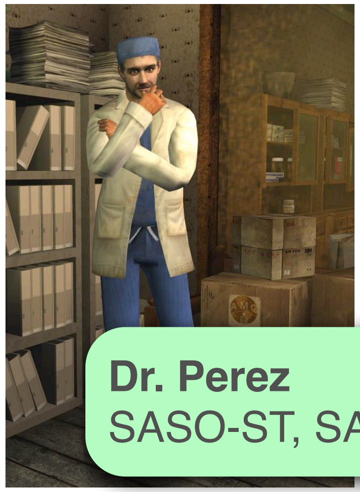
Natural Language Understanding: Statistical Approaches

Anton Leuski

USC

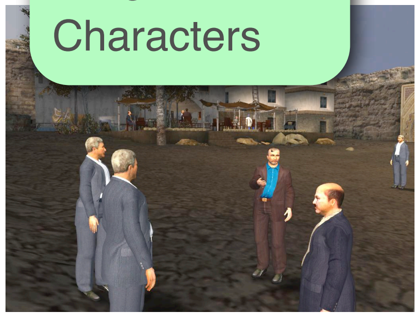
INSTITUTE FOR CREATIVE TECHNOLOGIES

ICT Virtual Humans

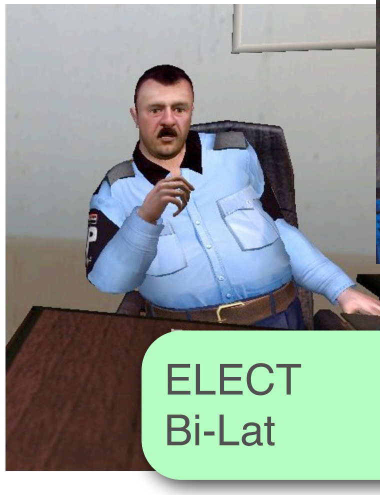


Dr. Perez
SASO-ST, SASO-EN

Elder-Al-Hassan
SASO-EN

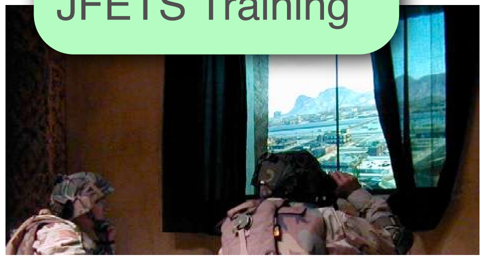


Minor
Characters



ELECT
Bi-Lat

C3IT
Cultural training



Radiobots
JFETS Training

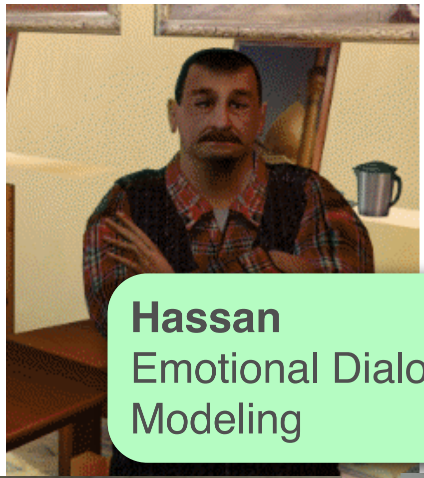
Virtual
Human
System

Research

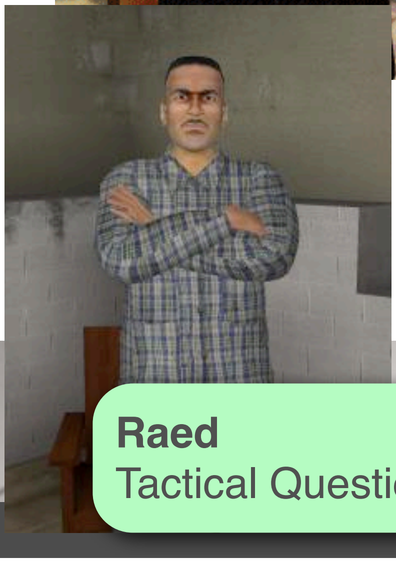
Training

Marketing

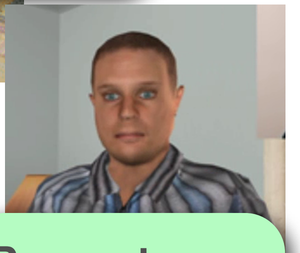
Clinical
Diagnosis



Hassan
Emotional Dialog
Modeling

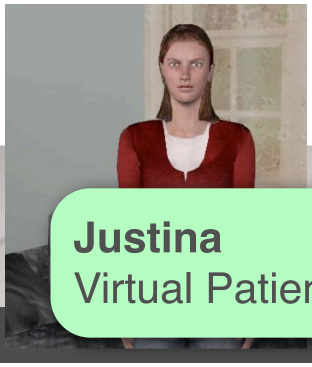


Raed
Tactical Questioning



Rapport
Agent Study

Justin
Virtual Patient

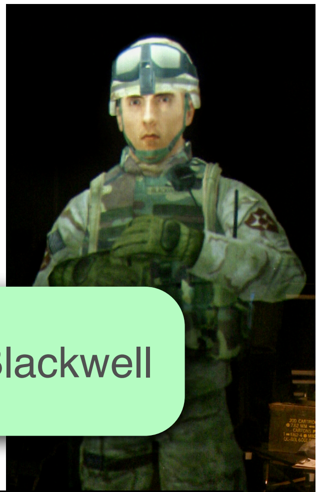


Justina
Virtual Patient

Gunslinger

Museum
Docent

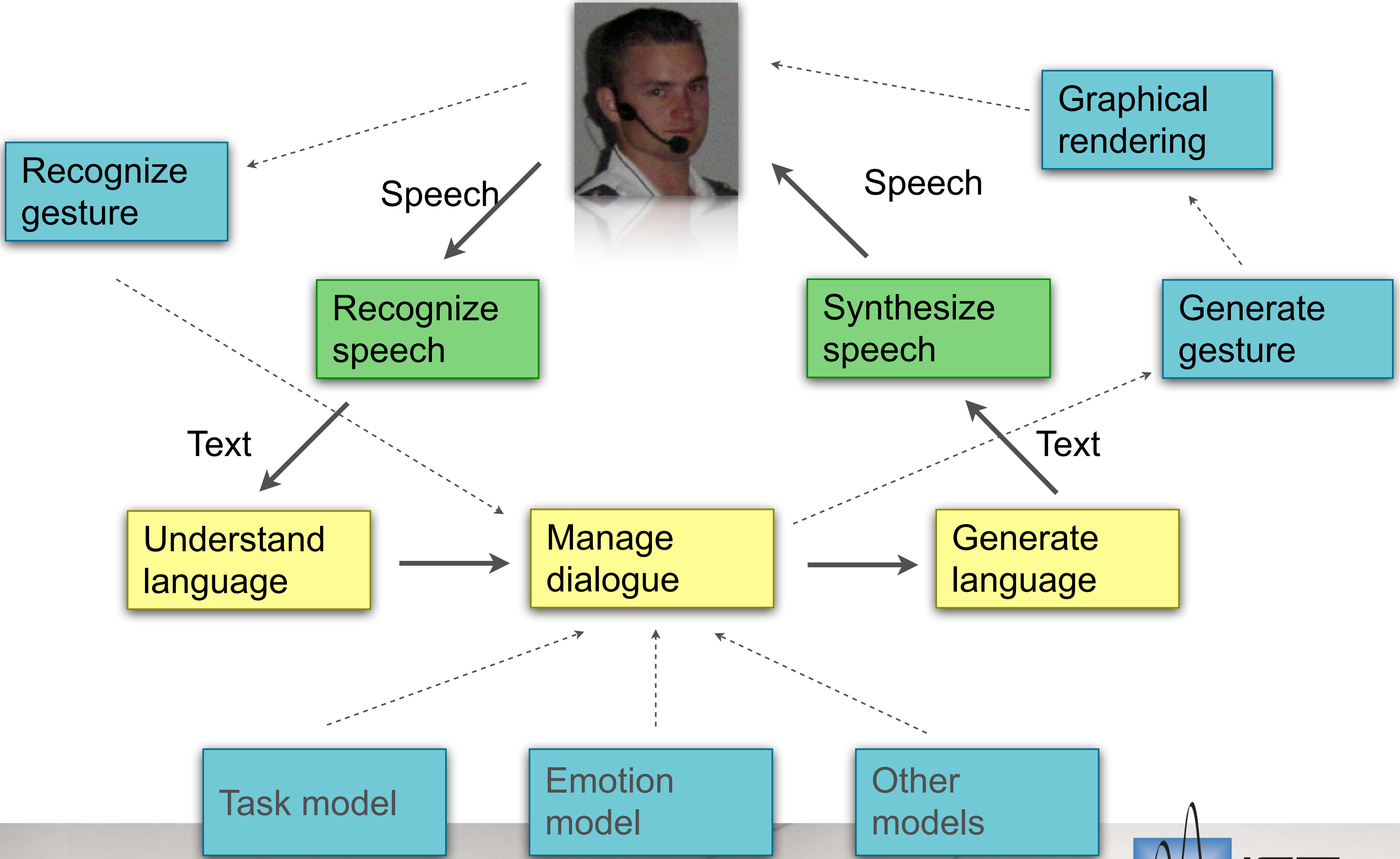
Be a Reporter



Sgt. Blackwell

Sgt. Star

Virtual Human Dialogue System



NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification →	Keep history →	Recorded answers
	**Transaction Dialogue**	Information extraction →	Follow Protocol →	Template-based
	**Negotiation**	Semantic parsing →	Rule-based reasoning →	Statistical & Grammar-based

complexity

NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification →	Keep history →	Recorded answers
	**Transaction Dialogue**	Information extraction →	Follow Protocol →	Template-based
	**Negotiation**	Semantic parsing →	Rule-based reasoning →	Statistical & Grammar-based

complexity

Language Understanding

- **Text classification**
  - "What is your name?" →
    "Sergeant John Blackwell. "Charlie" Company, sir."
- **Information Extraction**

"Alpha	one	six	this	is	Bravo	two	five	adjust	fire	over" →
FDC	FDC	FDC	O	O	FO	FO	FO	WO	WO	K

- **Semantic Parsing**
  - "We have to move the clinic" →

mood	declarative
speechact.type	statement
modal.deontic	must
task	move-clinic
type	event
event	move
theme	clinic
source	here
destination	there

Language Understanding

- **Problem: Speech input is often unpredictable**
  - language flexibility
  - spoken language disfluencies
  - speech recognition errors
- **Solution: Automatically train machines from input-output pairs**

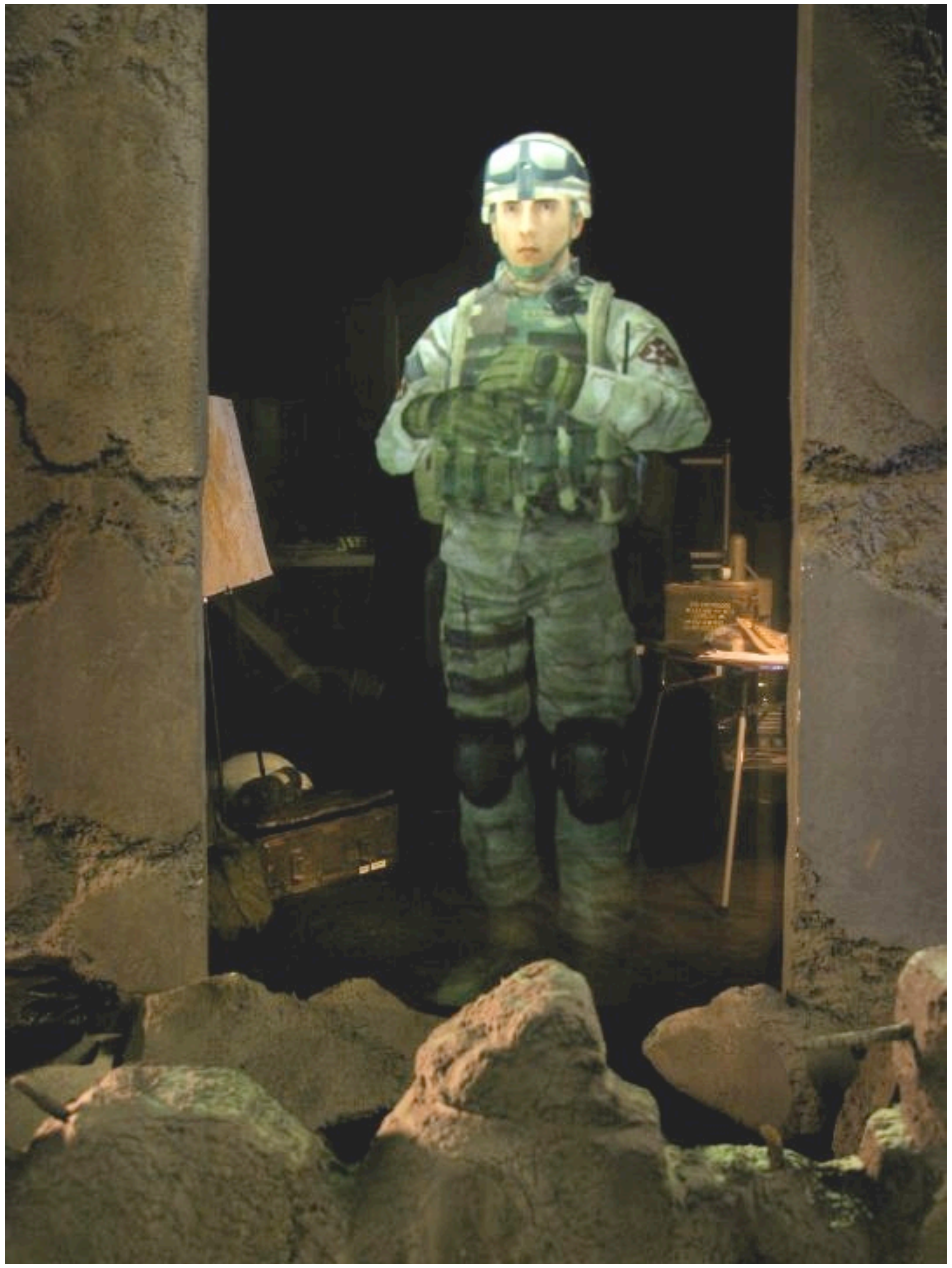
NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification	Keep history	Recorded answers
	**Transaction Dialogue**	Information extraction	Follow Protocol	Template-based
	**Negotiation**	Semantic parsing	Rule-based reasoning	Statistical & Grammar-based

complexity

Text Classification



- **The system has ...**
  - fixed set of answers
  - a set of questions linked to the answers - training
  - qa mapping is many-to-many
  - a test question
- **It must select the correct answer**
- **3 approaches**

Approach 1: Traditional Classification

- **Answer = class label**
- **Question = instance**
  - vector of words extracted from the text (tf•idf)
- **Training questions = training instances**
  - instances assigned to a class
- **Classification algorithm will assign class to a test question**
  - SVM (e.g., SVM$^{\textit{light}}$) state-of-the-art technique

Approach 1: Traditional Classification

- **Answer = class label**
- **Question = instance**
  - vector of words extracted from the text (tf•idf)
- **Training questions = training instances**
  - instances assigned to a class
- **Classification algorithm will assign class to a test question**
  - SVM (e.g., SVM*light*) state-of-the-art technique

Limitations:

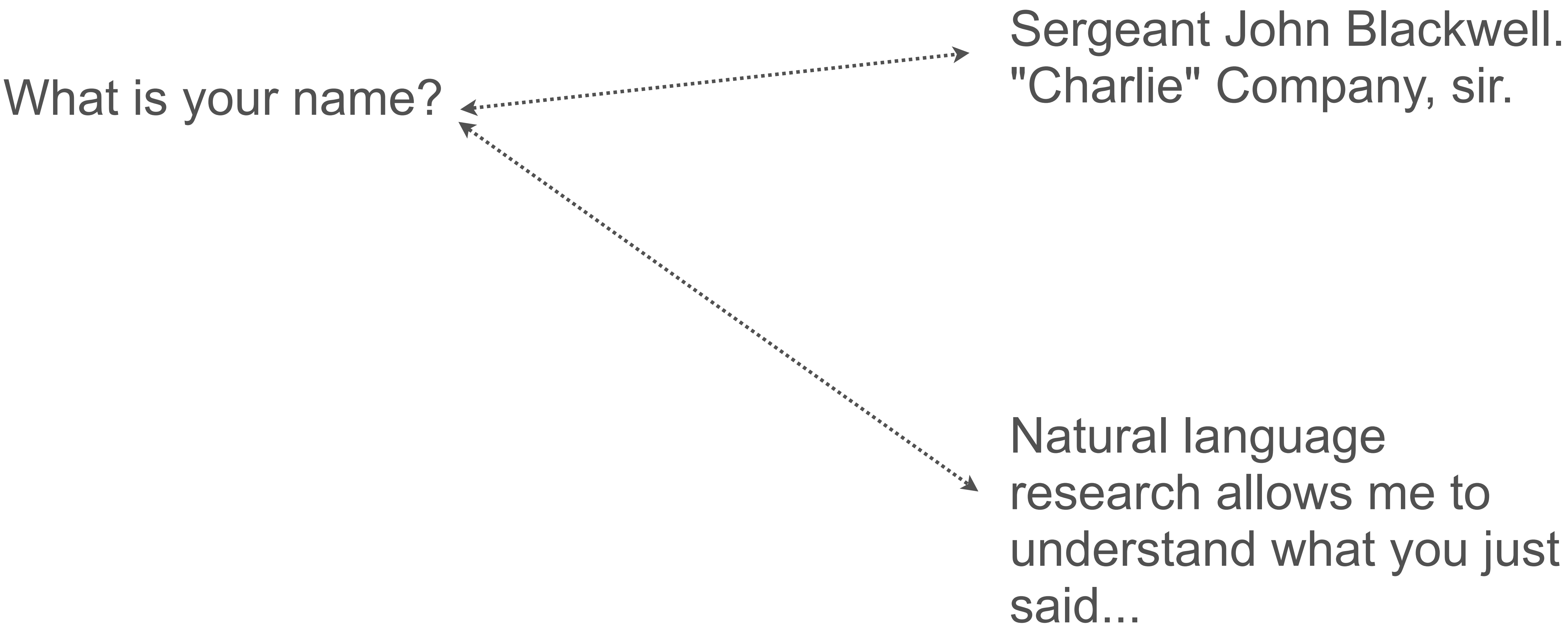
- ignores answer text
- ad-hoc features
- binary classification - cannot easily handle many-to-many relations

Can we do better?

Information Retrieval

- **Answer = document**
  - text is important
- **Question = query**
- **Compare question text against answer texts and select the most similar answer**
  - Language Modeling (others exist)
- **Tune text representation using training questions**

Language Models in IR

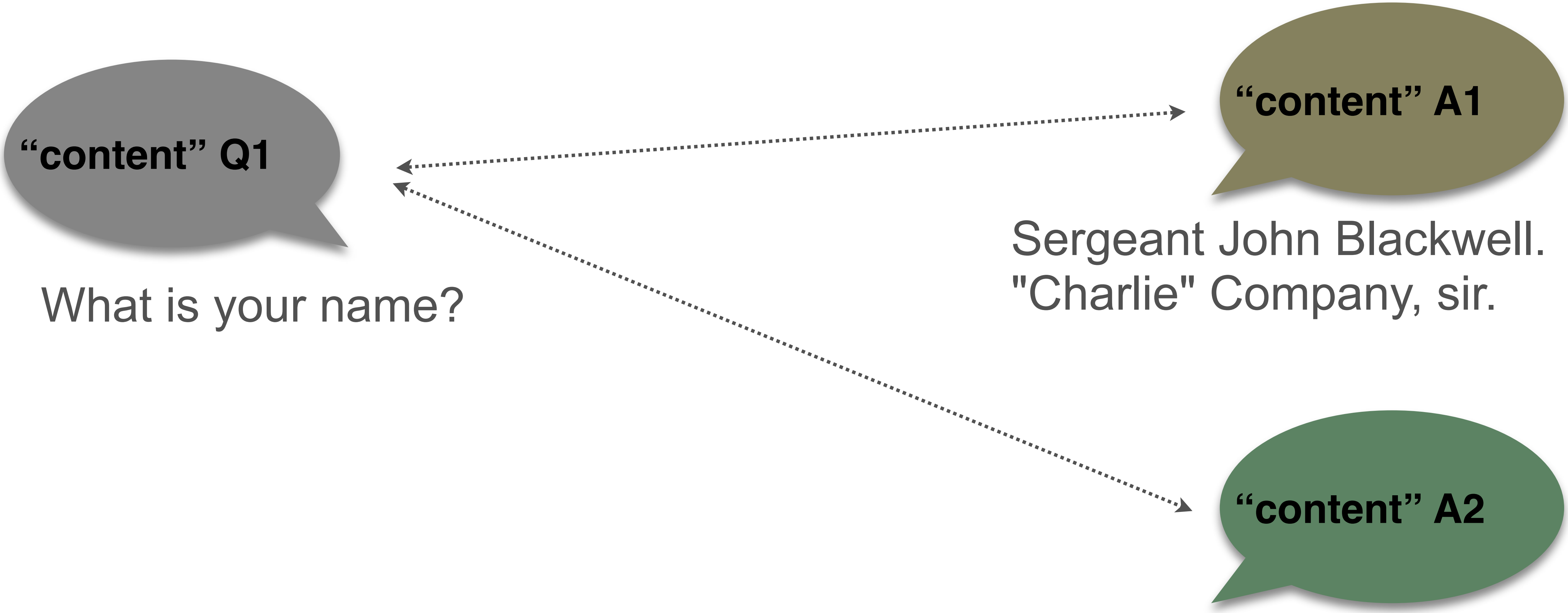


What is your name?

Sergeant John Blackwell. "Charlie" Company, sir.

Natural language research allows me to understand what you just said...

Language Models in IR

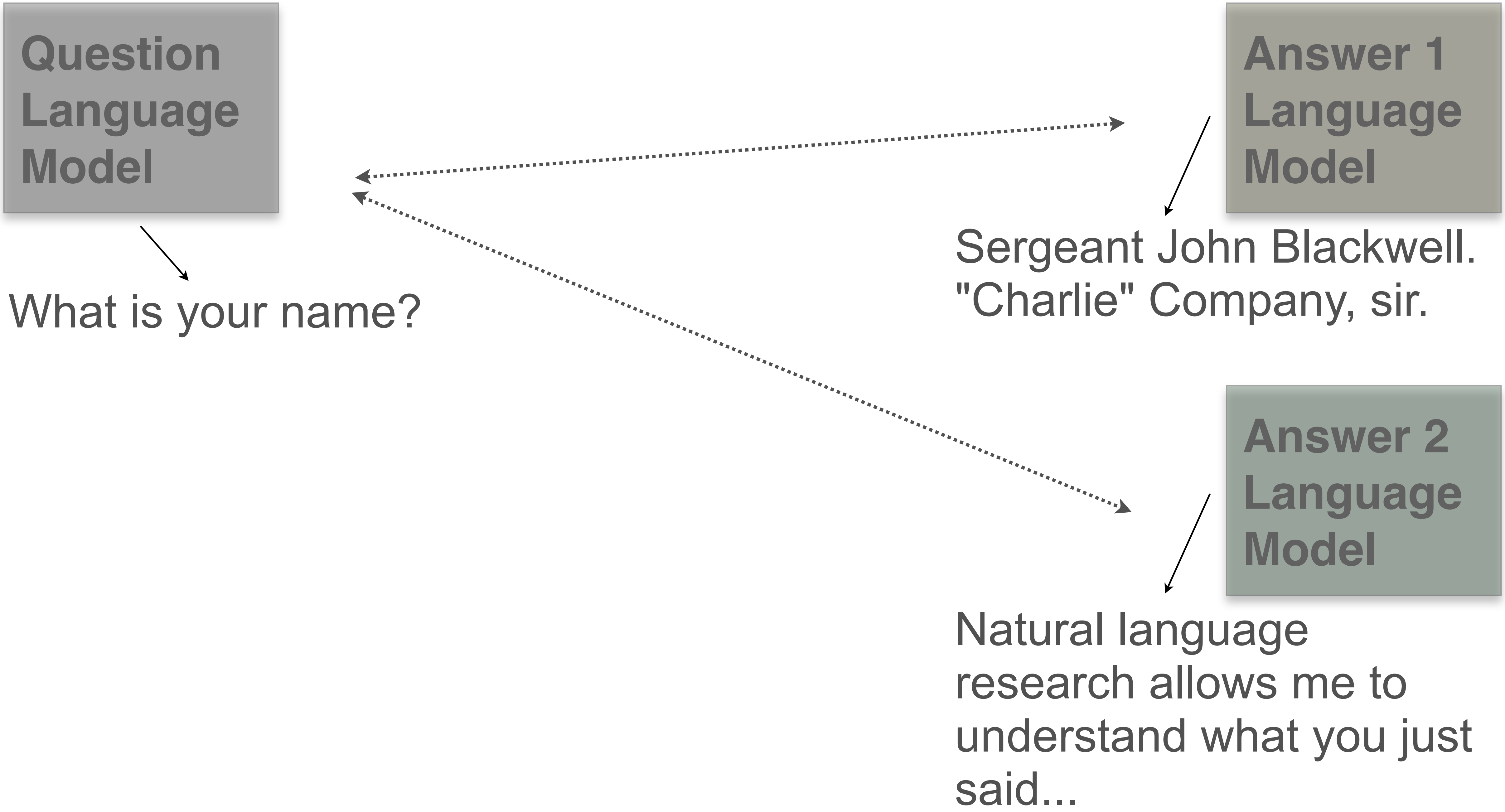


What is your name?

Sergeant John Blackwell. "Charlie" Company, sir.

Natural language research allows me to understand what you just said...

Language Models in IR



Question Language Model

What is your name?

Answer 1 Language Model

Sergeant John Blackwell. "Charlie" Company, sir.

Answer 2 Language Model

Natural language research allows me to understand what you just said...

Language Model

Name?

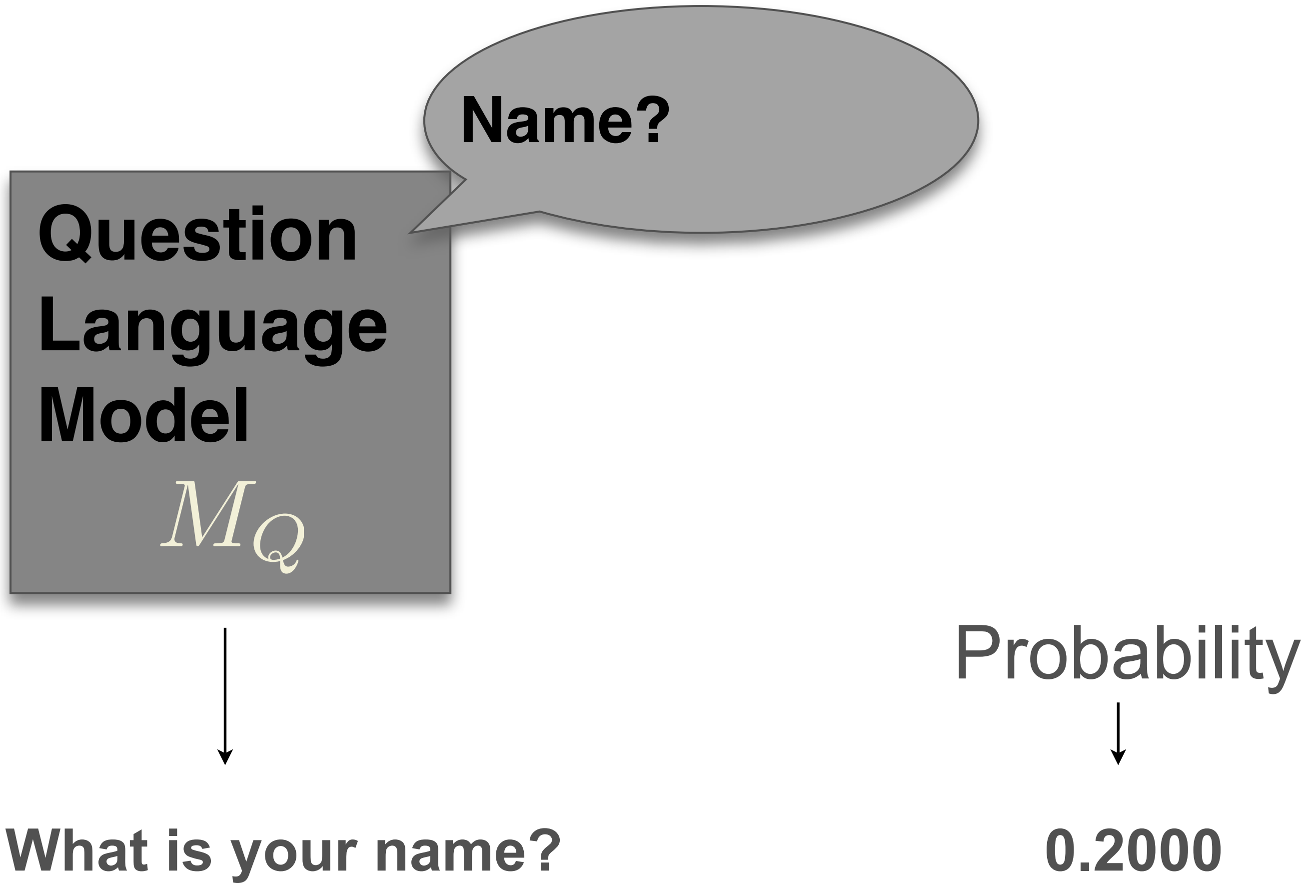
Question Language Model

M_Q

↓

What is your name?

Language Model



Name?

Question Language Model

M_Q

Probability

What is your name?

0.2000

Language Model

Name?

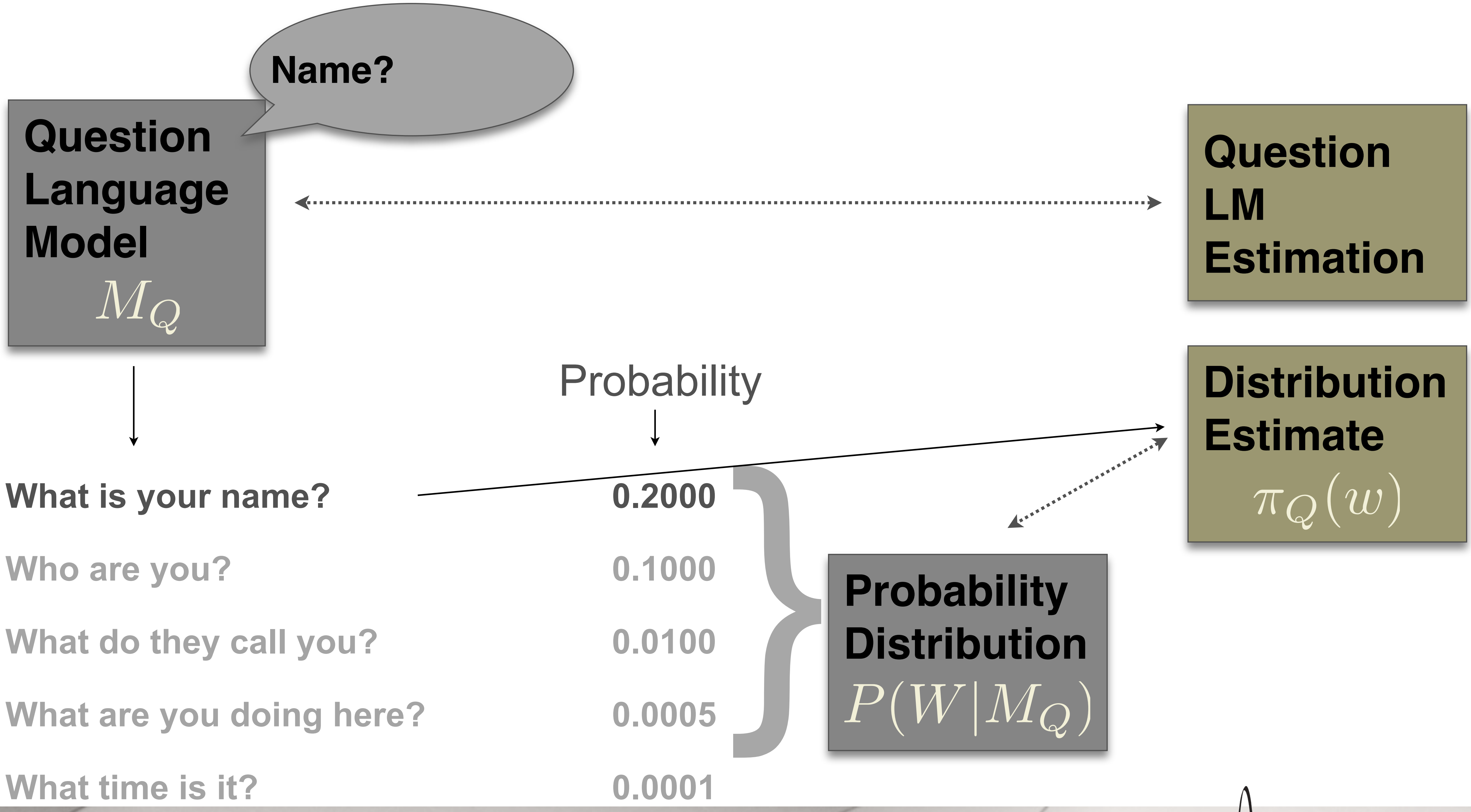
Question Language Model
M_Q

Probability

What is your name?	**0.2000**
Who are you?	0.1000
What do they call you?	0.0100
What are you doing here?	0.0005
What time is it?	0.0001

Probability Distribution
$P(W|M_Q)$

Language Model



Name?

Question Language Model M_Q

Question LM Estimation

Distribution Estimate $\pi_Q(w)$

Probability Distribution $P(W|M_Q)$

Probability

What is your name?	**0.2000**
Who are you?	0.1000
What do they call you?	0.0100
What are you doing here?	0.0005
What time is it?	0.0001

...

Language Models in IR

Question Language Model

What is your name?

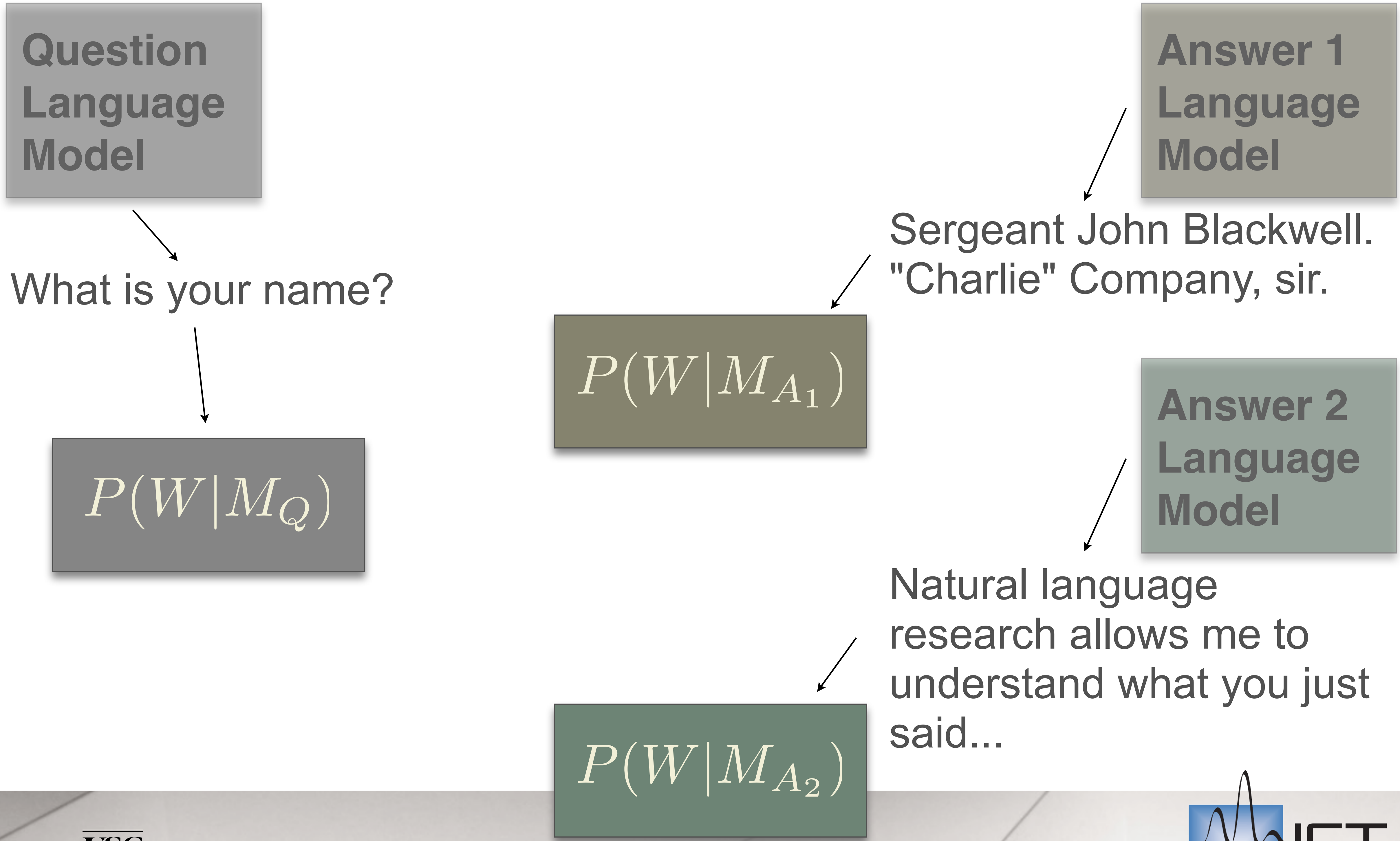
Answer 1 Language Model

Sergeant John Blackwell. "Charlie" Company, sir.

Answer 2 Language Model

Natural language research allows me to understand what you just said...

Language Models in IR



Question Language Model

What is your name?

$P(W|M_Q)$

Answer 1 Language Model

Sergeant John Blackwell. "Charlie" Company, sir.

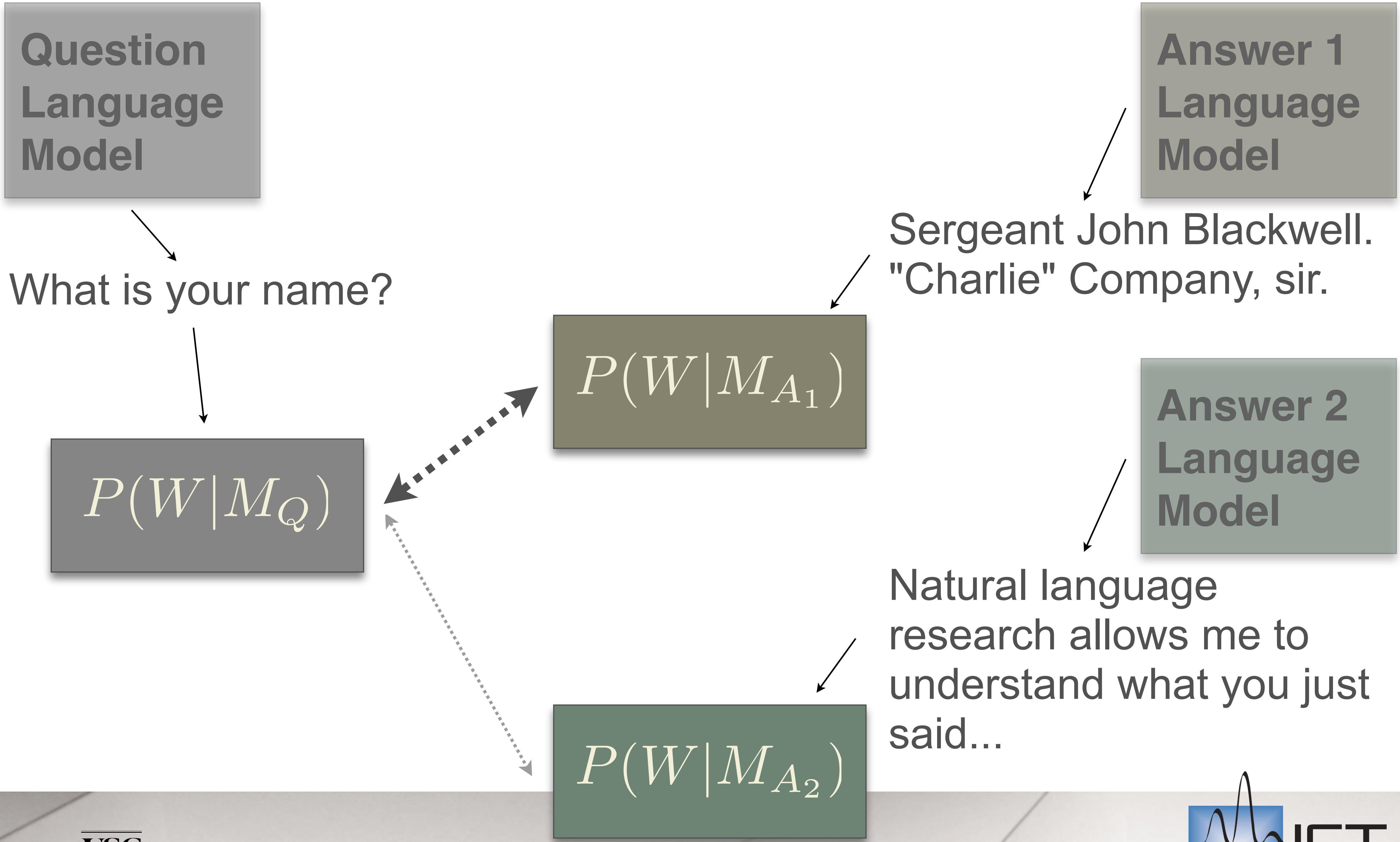
$P(W|M_{A_1})$

Answer 2 Language Model

Natural language research allows me to understand what you just said...

$P(W|M_{A_2})$

Language Models in IR



Question Language Model

What is your name?

$P(W|M_Q)$

Answer 1 Language Model

Sergeant John Blackwell. "Charlie" Company, sir.

$P(W|M_{A_1})$

Answer 2 Language Model

Natural language research allows me to understand what you just said...

$P(W|M_{A_2})$

Information Retrieval

- **Answer (document in IR)**
  - – estimate M_A: $P(w|M_A)$
- **Question (query)**
  - – estimate M_Q: $P(w|M_Q)$
- **Compare questions against answers: similarity score**
  - – cross-entropy: number of bits to “encode” M_Q with M_A

$$H(M_Q||M_A) = -\sum_w P(w|M_Q) \log P(w|M_A)$$

- **Rank all answers by the similarity score**
- **Cut the ranking at some threshold**
- **Return the set (or you can return the top ranked answer)**

Estimation

- **Unigram language model**

$$P(W) = P(w_1 ... w_n) = \prod_{i=1}^{n} P(w_i)$$

- **Jelinek-Mercer estimation**
  - Interpolated Maximum-likelihood

$$P(w|M_A) = \pi_A(w)$$

$$\pi_s(w) = \lambda_\pi \cdot \frac{\#(w,s)}{|s|} + (1-\lambda_\pi) \cdot \frac{\sum_s \#(w,s)}{\sum_s |s|}$$

- **Other approaches exist**

Text as Vector

“What happened here?”

what	**happened**	**here**

Term (w)	#(w,s)
what	1
happened	1
here	1
...	...

- **“Bag of words”**
- **Stopping - remove frequent words, e.g., “a”, “the”, ...**
- **Stemming - find word root**
- **N-grams to capture order:**

what happened	**happened here**

what happened here

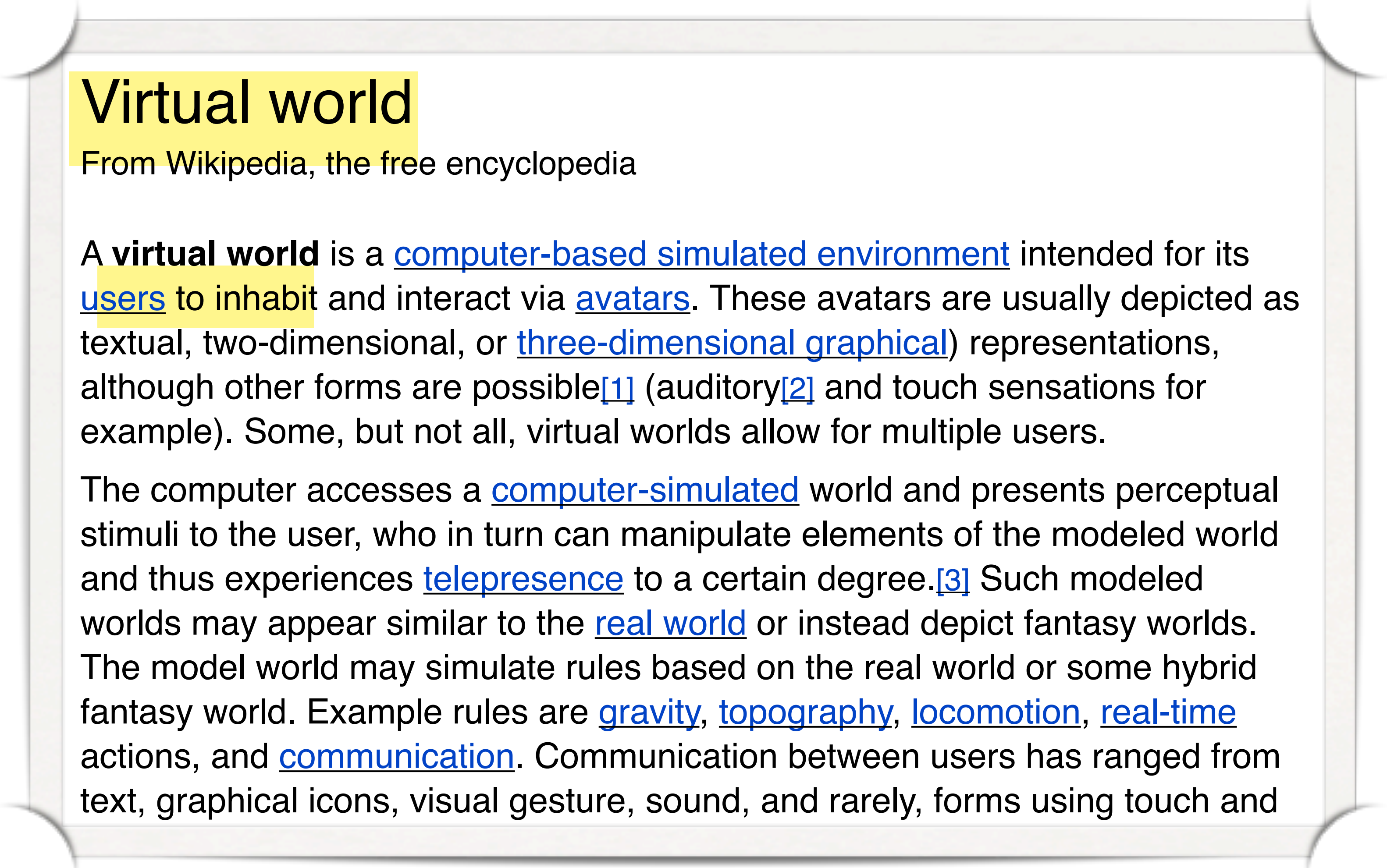
LM in IR: Assumption

- **IR assumes that language of queries is the same as language of documents**
  - a query is like a document - will have common words

“Virtual World”

↓

→



Virtual world

From Wikipedia, the free encyclopedia

A **virtual world** is a computer-based simulated environment intended for its users to inhabit and interact via avatars. These avatars are usually depicted as textual, two-dimensional, or three-dimensional graphical) representations, although other forms are possible[1] (auditory[2] and touch sensations for example). Some, but not all, virtual worlds allow for multiple users.

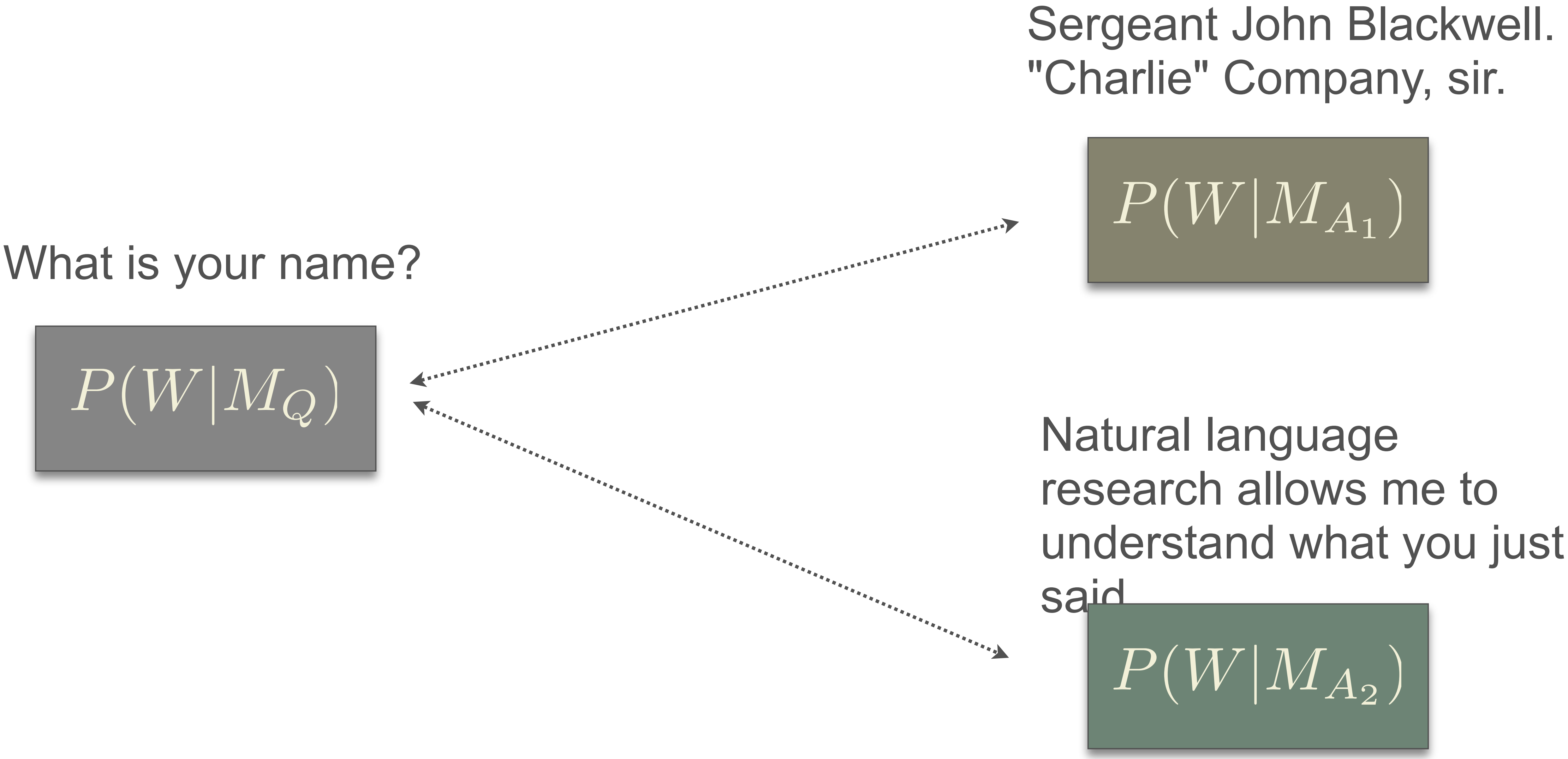
The computer accesses a computer-simulated world and presents perceptual stimuli to the user, who in turn can manipulate elements of the modeled world and thus experiences telepresence to a certain degree.[3] Such modeled worlds may appear similar to the real world or instead depict fantasy worlds. The model world may simulate rules based on the real world or some hybrid fantasy world. Example rules are gravity, topography, locomotion, real-time actions, and communication. Communication between users has ranged from text, graphical icons, visual gesture, sound, and rarely, forms using touch and

LM in Question Answering

- **This is incorrect for question answering**
  - questions and answers may have no common words:

    “What is your name?” →
    “Sergeant John Blackwell. "Charlie" Company, sir.”
  - questions have specific grammar
  - questions are not answers!
- **Questions and answers are two “languages”**

Approach 2: Single Language Retrieval



What is your name?

$P(W|M_Q)$

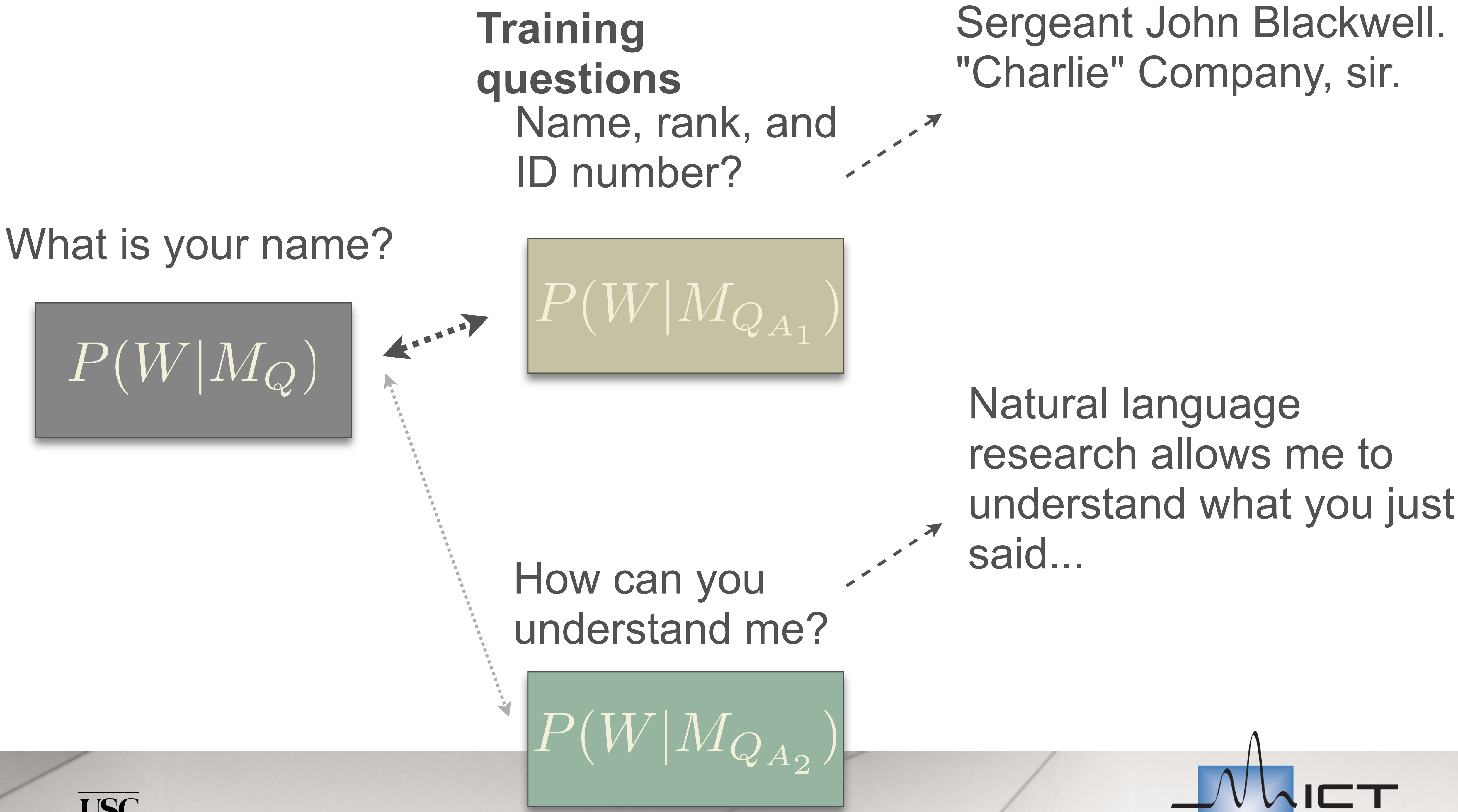
Sergeant John Blackwell. "Charlie" Company, sir.

$P(W|M_{A_1})$

Natural language research allows me to understand what you just said.

$P(W|M_{A_2})$

Approach 2: Single Language Retrieval



Training questions

What is your name?

$P(W|M_Q)$

Name, rank, and ID number?

$P(W|M_{Q_{A_1}})$

Sergeant John Blackwell. "Charlie" Company, sir.

How can you understand me?

$P(W|M_{Q_{A_2}})$

Natural language research allows me to understand what you just said...

Approach 2: Single Language Retrieval

- **Retrieve a training question, and select the matching answer**
  - document = a training question
  - document = text of all questions linked to a single answer
- **Limitation: ignores the answer text**

Approach 3: Cross-Language Retrieval

Tell me about your technology?

$P(W|M_Q)$

Sergeant John Blackwell. "Charlie" Company, sir.

$P(W|M_{A_1})$

Natural language research allows me to understand what you just sa

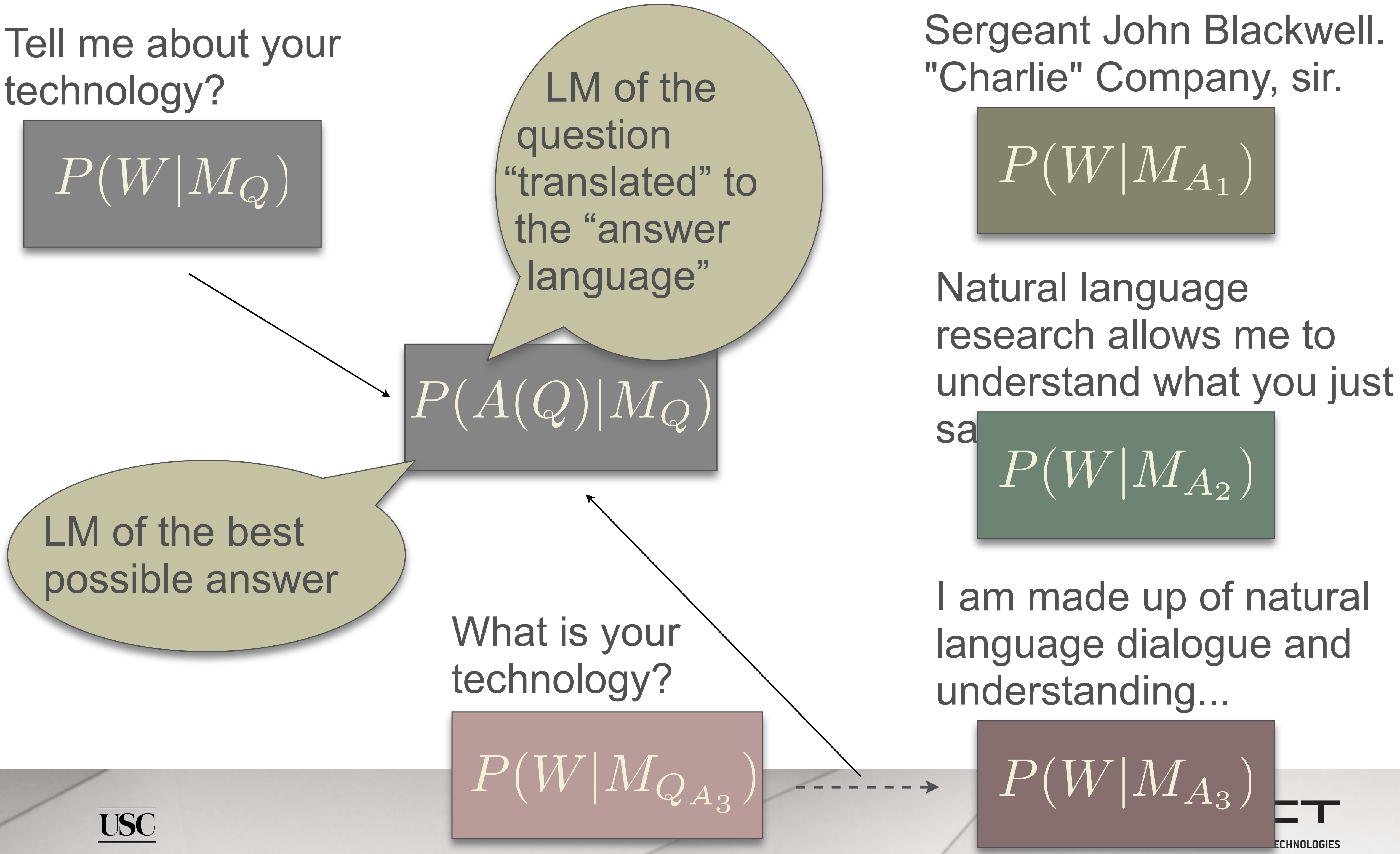
$P(W|M_{A_2})$

What is your technology?

$P(W|M_{Q_{A_3}}) \dashrightarrow P(W|M_{A_3})$

I am made up of natural language dialogue and understanding...

Approach 3: Cross-Language Retrieval



Tell me about your technology?

$P(W|M_Q)$

LM of the question “translated” to the “answer language”

$P(A(Q)|M_Q)$

LM of the best possible answer

What is your technology?

$P(W|M_{Q_{A_3}})$

Sergeant John Blackwell. "Charlie" Company, sir.

$P(W|M_{A_1})$

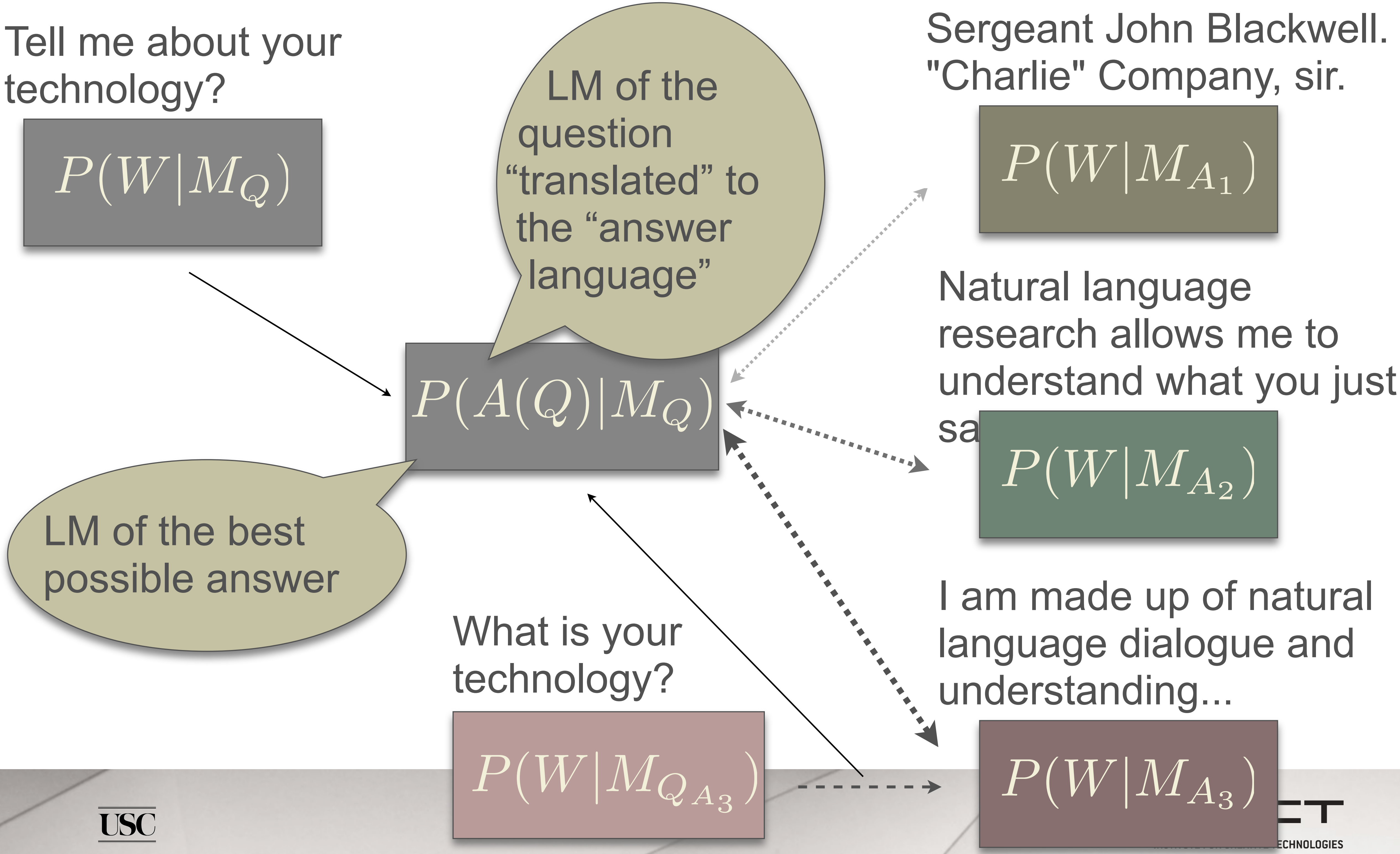
Natural language research allows me to understand what you just sa

$P(W|M_{A_2})$

I am made up of natural language dialogue and understanding...

$P(W|M_{A_3})$

Approach 3: Cross-Language Retrieval



Tell me about your technology?

$P(W|M_Q)$

LM of the question "translated" to the "answer language"

$P(A(Q)|M_Q)$

LM of the best possible answer

What is your technology?

$P(W|M_{Q_{A_3}})$

Sergeant John Blackwell. "Charlie" Company, sir.

$P(W|M_{A_1})$

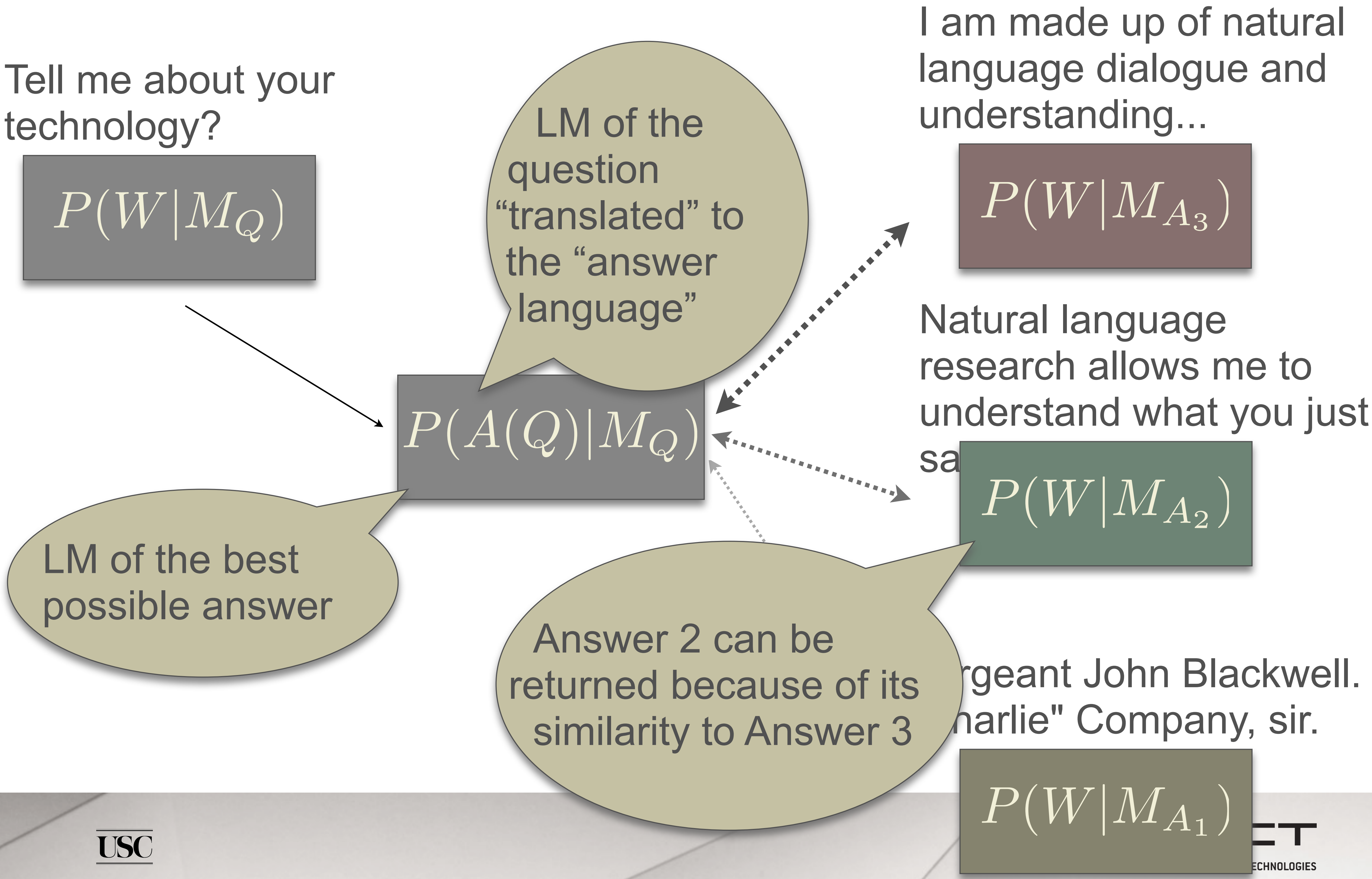
Natural language research allows me to understand what you just sa

$P(W|M_{A_2})$

I am made up of natural language dialogue and understanding...

$P(W|M_{A_3})$

Approach 3: Cross-Language Retrieval



Tell me about your technology?

$P(W|M_Q)$

LM of the question "translated" to the "answer language"

$P(A(Q)|M_Q)$

LM of the best possible answer

I am made up of natural language dialogue and understanding...

$P(W|M_{A_3})$

Natural language research allows me to understand what you just sa

$P(W|M_{A_2})$

Answer 2 can be returned because of its similarity to Answer 3

rgeant John Blackwell. harlie" Company, sir.

$P(W|M_{A_1})$

Approach 3: Cross-Language Retrieval

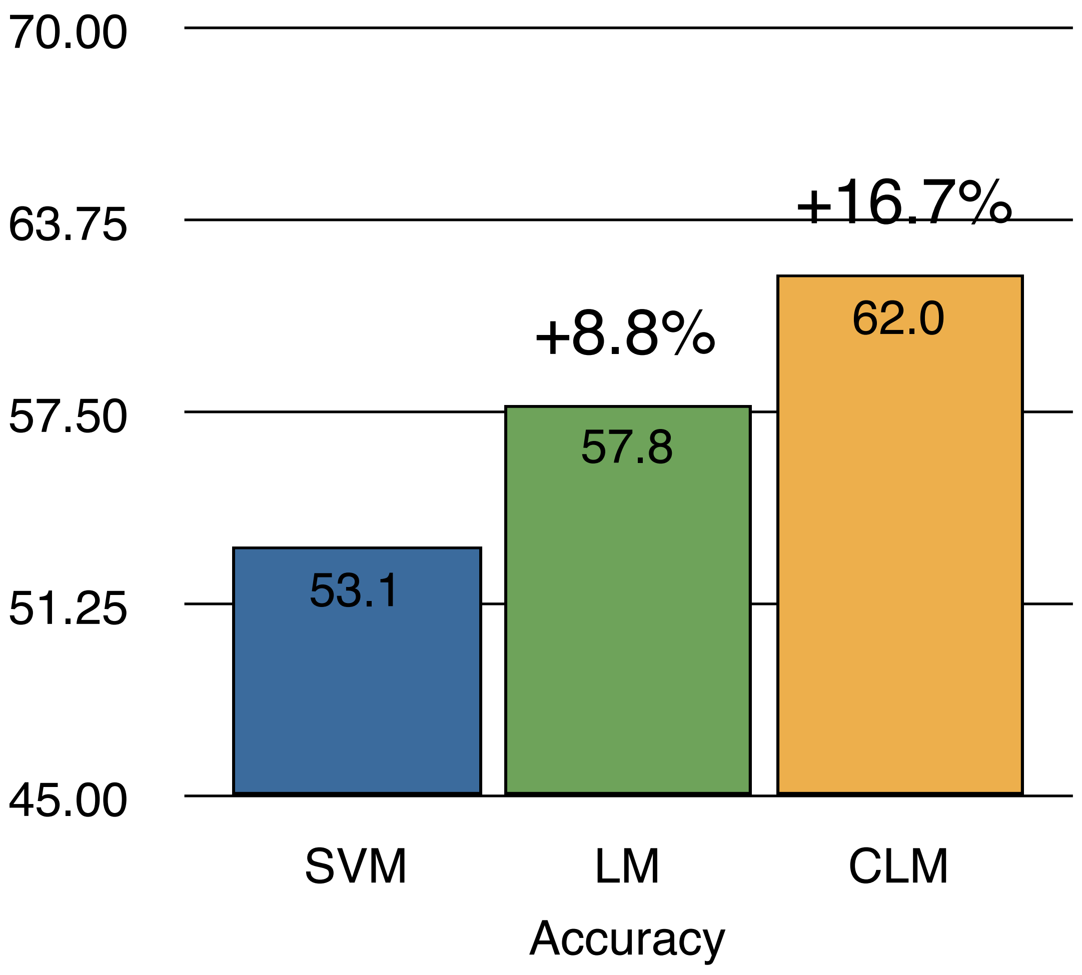
- **IR: find Chinese documents with English query**
- **Expected answer model:**

$$P(w|M_Q) = E_{(\mathcal{Q},\mathcal{A})_T}\pi_a(w) = \frac{\sum_{(q,a)\in(\mathcal{Q},\mathcal{A})_T} \pi_a(w) \prod_{i=1}^{|q|} \pi_q(q_i)}{\sum_{(q,a)\in(\mathcal{Q},\mathcal{A})_T} \prod_{i=1}^{|q|} \pi_q(q_i)}$$

- **Given a question, estimate the LM of the best possible answer, compare to known answers**
- **Similar to traditional QA, but we learn the question "translation" – they build it by hand**

Classifier Accuracy

- **1261 questions**
- **60 answers**
- **10-fold cross-validation**
- **t-test (p<0.05)**
- **differences are statistically significant**



Classifier Robustness

- **Character system based on cross-language approach**
- **20 users**
- **20 questions each = 10 predefined + 10 spontaneous**
- **2 sets of questions**
  - TRS - transcribed
  - ASR - speech recognition output
- **3 raters evaluated appropriateness of the answers on 1-6 scale**

Rating schema

Example: How do you think you can help the real army soldiers?

1. **Response is not related in any way the question**
   For me being virtual, Christmas is not a big deal. But I know it is for other real guys like me, who will be overseas this Christmas.

2. **Response has some discussion of the topic, but not the question itself.**
   Where was not I trained? That is the thing about being virtual. Infantry, Armor, Arty, Benning, Bragg, Airborne, Ranger, you name it, I can do it.

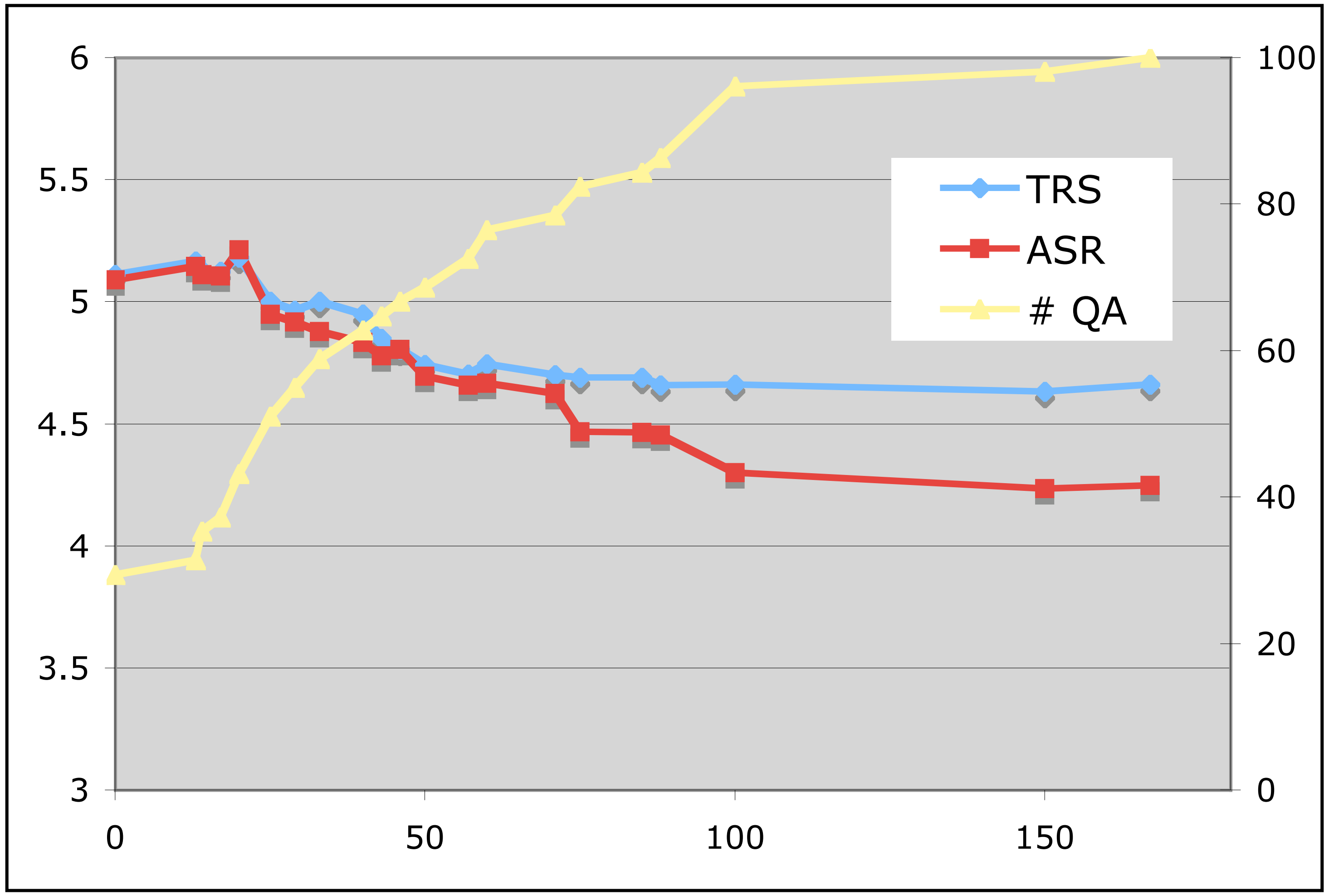
3. **Response partially addresses the question, but little or no coherence.**
   I am a virtual soldier. Which has got lots of big advantages. First of all I am readily deployable. And I can work 24-7 without a break - except periodic PMCS - maintenance. So I am pretty full-spectrum. And I am proud of it.

4. **Response addresses the question, but with major problems in coherence.**
   I am not actually flesh and blood, sir. I am pixels and polygons, 3D geometry. But that means I can be anywhere and do anything. I am a full-spectrum warrior. I can function in urban operations, food distributions, or SASO.

5. **Response does address the question, but the transition is awkward.**
   Why do you need me?! What you should be saying is "How did you get along without me?" I will show you how to be a leader, how to make critical decisions under stress...

6. **Response answers the question in a perfectly fluent manner.**
   Not to be too cocky - cause a lot of my technology is just starting to come on-line. But think of it this way: while I will never be able to do what real soldiers do, I can help my flesh and blood brethren learn how to better do their business out of the line of fire, so that they can be more capable and better prepared when they finally do get into it.

Classifier Robustness



- **Expected answer appropriateness as a function of speech recognition quality (WER, %)**

[SIGDIAL, IUI, IVA]

NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification	Keep history	Recorded answers
	**Transaction Dialogue**	Information extraction	Follow Protocol	Template-based
	**Negotiation**	Semantic parsing	Rule-based reasoning	Statistical & Grammar-based

complexity

Semantic Parsing

- **“We have move the clinic” →**

mood	declarative
speechact.type	statement
modal.deontic	must
task	move-clinic
type	event
event	move
theme	clinic
source	here
destination	there

- **We have a training set of text strings with matching semantic frames. Build the NLU.**
- **Traditional parsing fails due to ASR**
- **Two approaches**
  - frame retrieval
  - frame building

Frame Retrieval

- **Two languages:**
  - text strings (words)
  - semantic frames (slot-value pairs)
- **Given a string, retrieve the best frame**
- **Use cross-lingual LM approach**
  - Likelihood of observing a slot-value pair:

$$P(\sigma|M_S) = \frac{\sum_{(s,f)\in(\mathcal{S},\mathcal{F})_T} \pi_f(\sigma) \prod_{i=1}^{|s|} \pi_s(s_i)}{\sum_{(s,f)\in(\mathcal{S},\mathcal{F})_T} \prod_{i=1}^{|s|} \pi_s(s_i)}$$

- **Limitation: cannot produce new interpretations**

Frame Building

- **We only need to find slot-value pairs**
- **Use cross-lingual LM approach**
  - Slot-value pair likelihood

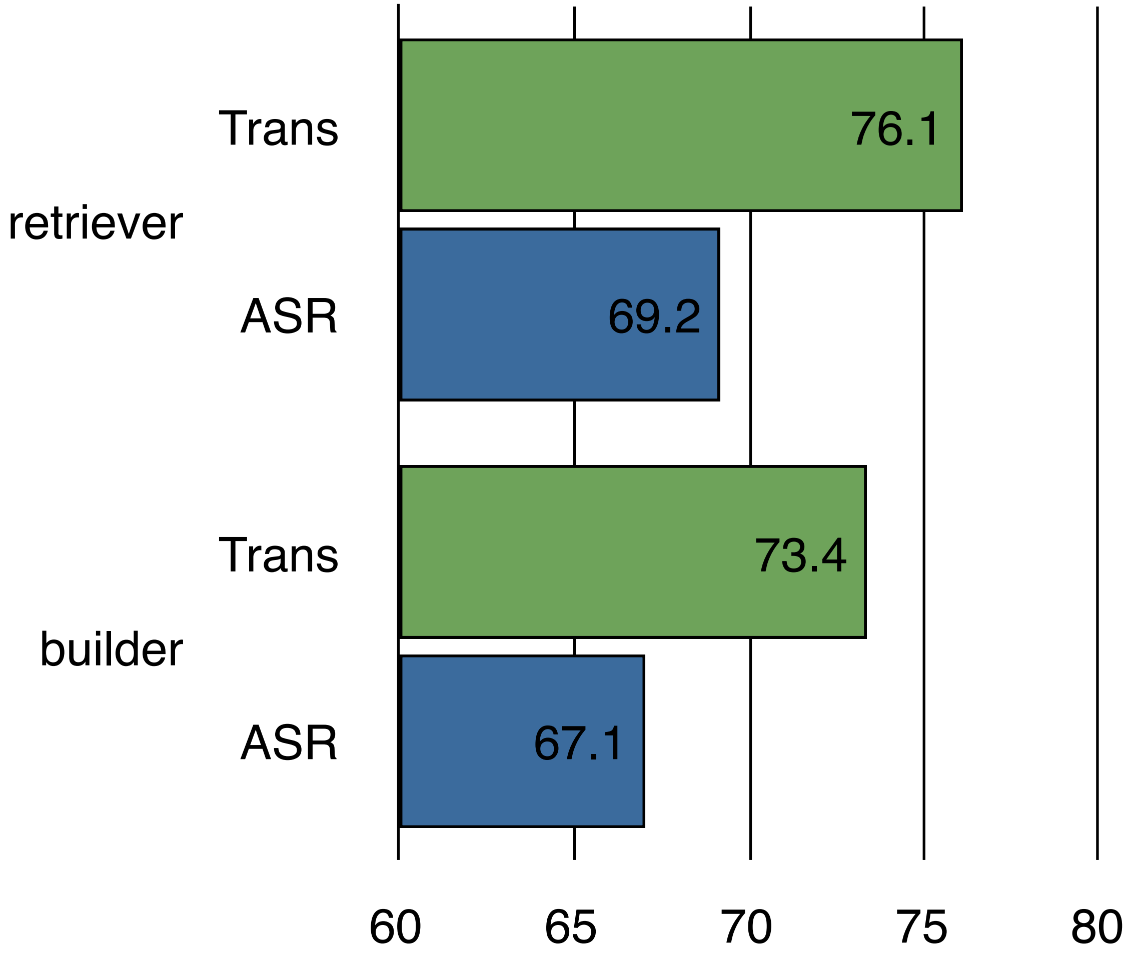
$$P(\sigma|M_S) = \frac{\sum_{(s,f)\in(\mathcal{S},\mathcal{F})_T} \pi_f(\sigma) \prod_{i=1}^{|s|} \pi_s(s_i)}{\sum_{(s,f)\in(\mathcal{S},\mathcal{F})_T} \prod_{i=1}^{|s|} \pi_s(s_i)}$$

- **Rank slot-value pairs**
- **Threshold the ranking**
- **Limitation: inconsistencies in frames**

mood	declarative
speechact.type	statement
modal.deontic	must
task	move-clinic
type	event
event	move
theme	clinic
source	here
destination	there
theme	patients
time	future
...	...

Comparison

- **1053 training utterances**
- **117 testing**
- **51 frame**
- **slot-value level accuracy: F-score**



NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification →	Keep history →	Recorded answers
	**Transaction Dialogue**	Information extraction →	Follow Protocol →	Template-based
	**Negotiation**	Semantic parsing →	Rule-based reasoning →	Statistical & Grammar-based

complexity

Language Generation

content.modality.type	desire
content.modality.desire	want
content.location	here
content.theme	clinic
content.event	operateFacility
content.type	action
content.time	future
action	assert
actor	doctor

→ **“i want to run the clinic here”**

- **Use the cross-lingual retrieval approach**
  - Likelihood of observing a word given a semantic frame

$$P(w|M_F) = \frac{\sum_{(f,s)\in(\mathcal{F},\mathcal{S})_T} \pi_s(w) \prod_{i=1}^{|f|} \pi_f(f_i)}{\sum_{(f,s)\in(\mathcal{F},\mathcal{S})_T} \prod_{i=1}^{|f|} \pi_f(f_i)}$$

- **Limitation: cannot handle totally new frames**

NL Dialogue Processing

best techniques for genre & sub-task

		Understand language	Manage dialog	Generate language
	**Question Answering**	Text classification →	Keep history →	Recorded answers
	**Transaction Dialogue**	Information extraction →	Follow Protocol →	Template-based
	**Negotiation**	Semantic parsing →	Rule-based reasoning →	Statistical & Grammar-based

complexity

Information Extraction

X:	Alpha	one	six	this	is	Bravo	two	five	adjust	fire	over
Y:	FDC	FDC	FDC	other	other	FO	FO	FO	WO	WO	K

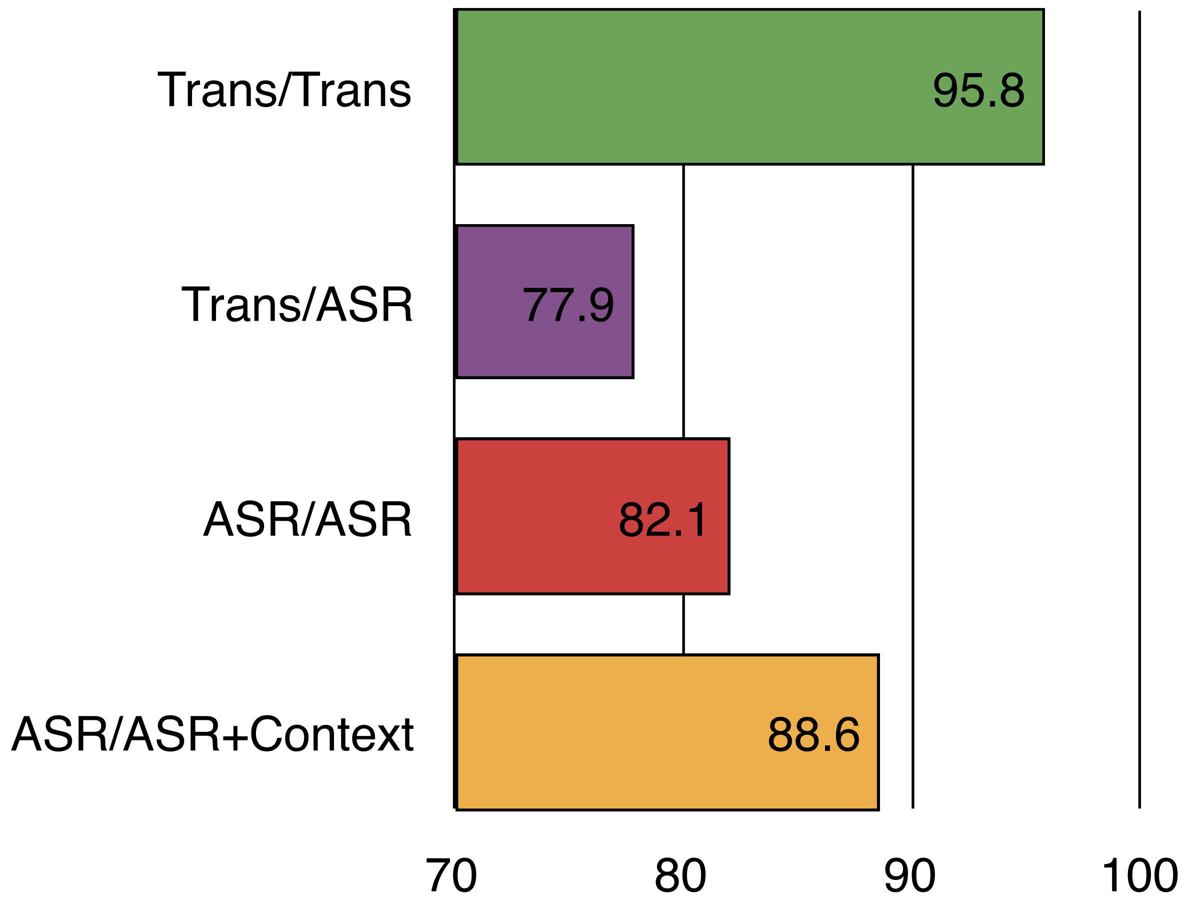
- **Markup important word sequences**
- **Maximize likelihood of observing a sequence of labels given a sequence of words: P(Y|X)**
- **Conditional Random Fields**

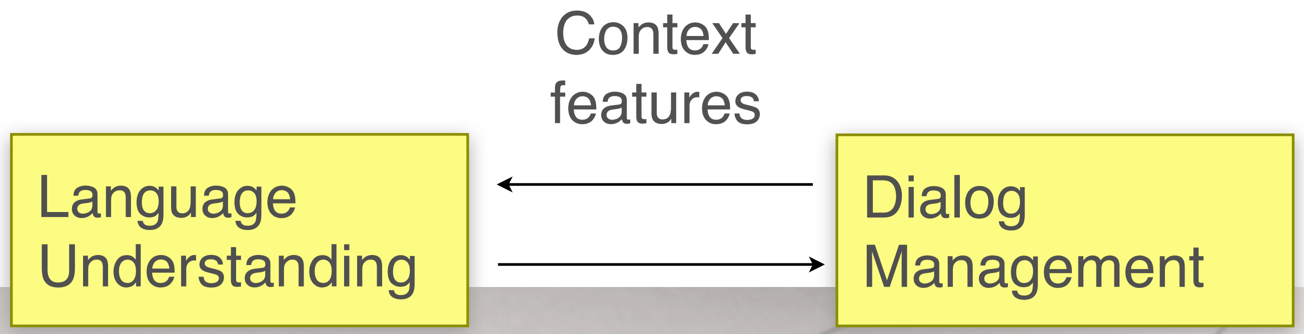
$$P(y|x) = \frac{1}{Z(x)} \exp\left\{ \sum_i \lambda_i f_i(y, x) \right\}$$

- **Lexical features – word occurrences, class**

Labeling Accuracy

- **1022 utterances**
- **10-fold cross-validation**
- **Transcribed vs ASR (WER 19%)**
- **Context features - dialog state information**





Context features

Language Understanding

Dialog Management

Language Understating Summary

- **Language Understating as a part of a virtual character**
- **Different approaches for different tasks**
  - – text classification, semantic parsing - cross-language LM
  - – information extraction - Conditional Random Fields
- **Statistical language models**
  - – accurate - outperforms state-of-the-art by 17%
  - – robust - input errors have no effect until 70% WER

Thank You

USC

INSTITUTE FOR CREATIVE TECHNOLOGIES