

Predicting Speech Acts in MOOC Forum Posts

Jaime Arguello and Kyle Shaffer

School of Information and Library Science
University of North Carolina at Chapel Hill



ICWSM 2015

Background and Motivation

- **Massive Open Online Course (MOOC):** an online course designed for open, large-scale enrollment.
 - ▶ students interact with each other and the instructor(s) through the MOOC's discussion forums
- **Challenge:** the large number of posts makes it difficult for instructors to know where to intervene to answer questions, resolve issues, and provide feedback
- **Objective:** automatically predicting the speech acts present in a MOOC forum post
- **Speech Act:** the function or purpose of a post in the thread
- **Motivation:** speech acts can help identify posts that require an instructor's attention

Speech Acts

- **Question:** requests information about the course content
- **Answer:** contributes information in response to a question
- **Issue:** expresses a problem with the course management
- **Issue Resolution:** attempts to resolve a previously raised issue
- **Positive Ack:** positive sentiment about a previous post
- **Negative Ack:** negative sentiment about a previous post
- **Other:** serves a different purpose

Speech Acts

Speech Act	Example
Question	“In Question 8 on the assignment I’m confused about the code formatting. In lectures, the instructor said syntax should be of the form X, but do you have to include Y? Any ideas what I’m doing wrong?”
Answer	“The answer here should follow the form of the practice problems. Hopefully that helps.”
Issue	“The wording for Question 6 was confusing and ambiguous. Please consider revising the wording or giving students the points for this question.”
Issue Res.	“We are aware of a glitch in our submission form for Homework 2. As a result, the last question has been awarded to each student as a free point.”
Positive Ack.	“I’m glad I’m not the only one stuck on this! That was definitely confusing me too!”
Negative Ack.	“The last question may have been difficult, but part of learning new material is working at it. No sense in complaining.”
Other	“Hi everyone! I’m a web designer and extremely interested in this course!”

Research Questions

- **RQ1:** Can non-expert crowdsourced workers reliably label MOOC forum posts using our speech act definitions?
- **RQ2:** Can speech acts help predict instructor interventions and student assignment completion and performance?
- **RQ3:** What are the most predictive types of features for different speech acts?

MOOC Forum Dataset



- Metadata (8 weeks, August - November 2013)
- ~27K students (~1.5K earned Statement of Accomplishment)
- 2,943 posts (2,754 from students, 189 from course staff)

Speech Act Annotation



- Each post was labeled by 5 redundant MTurk workers
- Workers were asked to select at least one speech act per post
- Posts could be associated with multiple speech acts
- A post was associated with a speech act if at least 3/5 agreed.

Speech Act Labeling

Please examine the thread below, paying special attention to the highlighted post. When you are ready to label the highlighted post, make your selection by clicking the appropriate checkbox(es) at the right of the page. Detailed instructions are available [here](#).

Anonymous - [6 months ago](#).

It seems to me that the book was used as an illustration of a controlled vocabulary thesaurus. He wasn't trying to teach us how to use the LCSH in a professional sense for cataloging purposes.

What is the difference in terminology that you mention?

↓ 0 ↑

Anonymous - [6 months ago](#).

Exactly. We are not being being taught to catalogue and LCSH is being used as an example. I imagine one consideration is that the formatting of the print version is clear and easy to read and also fits nicely onto the screen. If you were running the course you may have chosen a different example for a different reason, but there is no need to complain because the lecturer didn't do as you would.

↓ 0 ↑

☐ Question

☐ Answer

☐ Issue

☐ Issue Resolution

☐ Pos Acknowledgement

☐ Neg Acknowledgement

☐ Non-English or N/A

Please provide an explanation for your choice(s) below. Then click 'Submit'.

Submit

RQ1: Annotation Results

	MTurk Workers	MV and Expert
	κ_f	κ_c
Question	0.569	0.893
Answer	0.414	0.790
Issue	0.421	0.669
Issue Resolution	0.286	0.635
Positive Ack.	0.423	0.768
Negative Ack.	0.232	0.633
Other	0.337	0.625

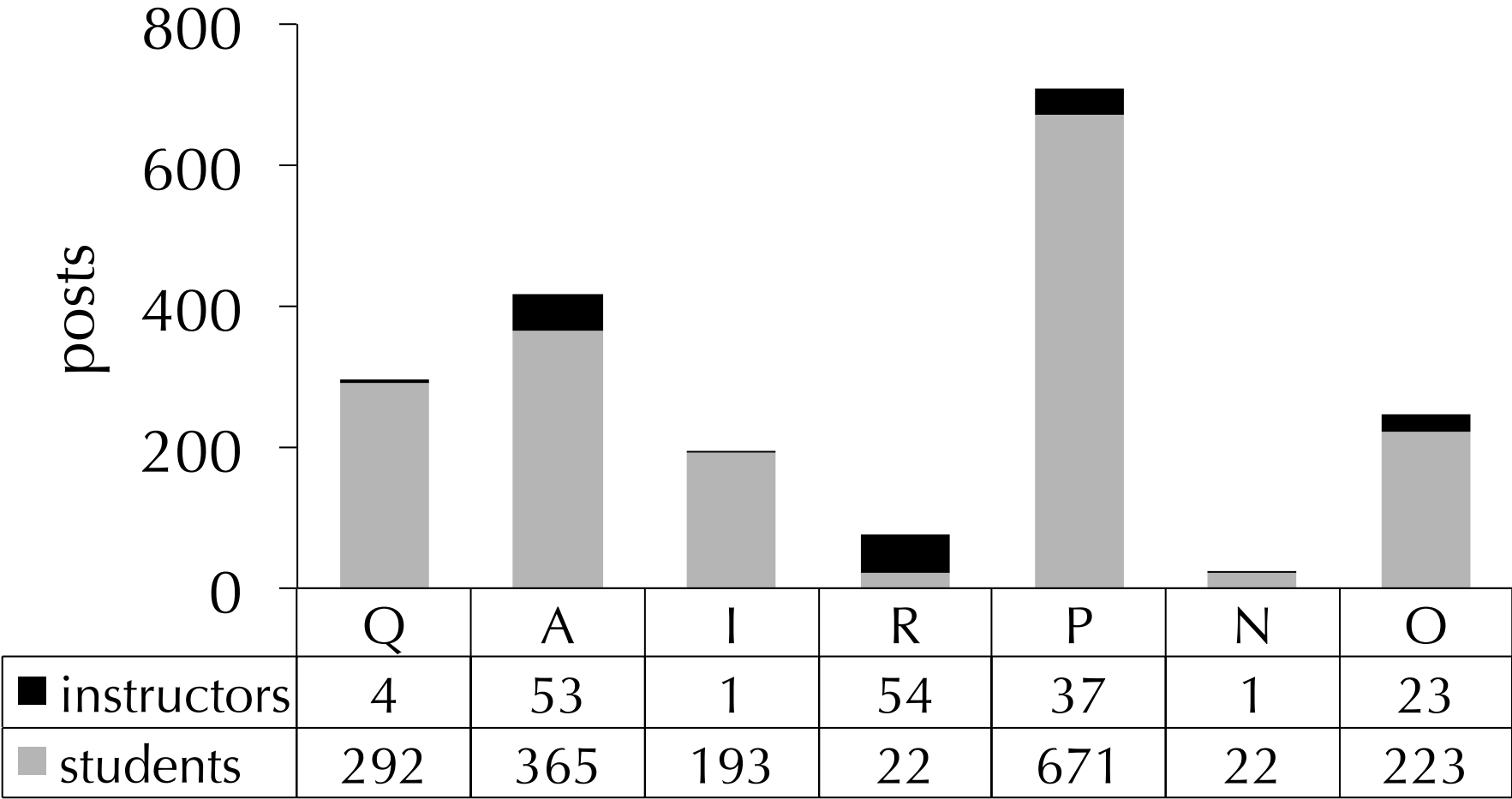
RQ1: Annotation Results

	MTurk Workers	MV and Expert
	κ_f	κ_c
Question	0.569	0.893
Answer	0.414	0.790
Issue	0.421	0.669
Issue Resolution	0.286	0.635
Positive Ack.	0.423	0.768
Negative Ack.	0.232	0.633
Other	0.337	0.625

RQ1: Annotation Results

	MTurk Workers	MV and Expert
	κ_f	κ_c
Question	0.569	0.893
Answer	0.414	0.790
Issue	0.421	0.669
Issue Resolution	0.286	0.635
Positive Ack.	0.423	0.768
Negative Ack.	0.232	0.633
Other	0.337	0.625

RQ1: Annotation Results



RQ2: Usefulness of Speech Acts

predicting instructor interventions

Input Variables	Odds Ratio	Wald's $\chi^2(1)$	<i>p</i> -value
Question (Prev)	1.391	1.345	.246
Answer (Prev)	1.566	3.361	.067
Issue (Prev)	2.278	8.782	.003
Resolution (Prev)	0.568	0.894	.345
Positive (Prev)	1.459	3.043	.081
Negative (Prev)	3.053	3.884	.049
Other (Prev)	2.001	2.992	.084
Question (Freq)	0.990	0.006	.938
Answer (Freq)	0.911	1.138	.286
Issue (Freq)	1.406	6.497	.011
Resolution (Freq)	0.876	0.349	.555
Positive (Freq)	0.930	0.768	.381
Negative (Freq)	0.685	2.843	.092
Other (Freq)	0.803	2.362	.124
NumPosts	0.981	0.081	.776
NumPostsInstruct.	1.562	24.545	.000

$$\chi^2(16) = 145.401, p < .001, \mathcal{R}^2 = 12.7\%$$

Speech Acts

- **Question:** requests information about the course content
- **Answer:** contributes information in response to a question
- **Issue:** expresses a problem with the course management
- **Issue Resolution:** attempts to resolve a previously raised issue
- **Positive Ack:** positive sentiment about a previous post
- **Negative Ack:** negative sentiment about a previous post
- **Other:** serves a different purpose

RQ3: Predicting Speech Acts

feature categories

- **Affective:** +/- sentiment, assent, anger, sadness, anxiety
- **Cognitive:** uncertainty, causality, comparing/contrasting
- **Personal concern:** topics of a personal nature
- **Linguistic:** # words, tenses, pronouns, quantifiers, negations
- **Perceptual:** references to things perceived by the senses
- **Social:** references to humans, friends, family
- **Spoken:** fillers, disfluencies
- **Unigrams:** terms with greatest chi-square correlation with SA
- **Text similarity:** cosine similarity with other posts in the thread
- **Temporal:** time to nearest deadline

RQ3: Predicting Speech Acts

feature categories

- **Sequential:** SA prediction confidence values in previous post
- **Author:** student vs. instructor
- **Hyperlinks:** number of hyperlinks
- **Modal:** number of modal verbs (could, should, shall)
- **Position:** position of post in the thread
- **Punctuation:** punctuation marks in the post
- **Votes:** up-votes versus down-votes

RQ3: Predicting Speech Acts

evaluation methodology

- Logistic regression classifiers (one per speech act)
- Feature ablation study
- 20-fold cross-validation
- Average precision

RQ3: Predicting Speech Acts

evaluation results

	question	answer	issue	issue resolution	positive ack.	negative ack.	other
all features	0.747	0.604	0.643	0.555	0.809	0.160	0.612
-affective	0.747 (0.00%)	0.590 (-2.32%)▼	0.631 (-1.87%)	0.575 (3.60%)▲	0.809 (0.00%)	0.152 (-5.00%)	0.613 (0.16%)
-author	0.746 (-0.13%)	0.605 (0.17%)▲	0.639 (-0.62%)	0.457 (-17.66%)▼	0.807 (-0.25%)	0.147 (-8.13%)▼	0.615 (0.49%)
-cognitive	0.753 (0.80%)	0.604 (0.00%)	0.648 (0.78%)	0.580 (4.50%)	0.811 (0.25%)	0.171 (6.88%)▲	0.620 (1.31%)
-cosine	0.737 (-1.34%)	0.589 (-2.48%)	0.634 (-1.40%)	0.570 (2.70%)	0.808 (-0.12%)	0.164 (2.50%)	0.605 (-1.14%)
-linguistic	0.728 (-2.54%)▼	0.594 (-1.66%)	0.644 (0.16%)	0.523 (-5.77%)	0.804 (-0.62%)▼	0.147 (-8.13%)	0.603 (-1.47%)
-links	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.612 (0.00%)
-modal	0.745 (-0.27%)	0.605 (0.17%)	0.640 (-0.47%)	0.575 (3.60%)▲	0.810 (0.12%)	0.160 (0.00%)	0.613 (0.16%)
-perceptual	0.746 (-0.13%)	0.605 (0.17%)	0.638 (-0.78%)	0.562 (1.26%)	0.810 (0.12%)	0.159 (-0.63%)	0.611 (-0.16%)
-personal_concerns	0.746 (-0.13%)	0.607 (0.50%)	0.640 (-0.47%)	0.550 (-0.90%)	0.810 (0.12%)	0.162 (1.25%)	0.609 (-0.49%)
-temporal	0.748 (0.13%)	0.605 (0.17%)	0.643 (0.00%)	0.592 (6.67%)▲	0.811 (0.25%)	0.166 (3.75%)	0.612 (0.00%)
-position	0.737 (-1.34%)▼	0.604 (0.00%)	0.645 (0.31%)	0.563 (1.44%)	0.811 (0.25%)	0.154 (-3.75%)	0.614 (0.33%)
-punctuation	0.561 (-24.90%)▼	0.584 (-3.31%)▼	0.643 (0.00%)	0.555 (0.00%)	0.804 (-0.62%)▼	0.166 (3.75%)	0.613 (0.16%)
-sentiment	0.746 (-0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.572 (3.06%)	0.806 (-0.37%)▼	0.161 (0.63%)	0.616 (0.65%)
-social	0.748 (0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.569 (2.52%)	0.809 (0.00%)	0.159 (-0.63%)	0.614 (0.33%)
-spoken	0.745 (-0.27%)	0.605 (0.17%)▲	0.640 (-0.47%)	0.563 (1.44%)	0.811 (0.25%)	0.157 (-1.88%)	0.615 (0.49%)
-unigram	0.685 (-8.30%)▼	0.547 (-9.44%)▼	0.460 (-28.46%)▼	0.426 (-23.24%)▼	0.749 (-7.42%)▼	0.090 (-43.75%)▼	0.506 (-17.32%)▼
-votes	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.611 (-0.16%)
+sequential corr.	0.764 (2.28%)	0.654 (8.28%)▲	0.657 (2.18%)	0.596 (7.39%)	0.816 (0.87%)	0.153 (-4.38%)	0.664 (8.50%)▲

RQ3: Predicting Speech Acts

evaluation results

	question	answer	issue	issue resolution	positive ack.	negative ack.	other
all features	0.747	0.604	0.643	0.555	0.809	0.160	0.612
-affective	0.747 (0.00%)	0.590 (-2.32%)▼	0.631 (-1.87%)	0.575 (3.60%)▲	0.809 (0.00%)	0.152 (-5.00%)	0.613 (0.16%)
-author	0.746 (-0.13%)	0.605 (0.17%)▲	0.639 (-0.62%)	0.457 (-17.66%)▼	0.807 (-0.25%)	0.147 (-8.13%)▼	0.615 (0.49%)
-cognitive	0.753 (0.80%)	0.604 (0.00%)	0.648 (0.78%)	0.580 (4.50%)	0.811 (0.25%)	0.171 (6.88%)▲	0.620 (1.31%)
-cosine	0.737 (-1.34%)	0.589 (-2.48%)	0.634 (-1.40%)	0.570 (2.70%)	0.808 (-0.12%)	0.164 (2.50%)	0.605 (-1.14%)
-linguistic	0.728 (-2.54%)▼	0.594 (-1.66%)	0.644 (0.16%)	0.523 (-5.77%)	0.804 (-0.62%)▼	0.147 (-8.13%)	0.603 (-1.47%)
-links	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.612 (0.00%)
-modal	0.745 (-0.27%)	0.605 (0.17%)	0.640 (-0.47%)	0.575 (3.60%)▲	0.810 (0.12%)	0.160 (0.00%)	0.613 (0.16%)
-perceptual	0.746 (-0.13%)	0.605 (0.17%)	0.638 (-0.78%)	0.562 (1.26%)	0.810 (0.12%)	0.159 (-0.63%)	0.611 (-0.16%)
-personal_concerns	0.746 (-0.13%)	0.607 (0.50%)	0.640 (-0.47%)	0.550 (-0.90%)	0.810 (0.12%)	0.162 (1.25%)	0.609 (-0.49%)
-temporal	0.748 (0.13%)	0.605 (0.17%)	0.643 (0.00%)	0.592 (6.67%)▲	0.811 (0.25%)	0.166 (3.75%)	0.612 (0.00%)
-position	0.737 (-1.34%)▼	0.604 (0.00%)	0.645 (0.31%)	0.563 (1.44%)	0.811 (0.25%)	0.154 (-3.75%)	0.614 (0.33%)
-punctuation	0.561 (-24.90%)▼	0.584 (-3.31%)▼	0.643 (0.00%)	0.555 (0.00%)	0.804 (-0.62%)▼	0.166 (3.75%)	0.613 (0.16%)
-sentiment	0.746 (-0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.572 (3.06%)	0.806 (-0.37%)▼	0.161 (0.63%)	0.616 (0.65%)
-social	0.748 (0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.569 (2.52%)	0.809 (0.00%)	0.159 (-0.63%)	0.614 (0.33%)
-spoken	0.745 (-0.27%)	0.605 (0.17%)▲	0.640 (-0.47%)	0.563 (1.44%)	0.811 (0.25%)	0.157 (-1.88%)	0.615 (0.49%)
-unigram	0.685 (-8.30%)▼	0.547 (-9.44%)▼	0.460 (-28.46%)▼	0.426 (-23.24%)▼	0.749 (-7.42%)▼	0.090 (-43.75%)▼	0.506 (-17.32%)▼
-votes	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.611 (-0.16%)
+sequential corr.	0.764 (2.28%)	0.654 (8.28%)▲	0.657 (2.18%)	0.596 (7.39%)	0.816 (0.87%)	0.153 (-4.38%)	0.664 (8.50%)▲

RQ3: Predicting Speech Acts

evaluation results

	question	answer	issue	issue resolution	positive ack.	negative ack.	other
all features	0.747	0.604	0.643	0.555	0.809	0.160	0.612
-affective	0.747 (0.00%)	0.590 (-2.32%)▼	0.631 (-1.87%)	0.575 (3.60%)▲	0.809 (0.00%)	0.152 (-5.00%)	0.613 (0.16%)
-author	0.746 (-0.13%)	0.605 (0.17%)▲	0.639 (-0.62%)	0.457 (-17.66%)▼	0.807 (-0.25%)	0.147 (-8.13%)▼	0.615 (0.49%)
-cognitive	0.753 (0.80%)	0.604 (0.00%)	0.648 (0.78%)	0.580 (4.50%)	0.811 (0.25%)	0.171 (6.88%)▲	0.620 (1.31%)
-cosine	0.737 (-1.34%)	0.589 (-2.48%)	0.634 (-1.40%)	0.570 (2.70%)	0.808 (-0.12%)	0.164 (2.50%)	0.605 (-1.14%)
-linguistic	0.728 (-2.54%)▼	0.594 (-1.66%)	0.644 (0.16%)	0.523 (-5.77%)	0.804 (-0.62%)▼	0.147 (-8.13%)	0.603 (-1.47%)
-links	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.612 (0.00%)
-modal	0.745 (-0.27%)	0.605 (0.17%)	0.640 (-0.47%)	0.575 (3.60%)▲	0.810 (0.12%)	0.160 (0.00%)	0.613 (0.16%)
-perceptual	0.746 (-0.13%)	0.605 (0.17%)	0.638 (-0.78%)	0.562 (1.26%)	0.810 (0.12%)	0.159 (-0.63%)	0.611 (-0.16%)
-personal_concerns	0.746 (-0.13%)	0.607 (0.50%)	0.640 (-0.47%)	0.550 (-0.90%)	0.810 (0.12%)	0.162 (1.25%)	0.609 (-0.49%)
-temporal	0.748 (0.13%)	0.605 (0.17%)	0.643 (0.00%)	0.592 (6.67%)▲	0.811 (0.25%)	0.166 (3.75%)	0.612 (0.00%)
-position	0.737 (-1.34%)▼	0.604 (0.00%)	0.645 (0.31%)	0.563 (1.44%)	0.811 (0.25%)	0.154 (-3.75%)	0.614 (0.33%)
-punctuation	0.561 (-24.90%)▼	0.584 (-3.31%)▼	0.643 (0.00%)	0.555 (0.00%)	0.804 (-0.62%)▼	0.166 (3.75%)	0.613 (0.16%)
-sentiment	0.746 (-0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.572 (3.06%)	0.806 (-0.37%)▼	0.161 (0.63%)	0.616 (0.65%)
-social	0.748 (0.13%)	0.604 (0.00%)	0.641 (-0.31%)	0.569 (2.52%)	0.809 (0.00%)	0.159 (-0.63%)	0.614 (0.33%)
-spoken	0.745 (-0.27%)	0.605 (0.17%)▲	0.640 (-0.47%)	0.563 (1.44%)	0.811 (0.25%)	0.157 (-1.88%)	0.615 (0.49%)
-unigram	0.685 (-8.30%)▼	0.547 (-9.44%)▼	0.460 (-28.46%)▼	0.426 (-23.24%)▼	0.749 (-7.42%)▼	0.090 (-43.75%)▼	0.506 (-17.32%)▼
-votes	0.746 (-0.13%)	0.604 (0.00%)	0.643 (0.00%)	0.560 (0.90%)	0.809 (0.00%)	0.161 (0.63%)	0.611 (-0.16%)
+sequential corr.	0.764 (2.28%)	0.654 (8.28%)▲	0.657 (2.18%)	0.596 (7.39%)	0.816 (0.87%)	0.153 (-4.38%)	0.664 (8.50%)▲

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- question words (how, what), modal verbs (can, would), words about confusion or gaps in knowledge (thought, wonder)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- course-related concepts (meta, tag), explanation (example, depend), terms indicating links to external content (http, href)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- terms about problems (issue, mistake), infrastructure (browser, chrome), course materials (answer, question, homework, quiz)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- terms about apologizing (apology, sorry), fixing problems (resolve, update), explaining errors (apparently, cause) and grading (credit, extra)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- terms indicating positive sentiment (agree, great, thank)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- terms indicating negative sentiment (disagree, disappoint, hate)

RQ3: Predicting Speech Acts

term stems with highest co-occurrence with target speech act

question	answer	issue	issue_resolution	positive_ack	negative_ack	other
anyon (96.6)	thank (171.8)	strong (350.8)	apolog (161.4)	thank (465.6)	disagre (35.2)	que (300.3)
doe (72.5)	exampl (83.9)	typo (324.6)	resolv (161.2)	cathi (110.1)	frustrat (35.0)	para (230.6)
what (61.9)	href (77.0)	factual (299.7)	fix (104.8)	agre (88.8)	anonym (32.5)	por (163.9)
how (38.5)	depend (71.9)	browser (293.1)	caus (100.2)	http (80.0)	less (27.1)	gracia (143.8)
anybodi (36.3)	describ (68.1)	omit (290.3)	sorri (76.8)	href (69.0)	code (25.8)	con (126.4)
ani (35.5)	http (65.3)	error (266.4)	now (67.3)	great (65.4)	necessarili (24.5)	pero (125.2)
can (32.3)	target (61.6)	problem (261.3)	remov (59.5)	target (58.0)	exercis (23.0)	todo (125.2)
question (29.3)	object (60.3)	detail (253.5)	001 (51.8)	org (46.7)	date (22.2)	curso (125.2)
why (24.2)	meta (58.4)	addit (236.9)	updat (49.3)	can (43.3)	white (21.7)	del (114.4)
would (20.4)	tag (57.1)	window (221.7)	homework (45.9)	not (43.0)	hate (21.6)	the (107.1)
take (20.3)	scheme (55.8)	chrome (210.9)	been (34.6)	pomerantz (42.8)	not (20.0)	como (105.8)
recip (15.9)	org (55.2)	mistak (206.0)	encount (28.5)	which (40.3)	turn (19.3)	mucha (104.8)
note (15.9)	content (53.8)	firefox (178.0)	score (27.5)	page (40.1)	consider (19.3)	est (85.7)
where (15.3)	mean (51.9)	homework (155.3)	appar (26.7)	question (39.3)	disappoint (19.3)	reput (67.6)
wonder (15.2)	element (51.7)	issu (136.6)	issu (25.4)	much (38.1)	accept (18.9)	thank (61.6)
email (14.9)	defin (51.4)	quiz (133.3)	sincer (23.7)	the (37.9)	get (18.5)	that (53.0)
know (14.8)	not (50.0)	messag (116.5)	instead (23.5)	problem (35.6)	challeng (16.5)	com (46.3)
thought (14.3)	format (48.5)	type (111.0)	credit (21.9)	name (35.5)	simpli (16.5)	you (39.0)
regard (14.0)	that (48.0)	question (100.8)	player (21.3)	titl (34.3)	express (16.5)	grei (37.9)
good (12.8)	name (47.9)	answer (99.0)	extra (14.1)	www (33.6)	should (15.8)	http (36.6)

- non-English words

Conclusions

- **RQ1:** using our speech act definitions, crowdsourced labels can be combined to approximate those of an expert.
 - ▶ SA labels can be collected reliably and inexpensively.
- **RQ2:** speech acts are helpful for predicting instructor interventions, but not assignment completion/performance.
 - ▶ SAs may be useful for alerting instructors, but additional factors are needed to predict student performance
- **RQ3:** different features are helpful for different speech acts
 - ▶ unigram features (which capture different phenomena) and sequential correlation features were consistently predictive

Thank you!
Questions?

