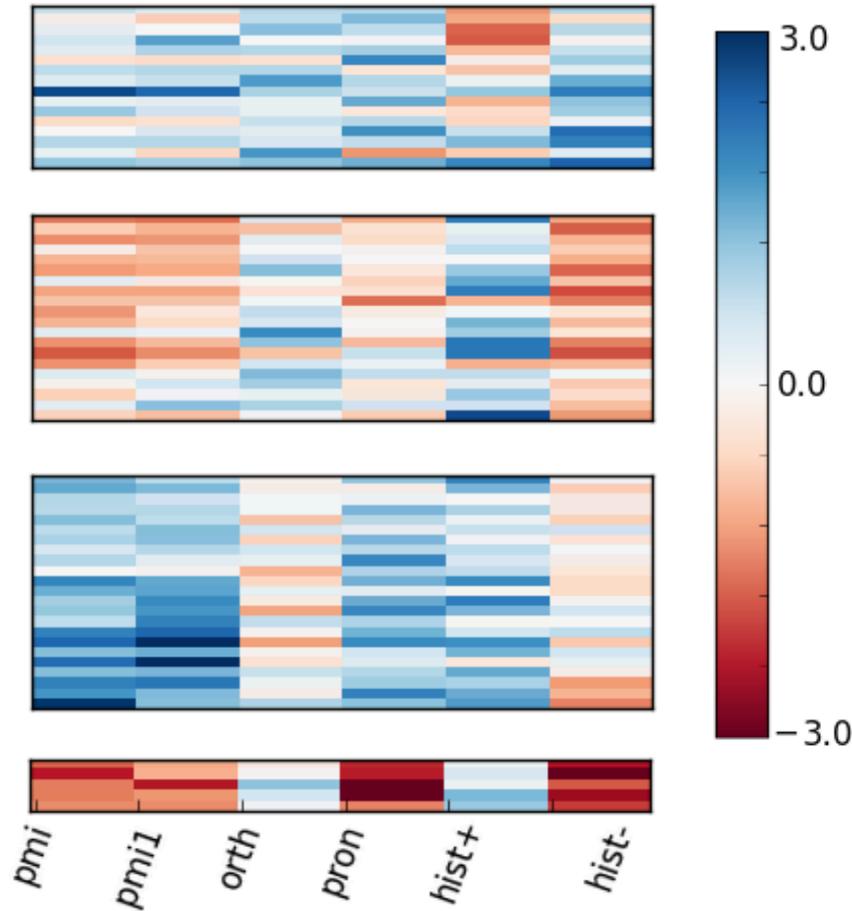
User Learning Styles

History Feature
weight



User Learning Styles

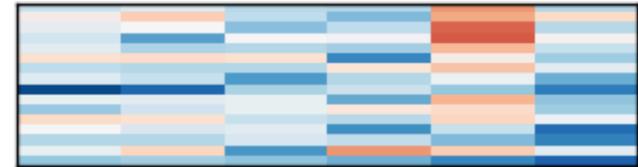
Clustered users
Into 4 groups.



User Learning Styles

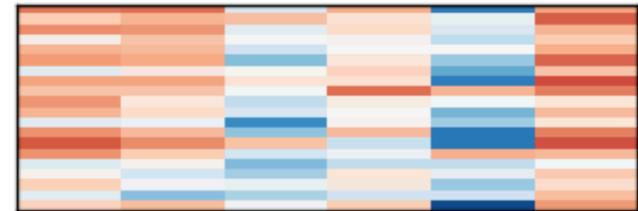
Using all the features to a similar degree

(A)



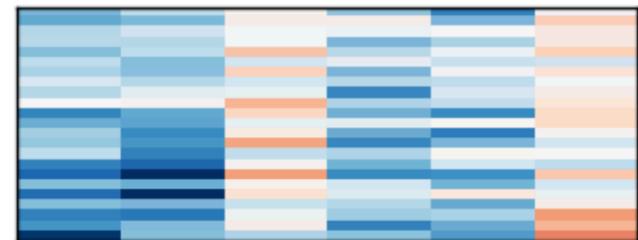
Using positive history and orthography but also some context

(B)



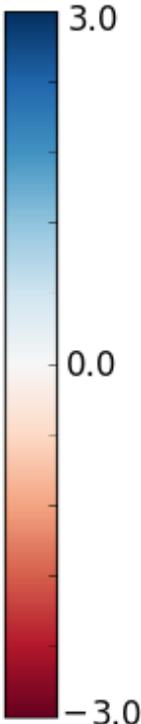
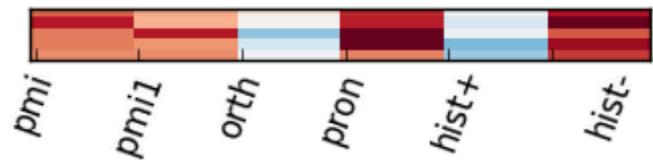
Using context, pronunciation and history!

(C)



Just seem to memorize a few words, only using orthography

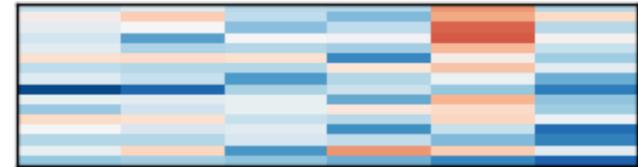
(D)



User Learning Styles

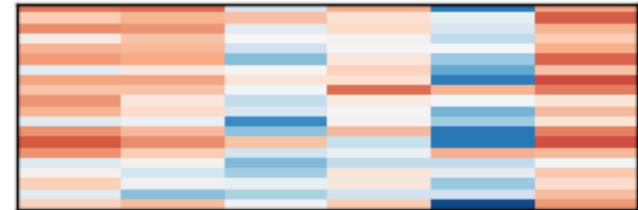
Using all the features to a similar degree

(A)



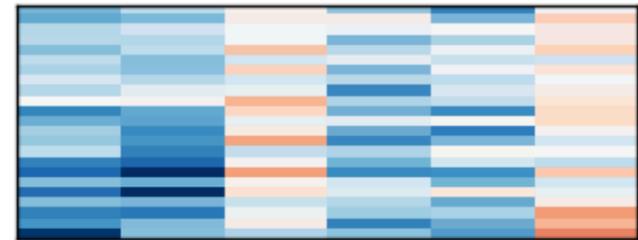
Using positive history and orthography but also some context

(B)



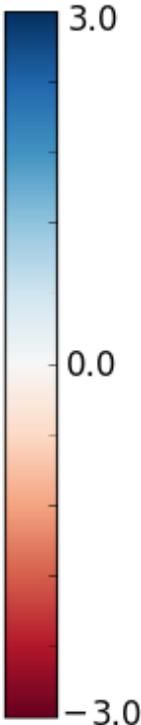
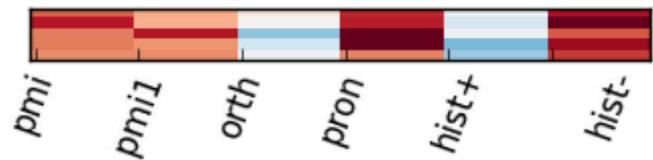
Using context, pronunciation and history!

(C)



Just seem to memorize a few words, only using orthography

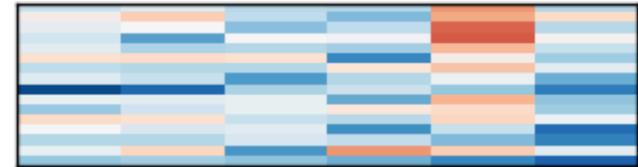
(D)



User Learning Styles

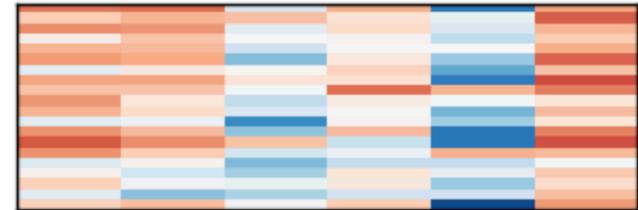
Using all the features to a similar degree

(A)



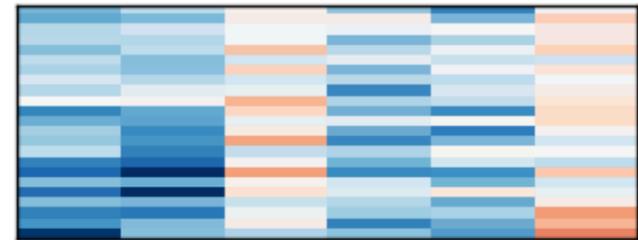
Using positive history and orthography but also some context

(B)



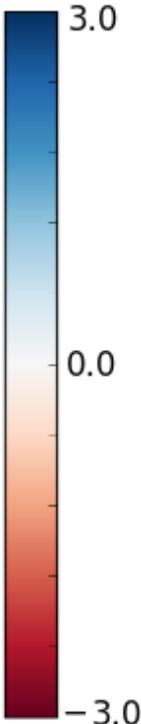
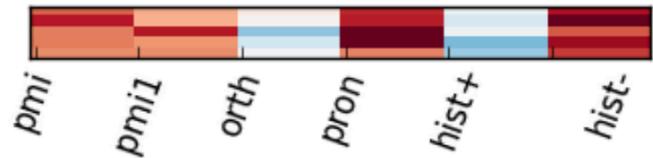
Using context, pronunciation and history!

(C)



Just seem to memorize a few words, only using orthography

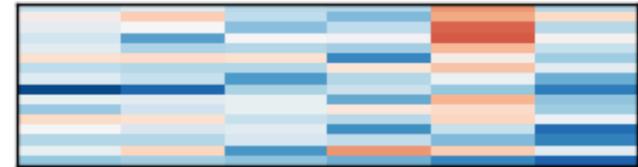
(D)



User Learning Styles

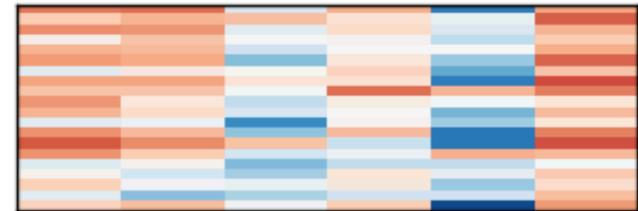
Using all the features to a similar degree

(A)



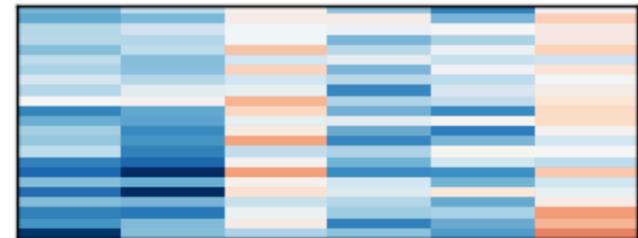
Using positive history and orthography but also some context

(B)



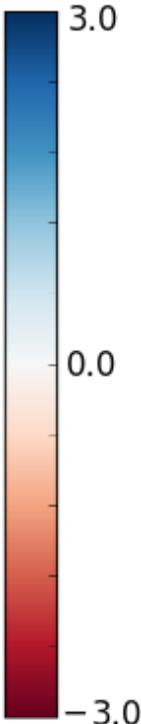
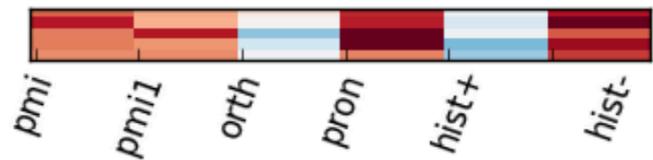
Using context, pronunciation and history!

(C)



Just seem to memorize a few words, only using orthography

(D)



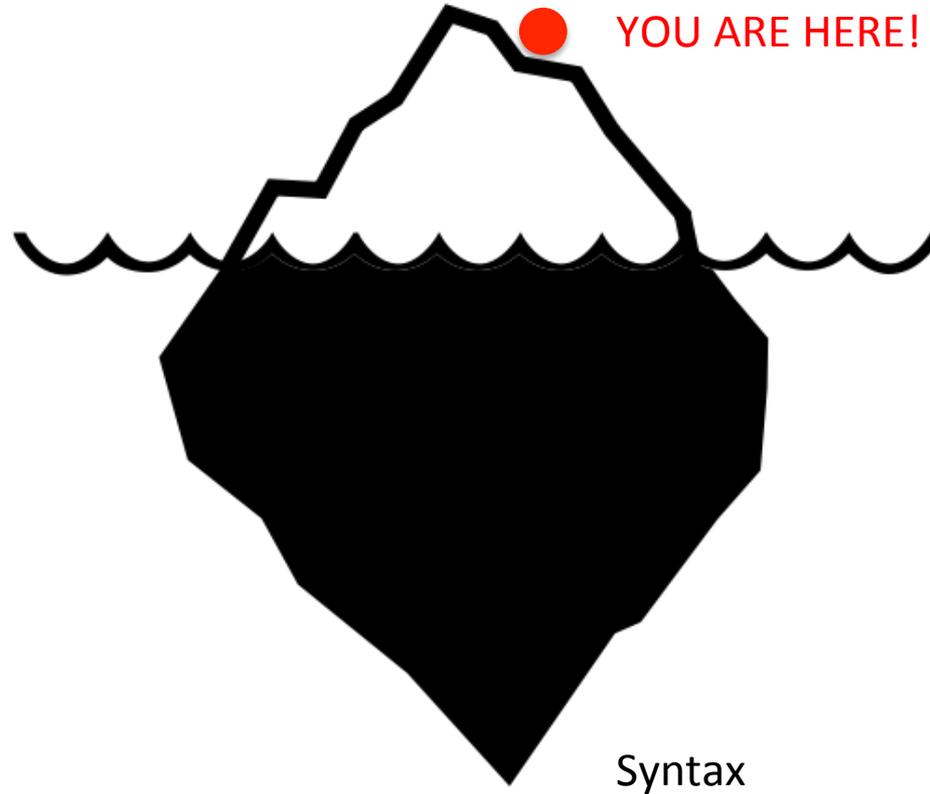
Recap

- General Problem:
 - Build automated systems to deliver personalized content at the appropriate macaronic level to a learner.
- Sub-problem:
 - Build a model to estimate a learner's comprehension of a macaronic sentence.

Open Challenges

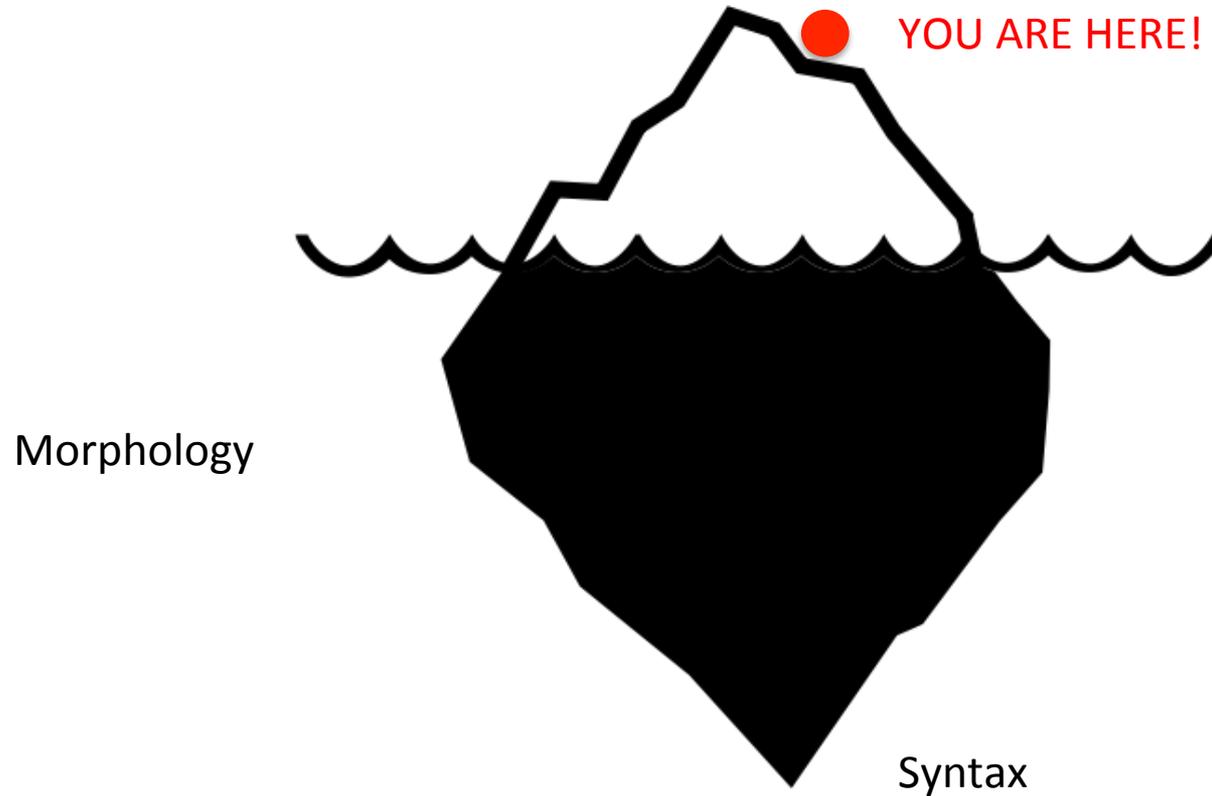


Open Challenges

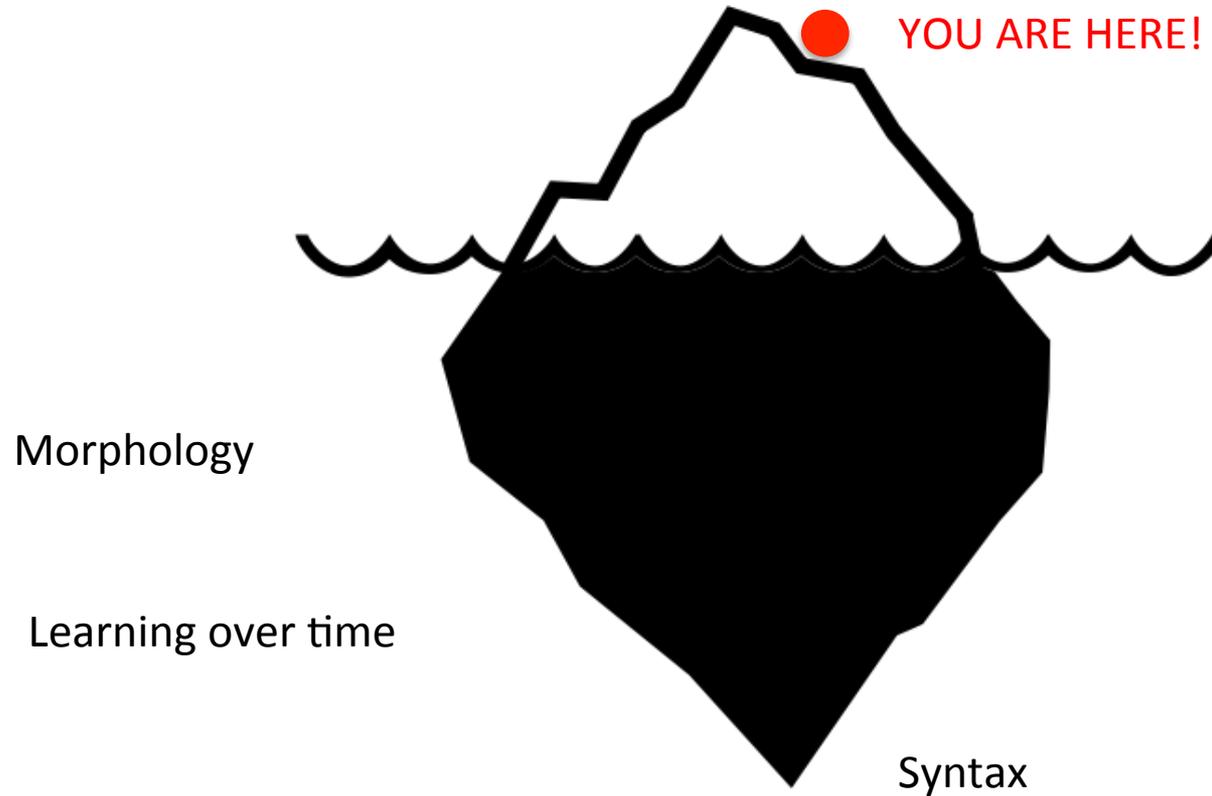


Syntax

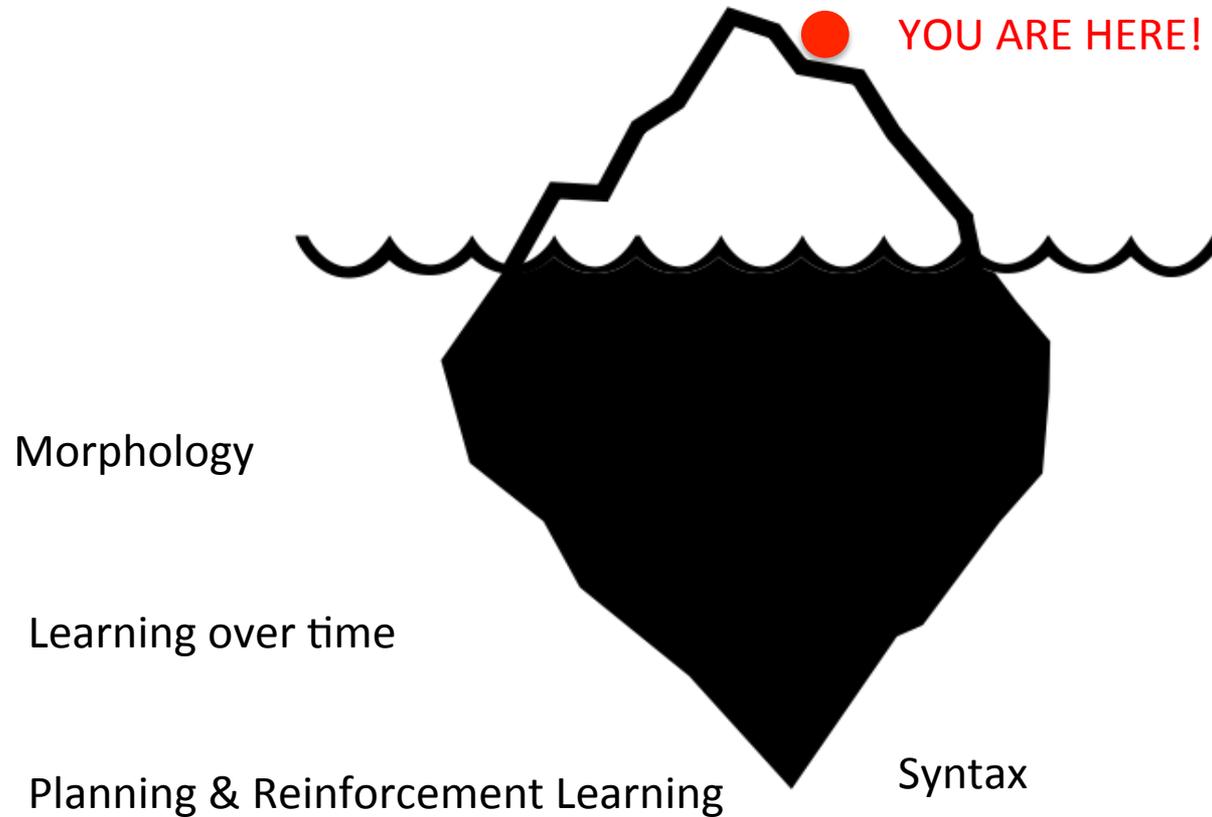
Open Challenges



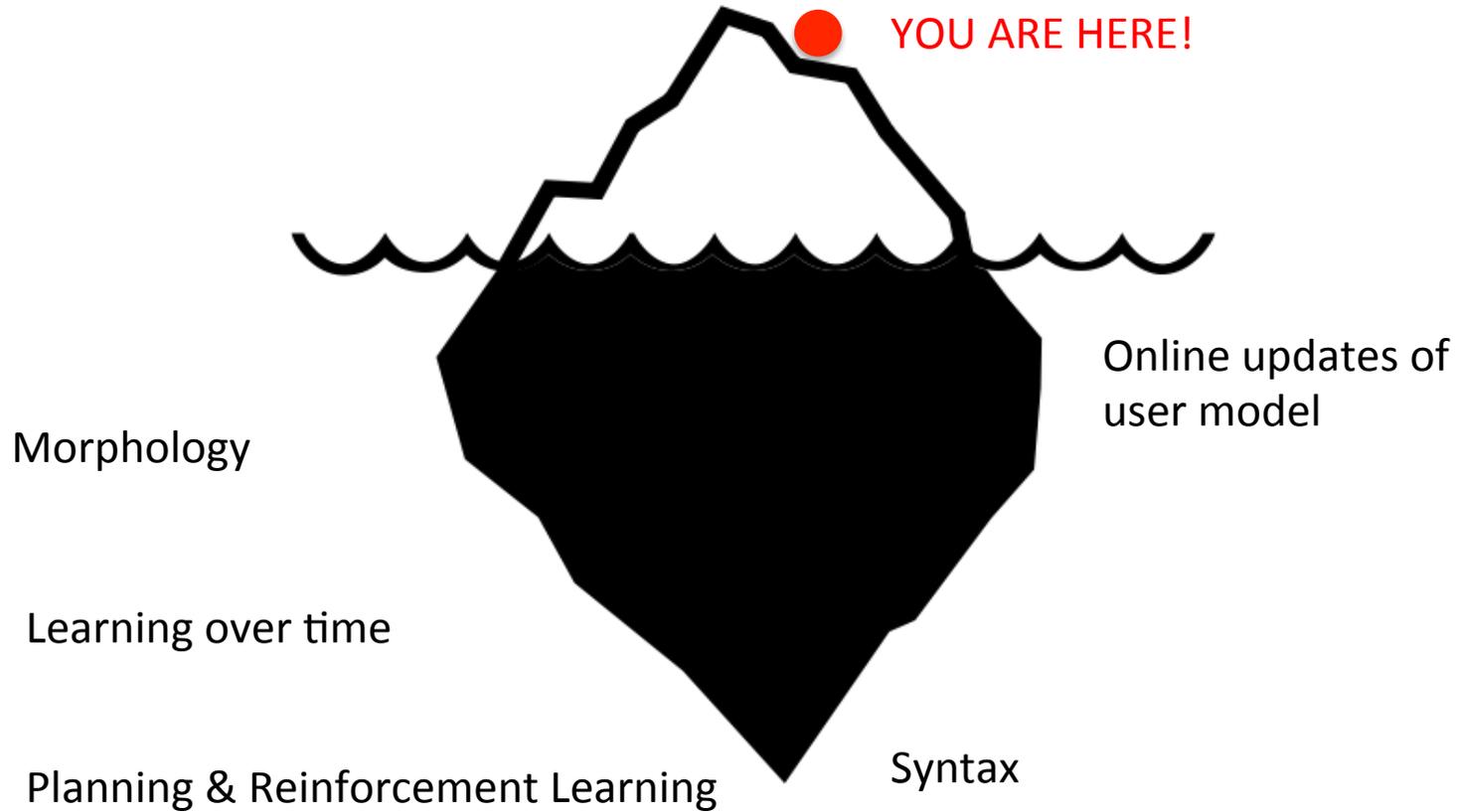
Open Challenges



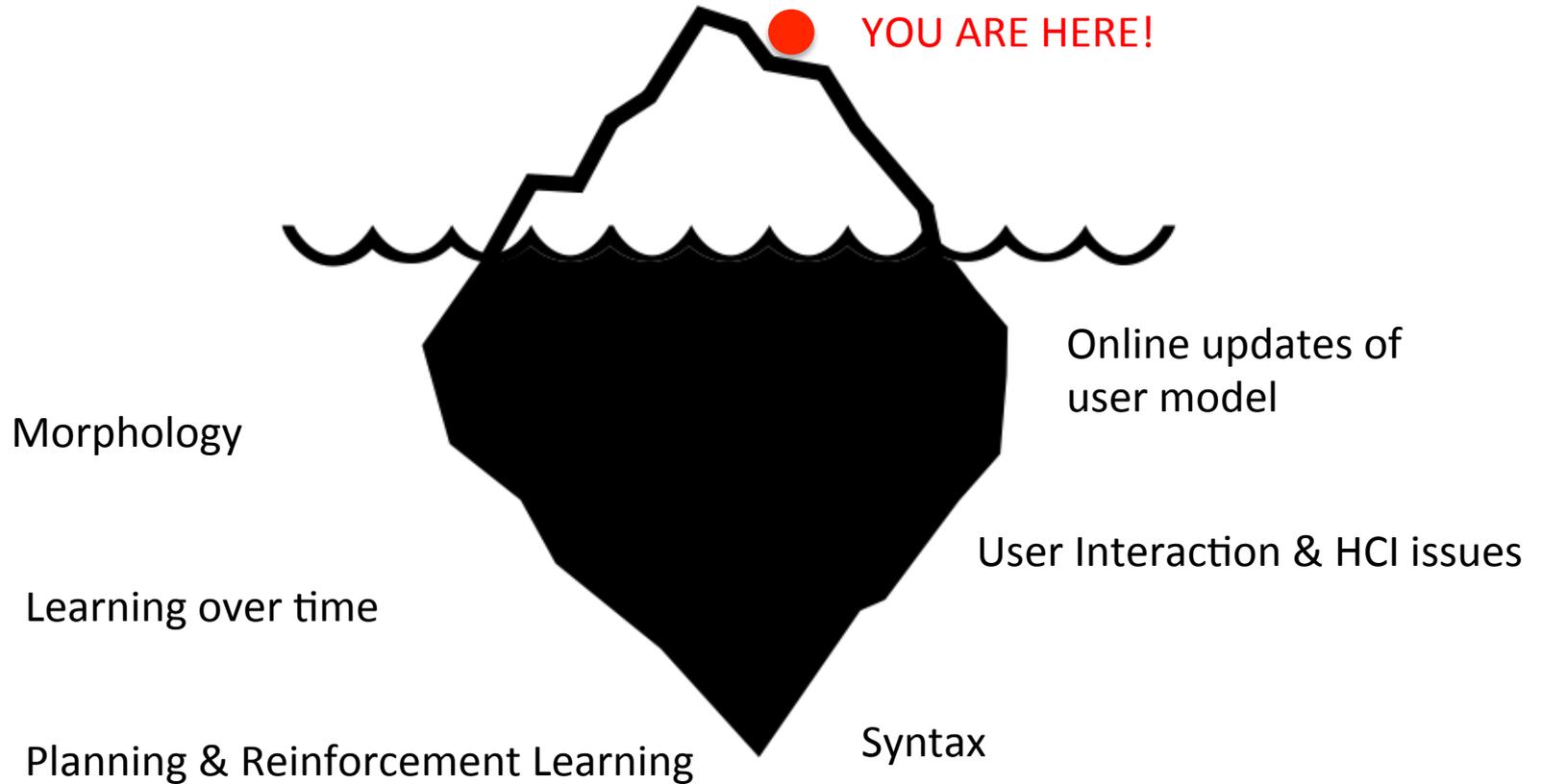
Open Challenges



Open Challenges



Open Challenges



Thank you!

Related Works at ACL

- **Demo session:**

“Creating Macaronic Interfaces for Language Learning”

5:30 – 7:00 pm Today!

Maritim Hotel

- **Companion paper:**

“Analyzing Learner Understanding of Novel L2 Vocabulary”

2 – 3:40 pm, Thursday

Room: 2.094