Detecting Small Talk

A Small Talk
by Robin Stewart
March 10, 2006

Cool (Useful) Things

- Knowledge Systems

- Question Answering

- Google

- Speech

Semantic Noise

[157] 1: Yeah, and for me, like, I'm closer to my college friends
than I am to large sections of my actual, like, biological family, so --
[158] 2: Well, you don't see 'em, probably.
[159] 1: Uh --
[160] 2: Just at holidays, do you.
[162] 1: -- I see 'em once in a while but not so much.
[163] 1: Oh gross, there's this huge bug with giant teeth on it
        outside of my house.
[164] 2: [LAUGH]
[165] 1: Wow.
[166] 1: It's got, like, these [MN] gripping teeth on it that are about
        half an inch long.
[167] 2: G- -- it's -- it's a bug?
[170] 1: Yeah, it's like a big beetle.
[171] 2: Oh, ick. [LAUGH]
[172] 2: They grow much longer down there, don't they?
[173] 1: Yeah, they do.
[174] 2: They have a longer growth. [LAUGH] I'm glad I don't --
[175] 1: Wow, I'm just starting to bite my keyring.
[176] 2: -- they -- they turn into critters, huh? [LAUGH]
[177] 1: [LAUGH] Yeah.
[178] 1: Ah, it's running away now.
[179] 1: It was flipped onto its back, so I turned it over.
[180] 2: [LAUGH]
[181] 1: Ew. [LAUGH]

[95110]

Semantic Noise

[157] 1: Yeah, and for me, like, I'm closer to my college friends
than I am to large sections of my actual, like, biological family, so --
[158] 2: Well, you don't see 'em, probably.
[159] 1: Uh --
[160] 2: Just at holidays, do you.
[162] 1: -- I see 'em once in a while but not so much.
[163] 1: Oh gross, there's this huge bug with giant teeth on it
         outside of my house.
[164] 2: [LAUGH]
[165] 1: Wow.
[166] 1: It's got, like, these [MN] gripping teeth on it that are about
         half an inch long.
[167] 2: G- -- it's -- it's a bug?
[170] 1: Yeah, it's like a big beetle.
[171] 2: Oh, ick. [LAUGH]
[172] 2: They grow much longer down there, don't they?
[173] 1: Yeah, they do.
[174] 2: They have a longer growth. [LAUGH] I'm glad I don't --
[175] 1: Wow, I'm just starting to bite my keyring.
[176] 2: -- they -- they turn into critters, huh? [LAUGH]
[177] 1: [LAUGH] Yeah.
[178] 1: Ah, it's running away now.
[179] 1: It was flipped onto its back, so I turned it over.
[180] 2: [LAUGH]
[181] 1: Ew. [LAUGH]

[95110]

Goals

Goals

Can we model this phenomenon?

Goals

Can we model this phenomenon?

Can we automatically detect irrelevant regions?

Approach

Related research on small talk and conversational speech
- Greetings and Partings
- Past tense stories, present tense small talk

Approach

Related research on small talk and conversational speech
- Greetings and Partings
- Past tense stories, present tense small talk

My approach: Apply machine learning techniques to conversations to see if these suggestions are actually useful, or if other attributes are better.
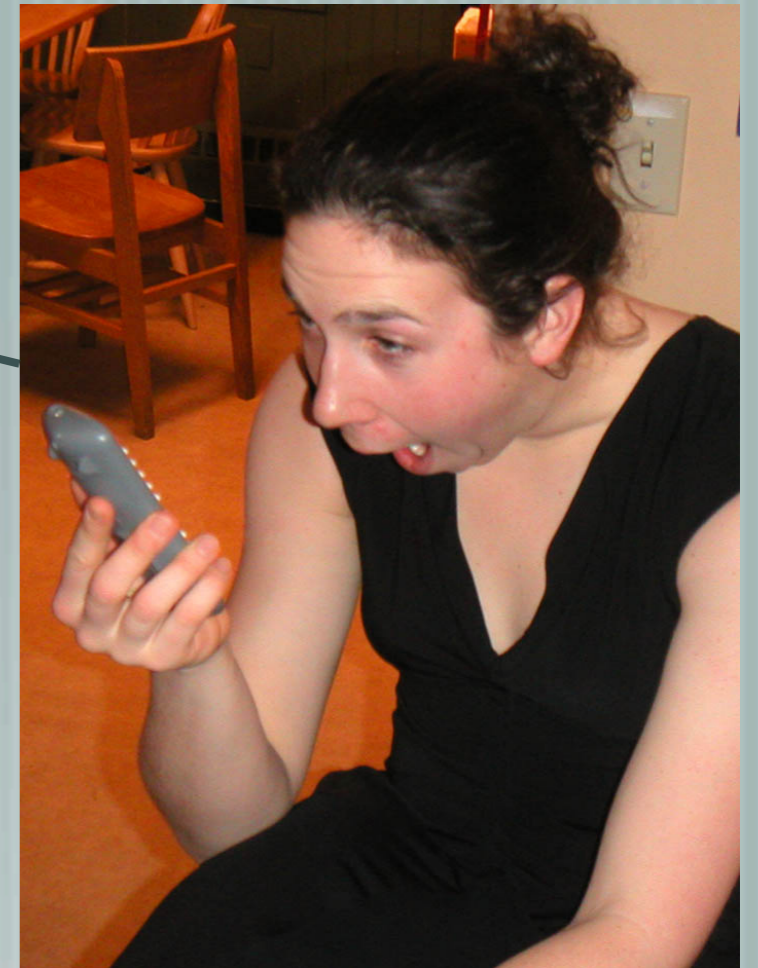
Data
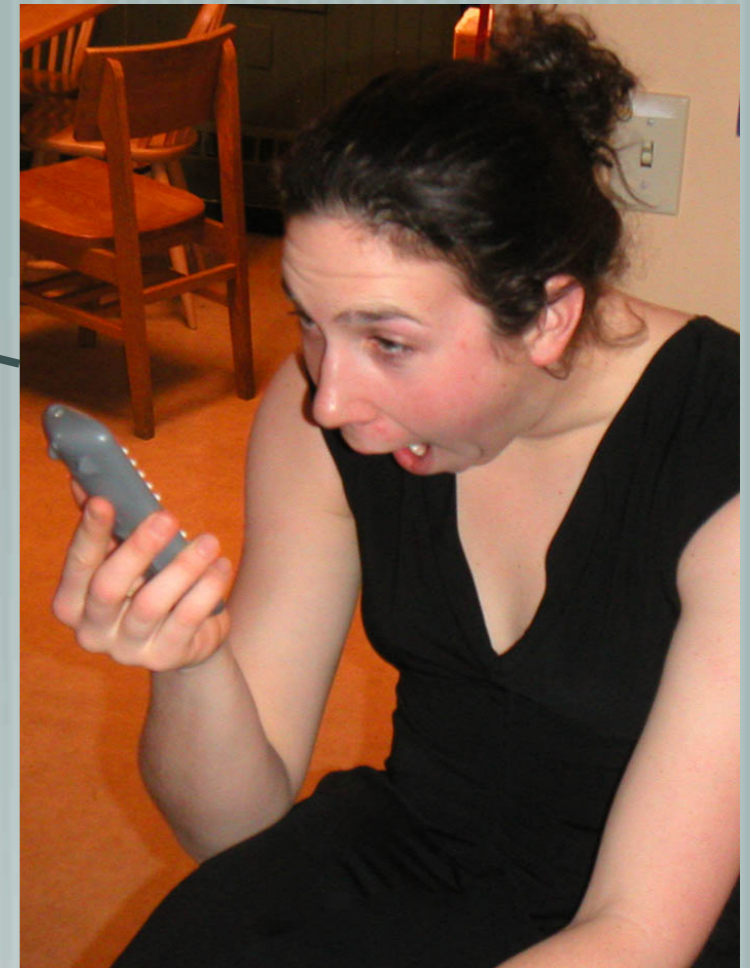
Data

Data

Automatic
Operator

Data

Automatic
Operator

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

Automatic Operator

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

Automatic Operator

10 mins

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

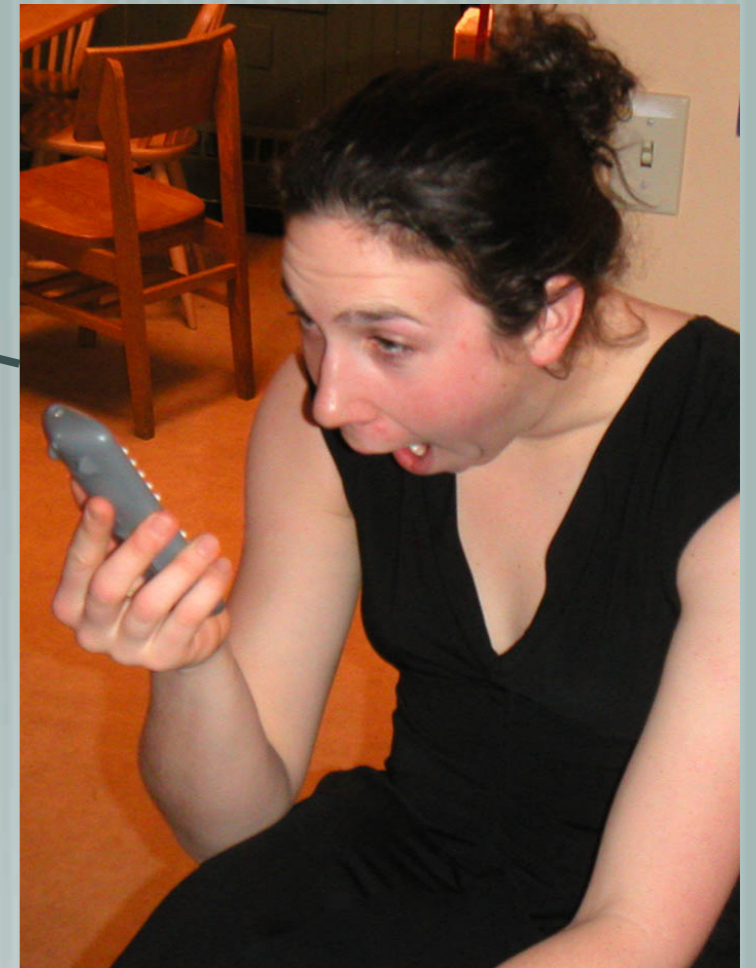Automatic Operator

10 mins

Recorded

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

Automatic Operator

10 mins
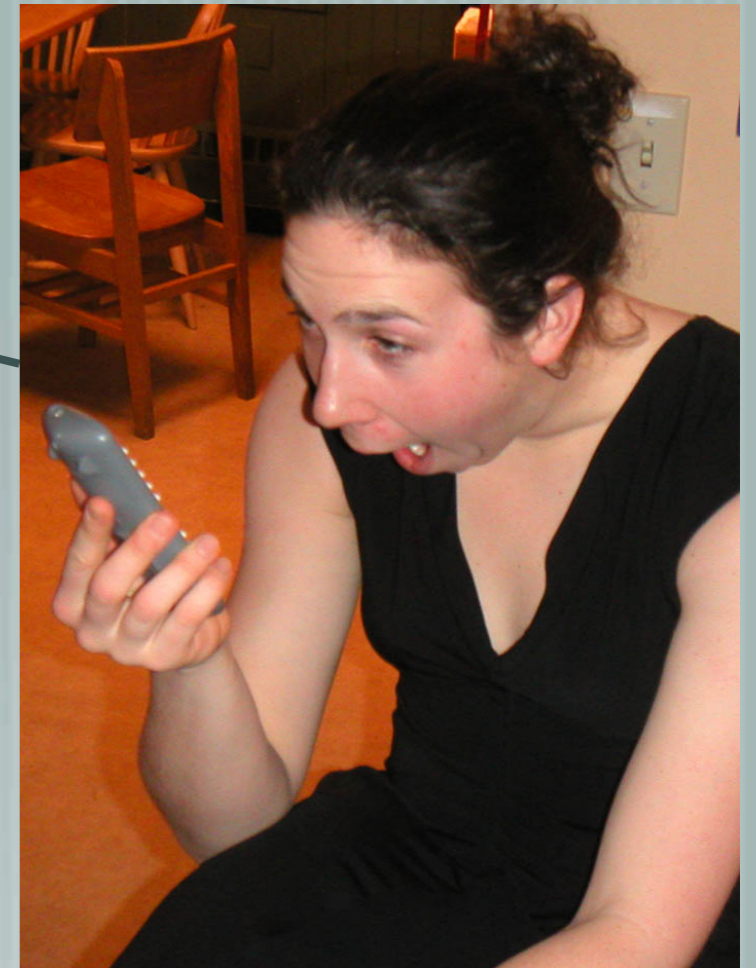
Recorded

Transcribed

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

Automatic Operator

10 mins

Recorded

Transcribed

Billing

Data

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"
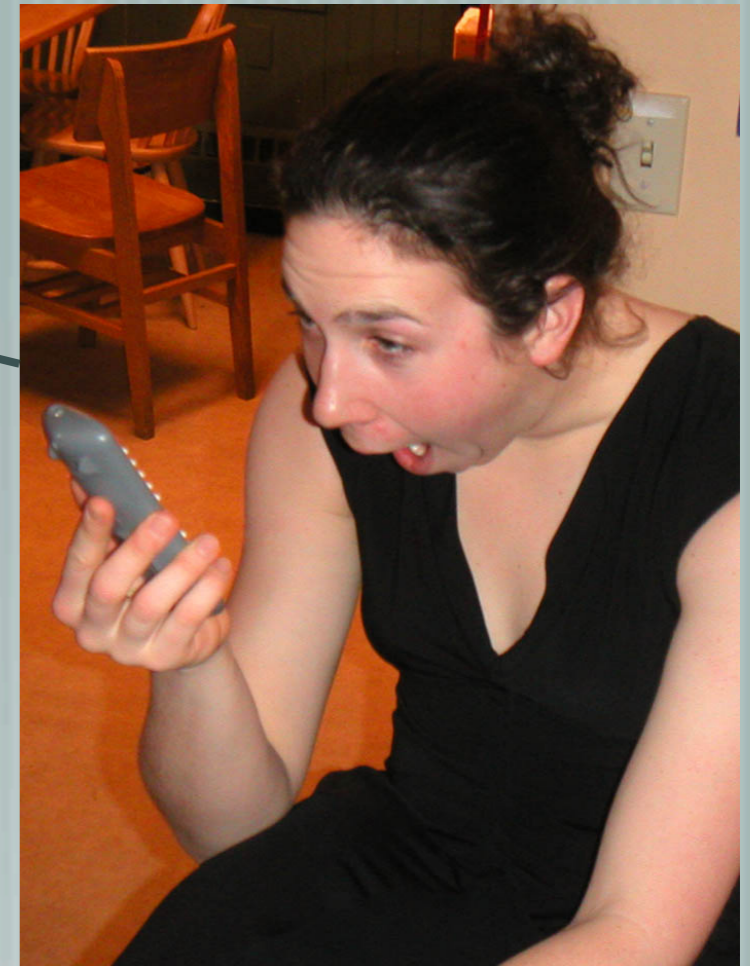
Automatic Operator

10 mins
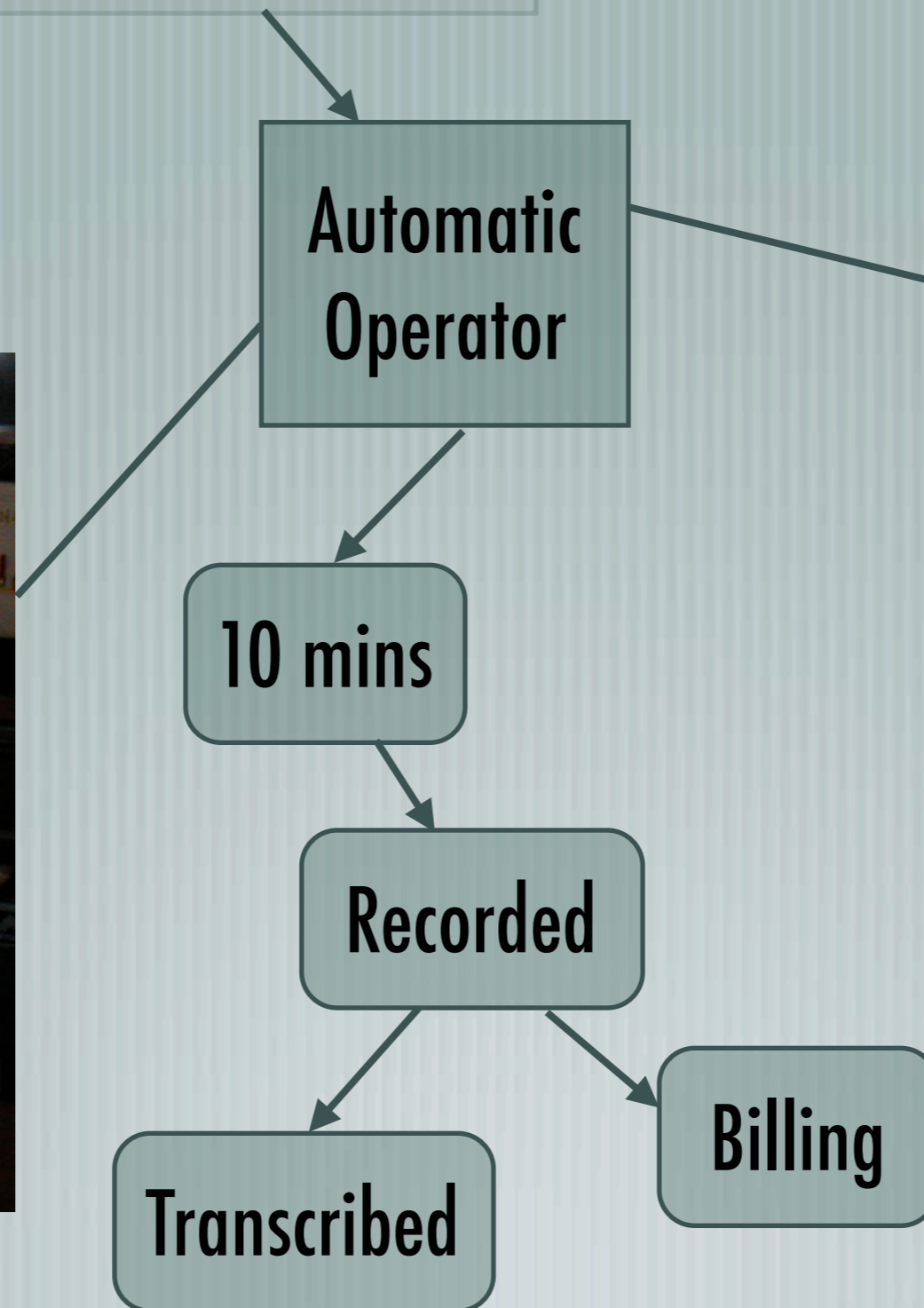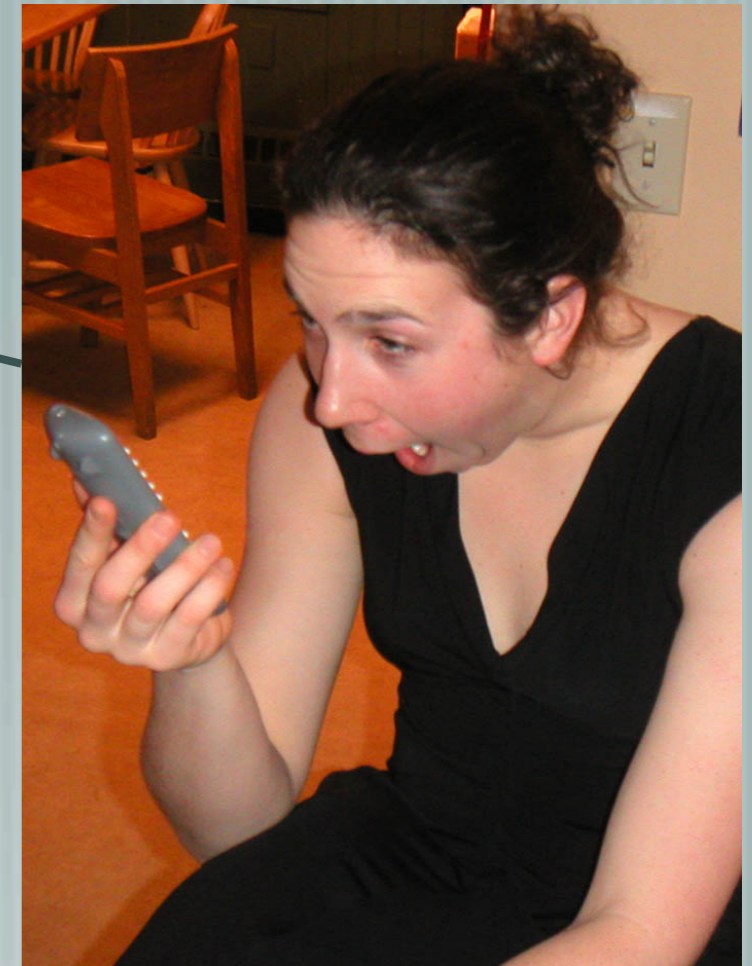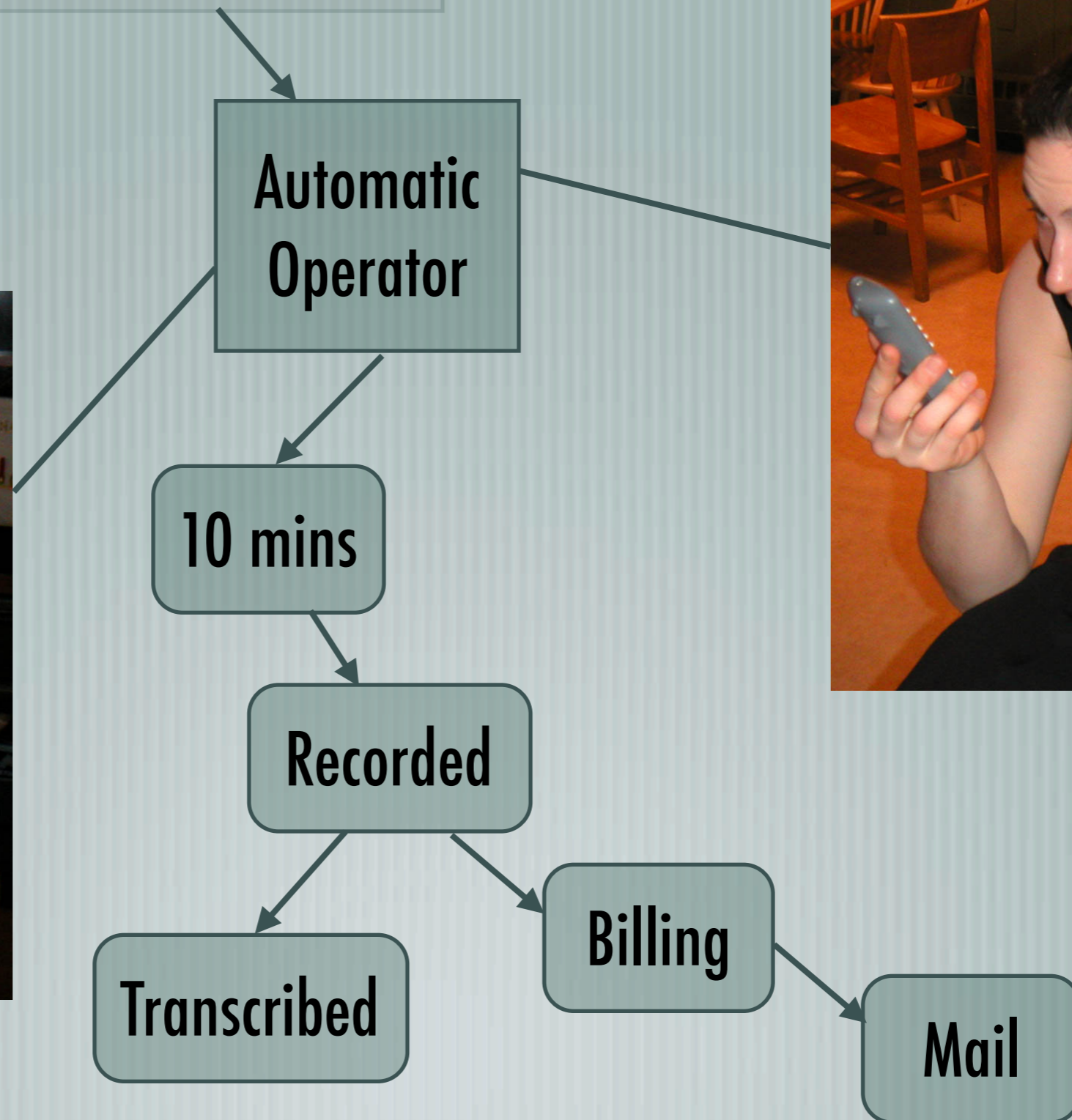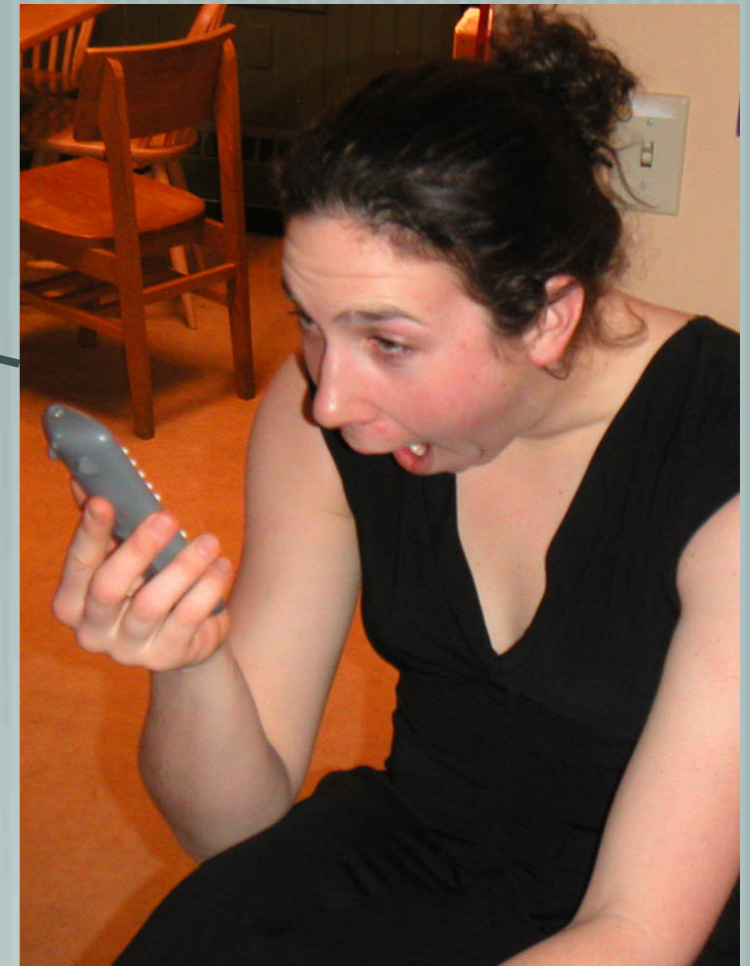
Recorded

Transcribed

Billing

Mail

$10

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

[1] 2: Hello.
[2] 1: Hello.
[3] 2: Hello.
[4] 2: I'm Richard.
[5] 1: Ah, my name is Devin.
[6] 2: Hi, Devin.
[7] 1: Hi.
[8] 2: Have you got a pet?
[9] 1: Huh?
[10] 2: Do you have a -- a pet?
[11] 1: I have two turtles.
[12] 2: You do?
[13] 1: Yeah.
[14] 2: Well, what are their names?
[15] 1: Um, I don't think they have names.
[16] 1: We never gave 'em names.
[17] 1: We had 'em for a long time and --
       I don't know.
[18] 1: Th- they don't have names.
[19] 2: Yeah -- yeah.
[20] 2: Well, I got two dogs.
[21] 1: Two dogs.
[22] 1: What's their names?
[23] 2: Ah, Hollywood and Twinkles.
[24] 1: Hollywood and --
[25] 2: Great names, aren't they?

> "Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

Small Talk
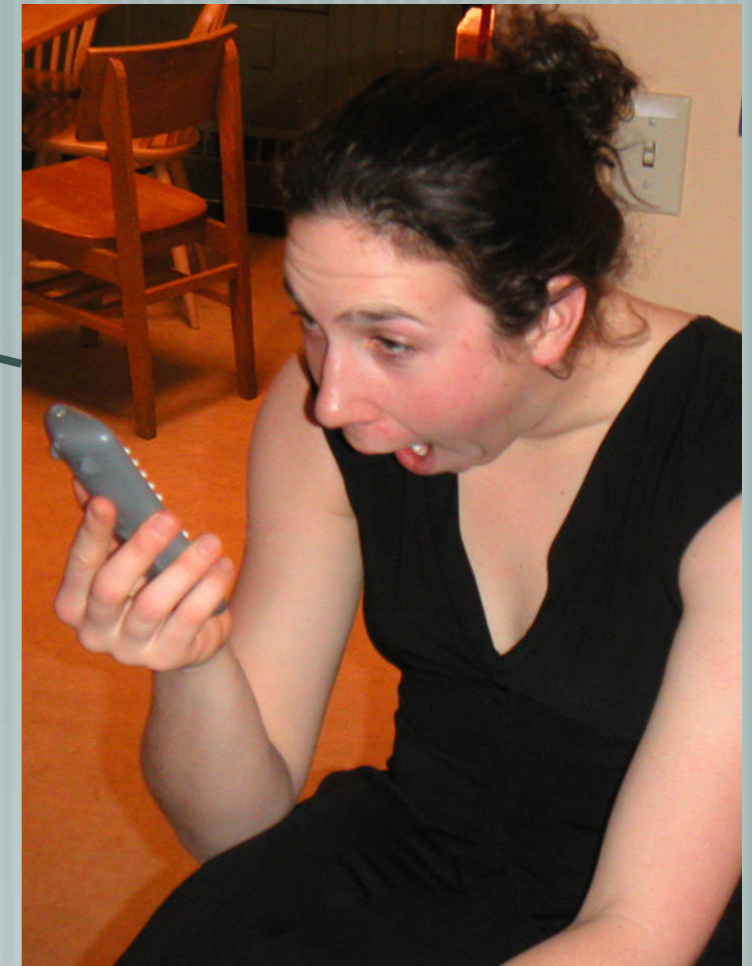
[1] 2: Hello.
[2] 1: Hello.
[3] 2: Hello.
[4] 2: I'm Richard.
[5] 1: Ah, my name is Devin.
[6] 2: Hi, Devin.
[7] 1: Hi.
[8] 2: Have you got a pet?
[9] 1: Huh?
[10] 2: Do you have a -- a pet?
[11] 1: I have two turtles.
[12] 2: You do?
[13] 1: Yeah.
[14] 2: Well, what are their names?
[15] 1: Um, I don't think they have names.
[16] 1: We never gave 'em names.
[17] 1: We had 'em for a long time and --
       I don't know.
[18] 1: Th- they don't have names.
[19] 2: Yeah -- yeah.
[20] 2: Well, I got two dogs.
[21] 1: Two dogs.
[22] 1: What's their names?
[23] 2: Ah, Hollywood and Twinkles.
[24] 1: Hollywood and --
[25] 2: Great names, aren't they?

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"
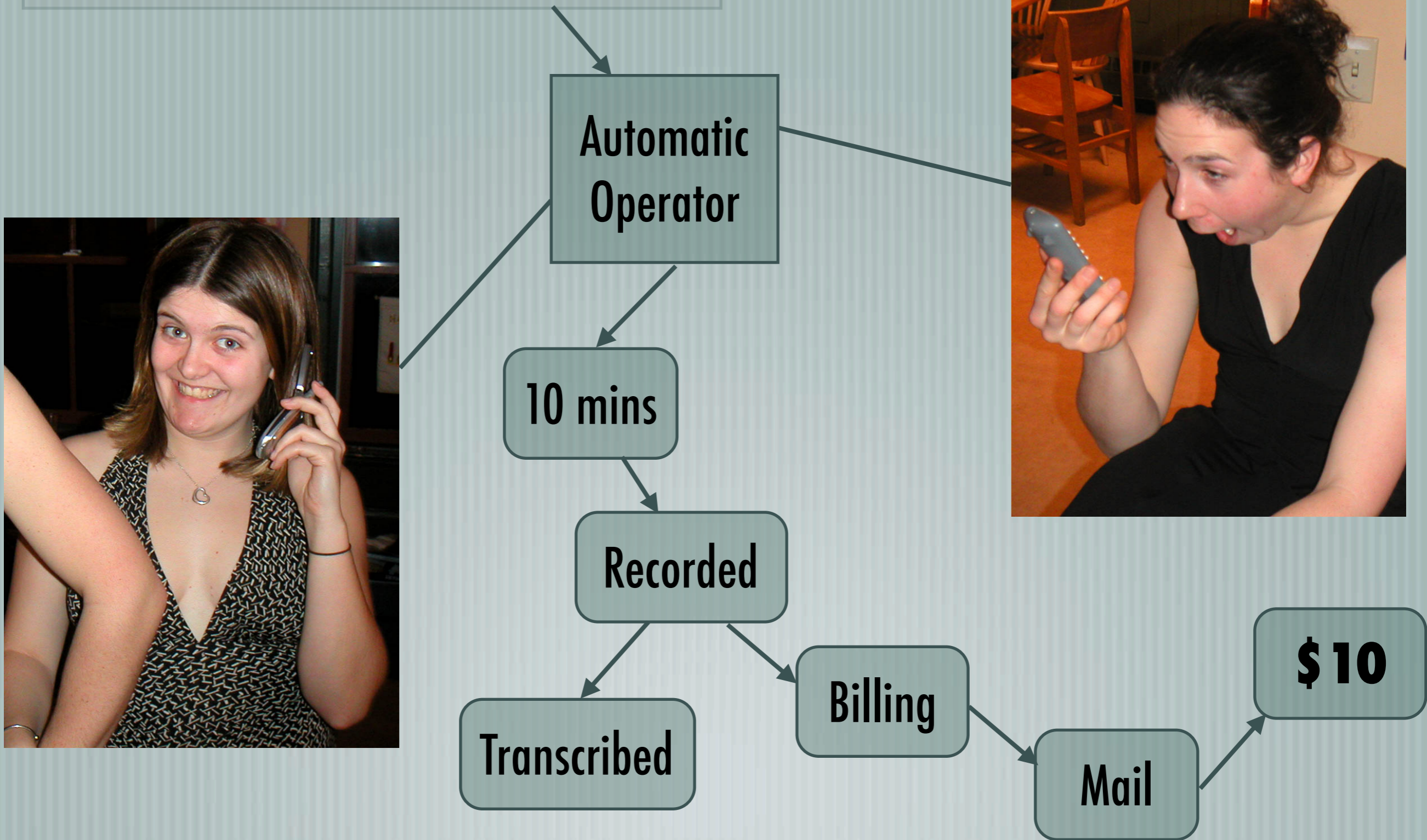
Small Talk

[1] 2: Hello.
[2] 1: Hello.
[3] 2: Hello.
[4] 2: I'm Richard.
[5] 1: Ah, my name is Devin.
[6] 2: Hi, Devin.
[7] 1: Hi.
[8] 2: Have you got a pet?
[9] 1: Huh?
[10] 2: Do you have a -- a pet?
[11] 1: I have two turtles.
[12] 2: You do?
[13] 1: Yeah.
[14] 2: Well, what are their names?
[15] 1: Um, I don't think they have names.
[16] 1: We never gave 'em names.
[17] 1: We had 'em for a long time and --
         I don't know.
[18] 1: Th- they don't have names.
[19] 2: Yeah -- yeah.
[20] 2: Well, I got two dogs.
[21] 1: Two dogs.
[22] 1: What's their names?
[23] 2: Ah, Hollywood and Twinkles.
[24] 1: Hollywood and --
[25] 2: Great names, aren't they?

[73085]

"Do either of you have a pet? If so, how much time each day do you spend with your pet? How important is your pet to you?"

[1] 2: Hello.
[2] 1: Hello.
[3] 2: Hello.
[4] 2: I'm Richard.
[5] 1: Ah, my name is Devin.
[6] 2: Hi, Devin.
[7] 1: Hi.
[8] 2: Have you got a pet?
[9] 1: Huh?
[10] 2: Do you have a -- a pet?
[11] 1: I have two turtles.
[12] 2: You do?
[13] 1: Yeah.
[14] 2: Well, what are their names?
[15] 1: Um, I don't think they have names.
[16] 1: We never gave 'em names.
[17] 1: We had 'em for a long time and --
         I don't know.
[18] 1: Th- they don't have names.
[19] 2: Yeah -- yeah.
[20] 2: Well, I got two dogs.
[21] 1: Two dogs.
[22] 1: What's their names?
[23] 2: Ah, Hollywood and Twinkles.
[24] 1: Hollywood and --
[25] 2: Great names, aren't they?

Small Talk

On-Topic

[73085]

Do either of you think public or private schools have the right to forbid students to read certain books?

[145] 1: That's a rough one. [LAUGH]
[146] 2: I know. [LAUGH]
[147] 1: I mean, I was on the message board about this thing and some people were talking about, like, pets; how much time they spend with their dogs and stuff like that.
[148] 1: I'm, like, I want that one. [LAUGH]
[149] 2: [LAUGH]
[150] 1: And this is my first one and I'm, like, holy crap. [LAUGH]
[151] 2: [LAUGH]
[152] 1: This is stuff that people debate over for, like, weeks, you know.
[153] 2: [LAUGH] I know.

Do either of you think public or private schools have the right to forbid students to read certain books?

[145] 1: That's a rough one. [LAUGH]
[146] 2: I know. [LAUGH]
[147] 1: I mean, I was on the message board about this thing and some people were talking about, like, pets; how much time they spend with their dogs and stuff like that.
[148] 1: I'm, like, I want that one. [LAUGH]
[149] 2: [LAUGH]
[150] 1: And this is my first one and I'm, like, holy crap. [LAUGH]
[151] 2: [LAUGH]
[152] 1: This is stuff that people debate over for, like, weeks, you know.
[153] 2: [LAUGH] I know.

> Do either of you think public or private schools have the right to forbid students to read certain books?

[145] 1: That's a rough one. [LAUGH]

[146] 2: I know. [LAUGH]

[147] 1: I mean, I was on the message board about this thing and some people were talking about, like, pets; how much time they spend with their dogs and stuff like that.

[148] 1: I'm, like, I want that one. [LAUGH]

[149] 2: [LAUGH]

[150] 1: And this is my first one and I'm, like, holy crap. [LAUGH]

[151] 2: [LAUGH]

[152] 1: This is stuff that people debate over for, like, weeks, you know.

[153] 2: [LAUGH] I know.

Metaconversation

[91930]

Detection

- Get conversations

- Annotate

- Generate Features

- Run classifier learning algorithm using training set

- Classify test set and evaluate accuracy

Detection

Get conversations

Annotate

Generate Features

Run classifier learning algorithm using training set

Classify test set and evaluate accuracy

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

- Bag of words (with counts)

Generating Features

- Words

	"hi"	"a"	"it"	"virginia"
[167] 2: G- -- it's -- it's a bug?	0	1	2	0
[170] 1: Yeah, it's like a big beetle.	0	1	1	0
[171] 2: Oh, ick. [LAUGH]	0	0	0	0

- Bag of words (with counts)

- Choosing words: Lewis (1994)

Additional Features

- ## Parts of speech

	PRP	UH	VBZ	.
g-/JJ --/: it/PRP 's/VBZ --/: it/PRP 's/VBZ a/DT bug/NN ?/.	2	0	2	0
yeah/UH ,/, it/PRP 's/VBZ like/IN a/DT big/JJ beetle/NN ./.	1	1	1	1
oh/UH ,/, ick/NN ./.	0	1	0	1

Additional Features

- ## Parts of speech

	PRP	UH	VBZ	.
g-/JJ --/: it/PRP 's/VBZ --/: it/PRP 's/VBZ a/DT bug/NN ?/.	2	0	2	0
yeah/UH ,/, it/PRP 's/VBZ like/IN a/DT big/JJ beetle/NN ./.	1	1	1	1
oh/UH ,/, ick/NN ./.	0	1	0	1

Additional Features

- ## Parts of speech

	PRP	UH	VBZ	.
g-/JJ --/: it/PRP 's/VBZ --/: it/PRP 's/VBZ a/DT bug/NN ?/.	2	0	2	0
yeah/UH ,/, it/PRP 's/VBZ like/IN a/DT big/JJ beetle/NN ./.	1	1	1	1
oh/UH ,/, ick/NN ./.	0	1	0	1

Additional Features

- ## Parts of speech

	PRP	UH	VBZ	.
g-/JJ --/: it/PRP 's/VBZ --/: it/PRP 's/VBZ a/DT bug/NN ?/.	2	0	2	0
yeah/UH ,/, it/PRP 's/VBZ like/IN a/DT big/JJ beetle/NN ./.	1	1	1	1
oh/UH ,/, ick/NN ./.	0	1	0	1

- ## Line number
- ## Sentence type
- ## Number of laughs
- ## Number of words
- ## Previous/Next 5 sentences

Detection

Get conversations

Annotate

Generate Features

Run classifier learning algorithm using training set

Classify test set and evaluate accuracy

Experimental setup

5 topics

1	2	3	4	5

Experimental setup

5 topics

4 conversations in each topic

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

Experimental setup

5 topics

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

4 conversations in each topic

= 20 conversations

Experimental setup

5 topics

4 conversations in each topic

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

= 20 conversations

- 4 fold cross-validation

Experimental setup

5 topics

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

4 conversations in each topic

= 20 conversations

- 4 fold cross-validation

- Classification of sentences

Experimental setup

5 topics

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

4 conversations in each topic

= 20 conversations

- ## 4 fold cross-validation

- ## Classification of sentences
  - Naive bayes, SVM, k-NN, C4.5 decision tree, Neural nets

Experimental setup

5 topics

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

4 conversations in each topic

= 20 conversations

- 4 fold cross-validation

- Classification of sentences
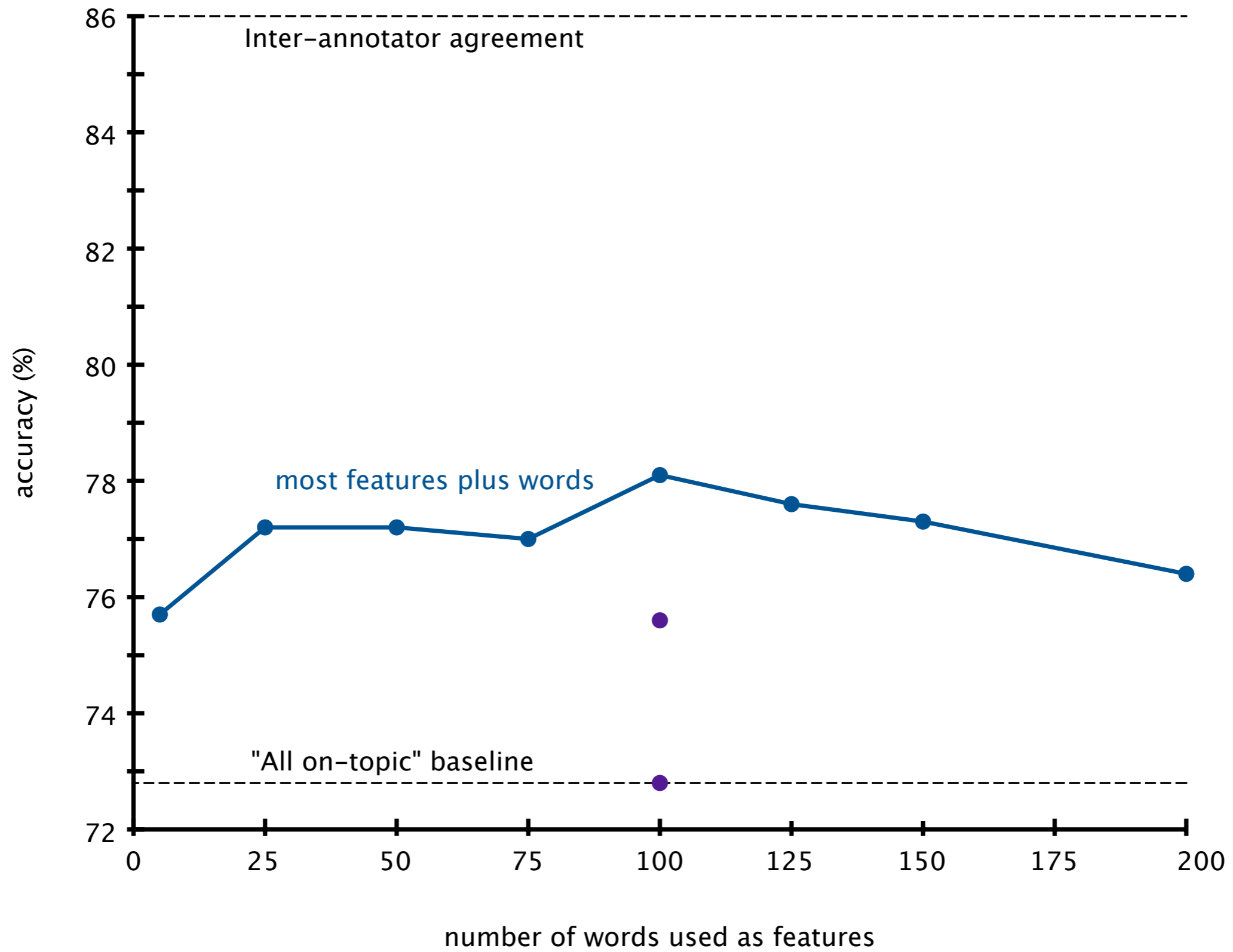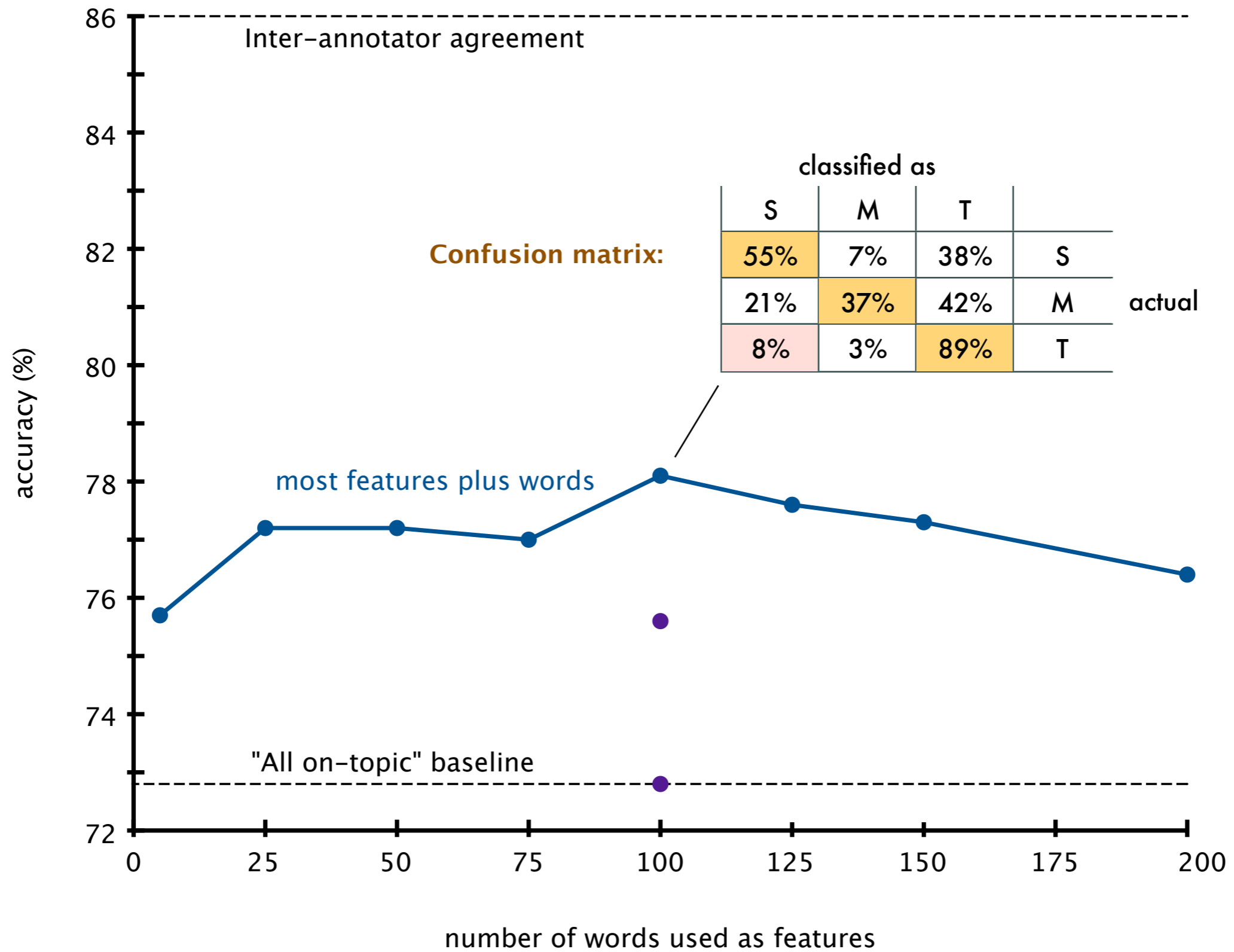 - Naive bayes, SVM, k-NN, C4.5 decision tree, Neural nets

- Inter-annotator agreement: 86% of sentences
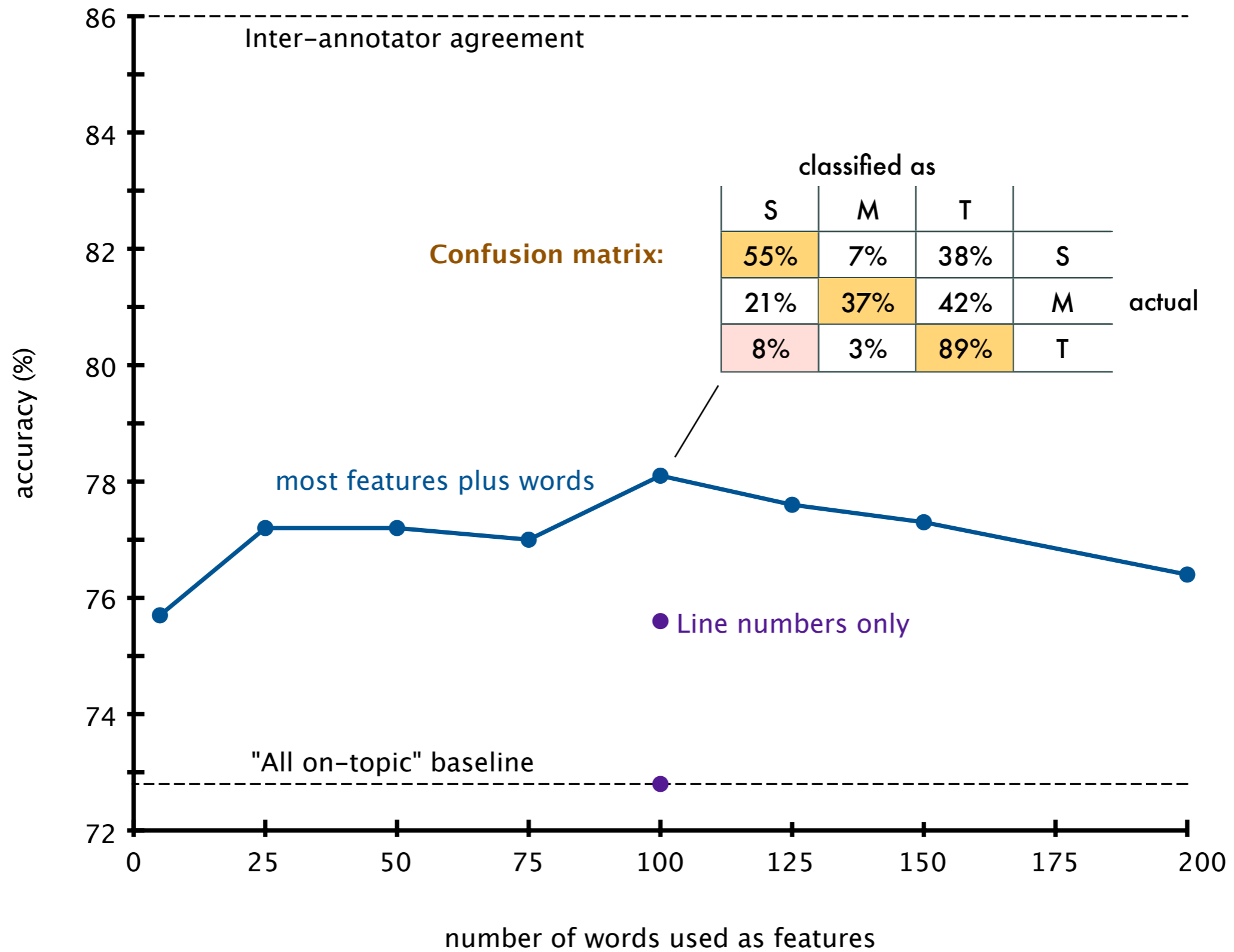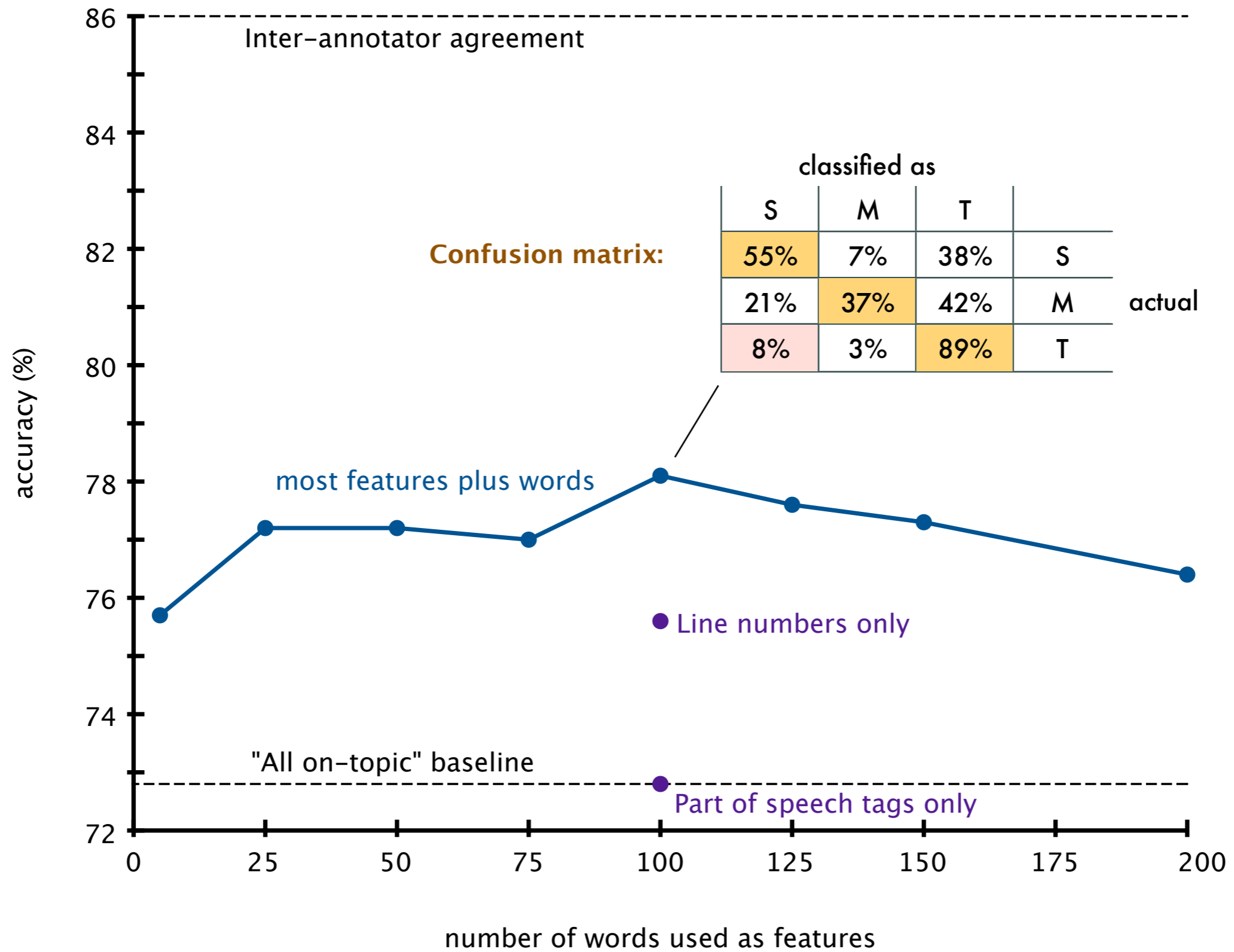
Top 20 most useful words for identifying regions

Top 20 most useful words for identifying regions

Small Talk

hi
hello
mike
missouri
possum
lisa
colder
chris
appointees
virginia
tarrant
rental
linda
lap
kaitlyn
illinois
gap
web
david
alabama

Top 20 most useful words for identifying regions

Small Talk

hi
hello
mike
missouri
possum
lisa
colder
chris
appointees
virginia
tarrant
rental
linda
lap
kaitlyn
illinois
gap
web
david
alabama

Metaconversation

cellphone
signed
strangers
protestors
fisher
topic
conversation
scam
recording
protests
penn
notion
internationally
handheld
calls
disconnected
was:
wallet
scams
refund

Top 20 most useful words for identifying regions

Small Talk	Metaconversation	On-Topic
hi	**cellphone**	**ahem**
hello	signed	happening
mike	strangers	black
missouri	protestors	skiing
possum	fisher	terms
lisa	**topic**	teachers
colder	**conversation**	keeping
chris	scam	fish
appointees	**recording**	sisters
virginia	protests	parent
tarrant	penn	employees
rental	notion	touch
linda	internationally	interests
lap	**handheld**	expect
kaitlyn	**calls**	cards
illinois	**disconnected**	solitaire
gap	was:	siamese
web	wallet	nuclear
david	scams	basic
alabama	refund	religion

Top 20 most useful words for identifying regions

Small Talk	Metaconversation	On-Topic
hi	**cellphone**	**ahem**
hello	signed	happening
mike	strangers	black
missouri	protestors	skiing
possum	fisher	terms
lisa	**topic**	teachers
colder	**conversation**	keeping
chris	scam	fish
appointees	**recording**	sisters
virginia	protests	parent
tarrant	penn	employees
rental	notion	touch
linda	internationally	interests
lap	**handheld**	expect
kaitlyn	**calls**	cards
illinois	**disconnected**	solitaire
gap	was:	siamese
web	wallet	nuclear
david	scams	basic
alabama	refund	religion

What's next

More annotated data

Active Learning

Segmentation-inspired approach

What's next

More annotated data

Active Learning

Segmentation-inspired approach

thanks!