ACL2020 Program Co-Chairs Report

Joyce Chai, Natalie Schluter and Joel Tetreault
July 21, 2020

1. Introduction

ACL2020 was a conference of records: a record number of submissions (3,429), a record number of accepted papers (778), a record number of attendees (4,972), and finally, the first ACL in its 58 years of going virtual. We focus this Program Chair report on the innovations introduced in the conference as well as reflect upon the decisions we made with an eye to support organizers of future conferences. This report has the following outline:

<outline>

Further information about ACL 2020 can be found here:

- Main Conference website: https://acl2020.org/
- ACL2020 blog information: https://acl2020.org/blog/
- Virtual Conference website: https://virtual.acl2020.org/
- ACL2020 Conference Anthology: https://www.aclweb.org/anthology/events/acl-2020/
- Proceedings Frontmatter (includes introductions, chairs, reviewers, etc.): https://www.aclweb.org/anthology/2020.acl-main.0.pdf

Reviewing Pipeline

Other conferences had this….

2. Reviewing Process

2.1 Four New Tracks

ACL2020 introduced four new tracks:(1) ***Ethics and NLP***. Ethical issues have become increasingly important as more advanced tools become available for NLP research and development. We dedicated a new track and explicitly invite contributions that study ethical

issues and impact regarding NLP research and applications. (2) ***Interpretation and Analysis of Models for NLP****.* As the community strives for pushing performance boundaries, understanding behaviors of STOA models becomes critical. (3) ***Theory and Formalism***. This track is designed to encourage submissions targeted to theoretical underpinning of NLP models which had little/small presence in the past ACL conferences. (4) ***Theme: Taking Stock of Where We've Been and Where We're Going****.* The last few years have witnessed an unprecedented growth in NLP since the field began over sixty years ago. This track is designed to invite submissions that can provide insight for the community to assess how much we have accomplished today with respect to the past and where the field should be heading to. As the theme track is different from other tracks, we used a slightly different review form.

2.2 Earlier Submission Deadline and Notification

To accommodate a more realistic workflow, given (1) the rapid growth in the number of submissions to ACL conferences, (2) together with avoiding the period for authors from Dec. 15-Jan. 15 while giving us more time to implement and test new implementations, we moved the submission deadline back to December 9. Specifically, previous PCs advised us to do this to set a precedent for future PCs, in accommodating a more realistic timeline. The timeline is still packed, but workable and gave us flexibility in case major problems arose. While this meant an earlier-than-normal deadline for paper writers, we feel that this decision was the right one, especially with the high number of submissions and changes in the world.

2.3 SAC, AC, and Reviewer Recruitment

ACL2020 saw a record number of submissions and we expect this number to grow for subsequent NLP conferences. While this is exciting for the field, it puts additional pressure on finding enough experienced reviewers. We looked for one Senior Area Chair (SAC) per track, with the option of two if the original SAC felt it was warranted. Given the size of some of the tracks, having two SACs turned out to be essential. We recommend future conferences make two SACs mandatory for all tracks, as was the case in ACL 2019.

To estimate the number of Area Chairs (ACs) and reviewers needed, we used the submission numbers per track from EMNLP19 multiplied by a constant of 1.25. We set a ceiling of 15 papers per AC, and 6 papers per reviewer, and used those thresholds to estimate the number of ACs and reviewers per track. The SACs were in charge of selecting their team of ACs. We feel these estimates were generally accurate and additionally motivated our team to recruit more people.

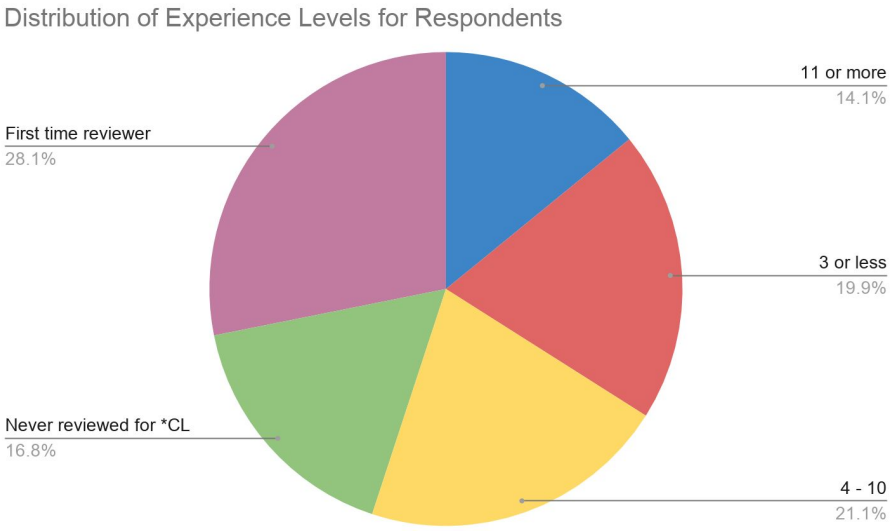
We used a multi-pronged approach to recruit more reviewers. Our solution rested on developing an author/reviewer form where all potential reviewers have to fill in information ranging from their experience level, language proficiency, research interests, etc. The approaches were:

1) Collect reviewer lists from several previous conferences as well.

2) Posted online asking for reviewers
3) SACs also recruited reviewers
4) Made it mandatory for **all** authors of a submission to sign up to review unless there were extenuating circumstances

Overall this netted nearly 10,000 submissions to the form! While the form required substantial manual and community effort, it was extremely useful throughout the conference process. We also asked reviewers if this information could be shared with other conferences, thus providing an additional resource to the ACL community. So we view this process as an overall positive.

As a side note, 50% of our reviewer pool has reviewed at 3 or fewer NLP conferences showing how many new researchers have joined recently. For ACL 2020, we draw primarily from the pool of reviewers that have reviewed for 4 or more conferences.



2.4 Initial submission reviews and desk rejects

All papers were carefully inspected to check for violations of ACL policies (ranging from formatting to anonymization to use of supplementary material). Similar to ACL2019, we used assistants to speed up an otherwise long process. All issues identified by assistants were cross-examined by two PCs. We noticed that many papers did not strictly follow the ACL style sheet. We have thus been lenient in terms of margin, line numbers, fonts, and other formatting issues. As a result 29 submissions were desk rejected for violating ACL policies on anonymity, page length, double blind review, etc.

2.5 Manual adjustment of submission tracks

Many papers were not submitted to the right track where they could receive reviews from most relevant reviewers. SACs were instructed to flag the papers that should be moved to a different track. We went through every single suggestion and moved papers around if warranted. This turned out to be a major effort. In total, 500-600 papers were moved across tracks as a result. Our recommendation is that authors and reviewers should be given clear descriptions of each track.

A more radical suggestion would be to remove tracks altogether! During the course of the review process we noted two things: first, some tracks are getting as large as conference with upwards of 300 submissions thus making it hard for SACs to know the backgrounds of everyone in their reviewing pool. Second, some papers can easily fit multiple tracks making placement and thus reviewer assignment more difficult. Thus, one change would be to do away with tracks completely and assign reviewers from the entire conference pool to each paper (instead of reviewers within the track) and automatically assign area chairs and senior area chairs.

2.6 Conflicts of Interest

Given the large number of papers and a tight reviewing timeline, we piloted a new COI detection system created by Arya McCarthy and Amanda Stent. It takes as input information from the Softconf Global Profile and Semantic Scholar of reviewers and authors and produces a list of reviewers who have a COI with each paper. This tool worked well and allowed us to do away with the bidding phase, and thus save time in the overall reviewing process. However, this process relies heavily on the community properly filling out their information. While the vast majority did do this correctly there were some gaps.

Going forward we believe that there should be a collective effort by the community to regularly update their profiles. In addition, extending the algorithm to ingest other sources will be helpful.

2.7 Reviewer Assignment

Next, we want to ensure that each submission is matched with the best reviewers possible. As in COI, we piloted a new algorithm created by Graham Neubig to serve as a first pass or pre-assignment of reviewers to papers. The Senior Area Chairs would then modify as necessary. In addition to this automatic step of Neubig's algorithm, we manually filtered potential reviewers based on their experience level from the author/reviewer form, as well as moved reviewers to tracks to better load balance across areas.

As the automatic reviewer assignment is not perfect, SACs did much manual work adjusting AC assignments as well as reviewer assignments. This effort varied among tracks. Given the current setup in Softconf, ACs' roles are fairly limited. ACs are essentially meta-reviewers who do not have access to the reviewer accounts, and therefore, cannot add reviewers, nor

make reviewer assignments, nor contact reviewers directly. We have given this feedback to Softconf and hopefully the system will be updated to support extended AC roles for future conferences.

2.8 Author Rebuttal

We also instituted the author rebuttal. While this adds an extra phase to the review process as well as more work for the program committee, we found that this information was extremely useful in the decision process. Area Chairs, reviewers and authors also reported positively on its inclusion. Our recommendation is to use the author rebuttal if time permits.

2.9 Ethical Issues

As ethical concerns and societal impacts are an important consideration for NLP research, we have explicitly asked reviewers to evaluate ethical implications of each submission. On the review form, we ask reviewers whether there are any ethical concerns about a submission that the area chairs and program chairs should be aware of. We also encourage reviewers to flag such concerns to the authors. A total of 187 reviews (out of 10111 reviews) have raised potential issues with ethical concerns. One paper, although rated highly based on its technical content, was not accepted into the program because of the potential ethical implications. This year we have asked reviewers to pay attention to ethical issues, we feel that some mechanism that would allow authors to explicitly discuss ethical implications of their work, e.g., as required by NeurIPS 2020, can better address ethical issues in the future.

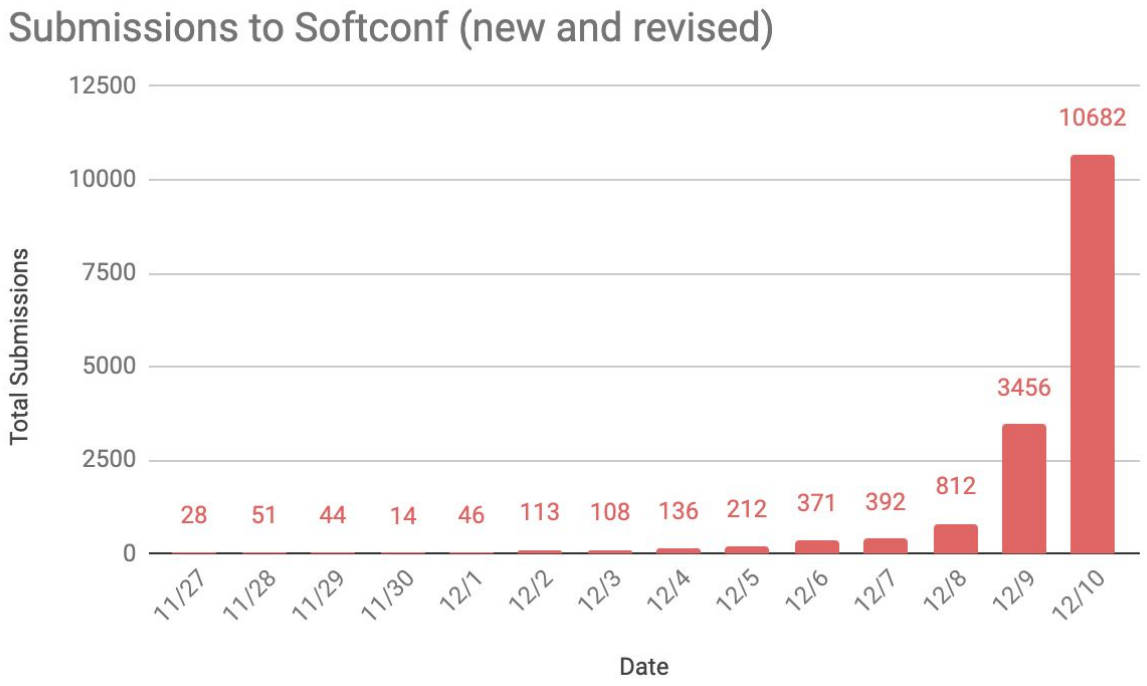
2.10 Other Observations

One major pain point for organizing ACL 2020 was the inability to send batch emails to thousands of recipients *not via Softconf.* This has recently been remedied by Nitin Madnani (ACL Information Technology Director).

ACL should perhaps have 4 PCs if the upward submission trend continues

Make sure to have emergency reviewers. On the author/reviewer form we asked if one would be available as an emergency reviewer. Having this in hand allowed the SACs to quickly contact eager reviewers to make sure all papers had at least three reviews.

As one might expect, a disproportionate number of submissions and revisions come in the 48 hours prior to the submission deadline. Future conference organizers should be aware of the load that this has on a conference management system and of course the last-minute questions many authors will have during this time. The chart below illustrates the number of submissions to Softconf in the two weeks leading up to the ACL 2020 deadline.



Finally, we feel that as the state of scientific publishing has changed quickly over the last few years, we feel that the ACL guidelines on anonymity should be reviewed as there are many corner cases to the policy that conference organizers must navigate. As a side statistic, we found that 14.4% of submissions had posted online ahead of the one-month anonymity date. 63% of that 14.4% posted on arxiv.

Because of several new initiatives implemented this year, extensive efforts have been made to communicate these changes to SACs, ACs, reviewers, as well as authors. Besides direct emails, we have used blog postings as well as twitters as our additional communication channels assisted by the publicity chair and the web chairs.

3. Reviewer Mentoring

3.1 Motivation

Given the rapid growth of NLP in terms of number of papers and new students, it is very important for our community to mentor and train our new reviewers. ACL2020 has launched a pilot program which calls for each AC to mentor at least one novice reviewer. Ultimately, the goal is to provide long-needed mentoring to new reviewers. At the very least, this process will inform ACL on constructing a reviewer mentoring program that is more scalable in the future.

Close to 300 area chairs and mentees participated in our mentoring program. This program had several phases.

- SACs and/or ACs selected potentially mentees from a large pool of reviewers based on their self-claimed experience reviewing for ACL (i.e., mostly first time reviewer for ACL).
- SAC/AC assigned mentees one paper each to review.
- After the reviews were due, ACs were asked to give an initial assessment on the quality of the reviews from their mentees.
- During the discussion period, ACs worked with their mentees who needed mentoring based on the initial assessment.
- After the final reviews were due, ACs were asked to give a follow-up assessment on the mentoring outcome.
- Mentees were also asked to fill out an exit survey to provide feedback to this program.

3.2 Mentees reviewing experience

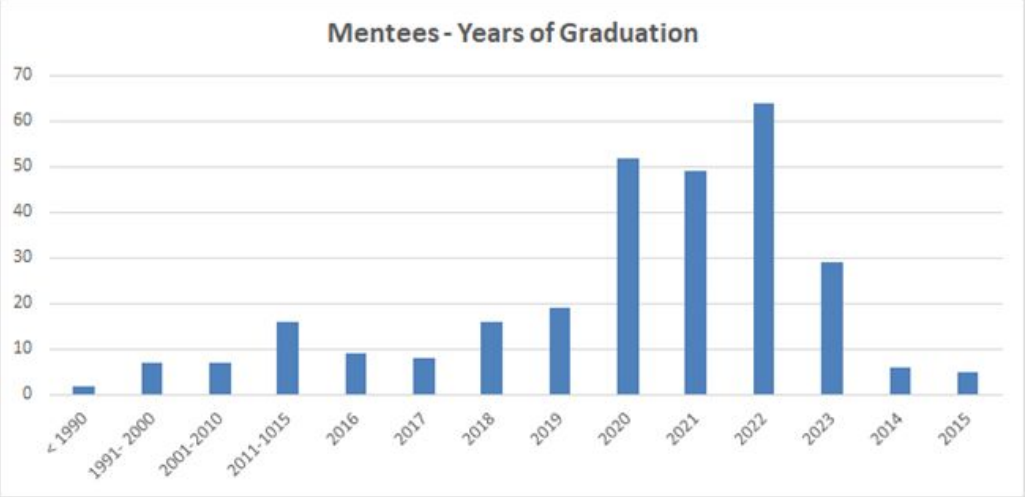


Figure 1: Distribution of Mentees' reviewing experience

Most of our mentees are self claimed first time reviewers for ACL. A number of them graduated a while ago. These are the experienced researchers in different disciplines and happened to review for ACL for the first time. We also have mentees who have just started their PhD study and will graduate in a few years down the road. A majority of mentees are PhD students in their mid to late stage of their graduate study.

3.3 How was the review quality?

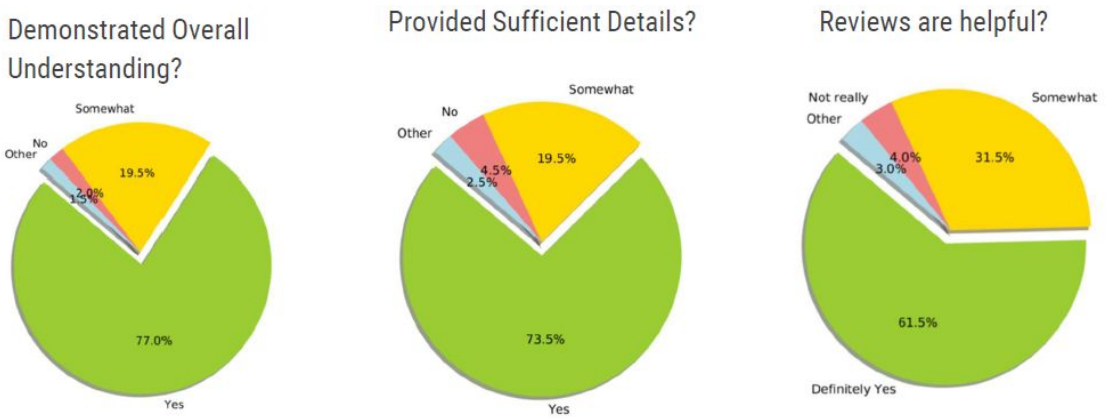


Figure 2: Responses to an initial assessment survey

After the reviews were due, we asked ACs to evaluate the reviews from their mentees and fill out an initial assessment survey. Over 200 ACs provided feedback. For questions like whether the reviews have demonstrated an overall understanding of the paper, whether the reviews have provided sufficient details, and whether the reviews are helpful in helping area chairs in their decision making process, the responses are quite positive (as shown in Figure 2).

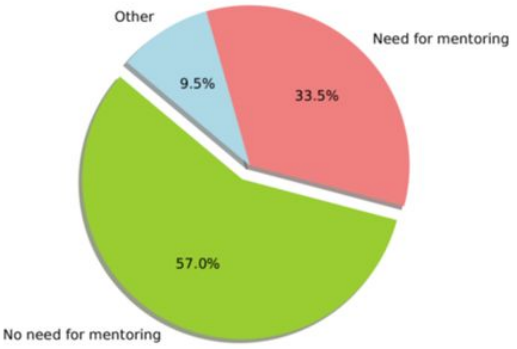


Figure 3: The need for mentoring

We also asked ACs whether their mentees needed mentoring based on their reviews. As shown in Figure 3, 57% of ACs responded that there was no need for any further mentoring. Based on this feedback, we asked the 43% of ACs (around 90 ACs) who indicated their mentees would benefit from some mentoring to continue with the mentoring activity. In particular, we asked these ACs to contact their mentees directly and provide specific comments/feedback to help their mentees to improve their reviews. So this process happened during the discussion period. At the end of the period, we did a follow-up survey on the mentoring outcome and 78 ACs provided feedback.

3.4 Did mentees update their reviews after being mentored?

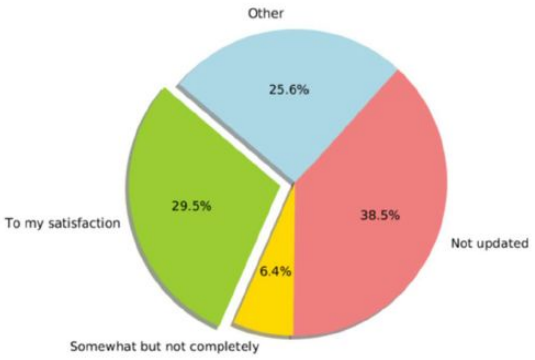


Figure 4: Review updating after mentoring

We specifically asked ACs whether their mentees have updated their reviews. As shown in Figure 4, less than 1/3 of ACs said that their mentees actually updated reviews to their satisfaction. More than 1/3 of mentees did not update their reviews after being engaged with their mentors. This outcome indicates that a better mechanism needs to be implemented to ensure a productive follow up.

3.5 Is the mentoring effort worthwhile?

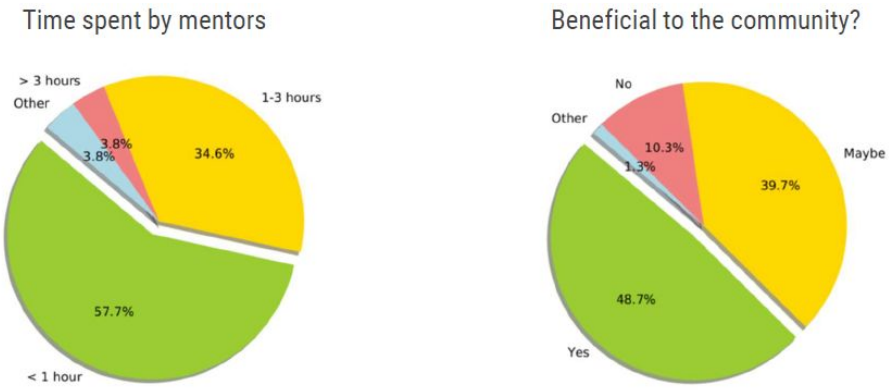


Figure 5: Perspectives from ACs on whether one-to-one mentoring is worthwhile

Mentoring takes time. Figure 5 shows the overall time spent by mentors in this effort. When asked whether this type of one-to-one mentoring interaction is beneficial to the community. Closed to half of the ACs (who responded to our second survey) said yes. About 10% of them said no and they thought there would be other mechanisms that would work out better than this one-to-one interaction.

We also asked mentees whether they thought this type of mentoring is helpful. As shown in Figure 6, for those mentees who actively engaged with their mentors and also responded to our survey, 3/4 of them said that this experience was rewarding.

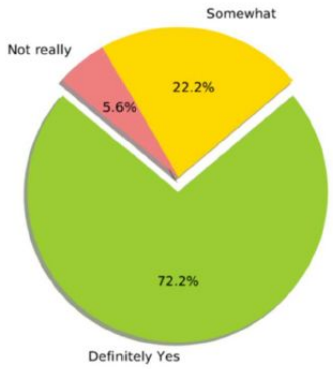


Figure 6: Mentees' responses to whether the mentoring is helpful

3.7 What can be done in the future?

A few things can be improved as we move forward.

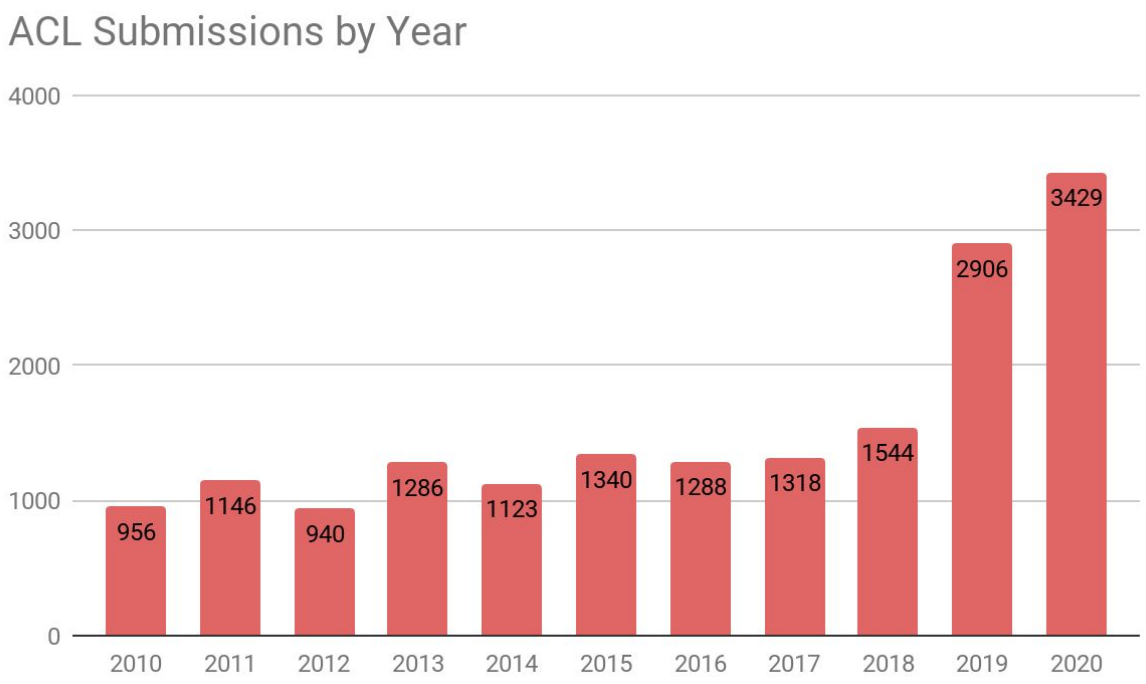
- Establish a better mentor-mentee matching system. For example, instead of selecting mentees based on their review experience, we could have junior reviewers enroll themselves in the mentoring program. This would lead to more motivated mentees which will ensure more productive follow up.
- Set up guidelines to help mentors engage in mentoring activities, e.g., what needs to be done, what is expected from the mentees.
- Establish a dedicated role of "mentoring chair" to organize and coordinate mentoring efforts.
- Improve infrastructure and communication channels. Most of communication related to the mentoring program was done using google doc, spreadsheet, and email, which made it cumbersome to manage and keep track of progress. An improved infrastructure (e.g., as part of softconf system) will help.
- Build a database of papers and examples and create tutorials with examples that can be helpful for new reviewers. This can also provide tools for mentors when they engage with mentees in one-to-one interaction.

4. Conference Statistics

4.1 General Statistics

ACL2020 had an acceptance rate of 22.7% based on 3,429 submissions and 778 accepted papers. 3,429 is a record number of submissions for ACL. To put that number in

perspective, and to show fast the field has grown, two years ago, the number of submissions was 1,544, or less than half. Ten years ago, there were “only” 956 submissions.

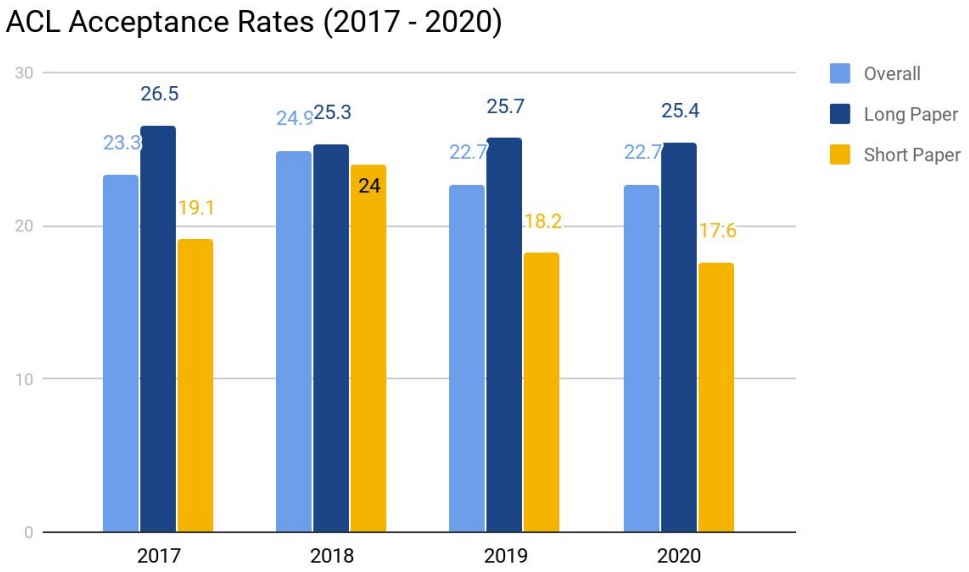


Of the 778 accepted papers, 570 were long papers and 208 were short papers. Note that when Desk Rejects and Withdrawals are removed, 29 and 312 papers respectively, the acceptance rate is 25.2%.

	Total Submissions	**Accepted**	**% Accepted**
Total	3429	779	22.7%
Long	2244	570	25.4%
Short	1185	208	17.6%

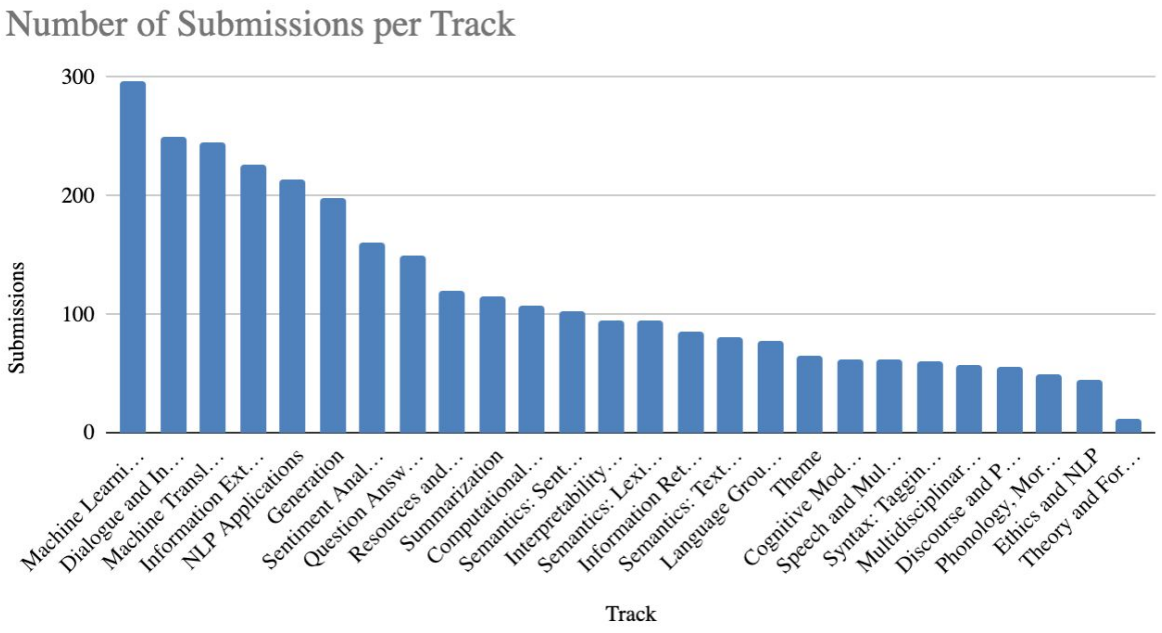
4.2 Comparison with Prior ACLs

This year’s ACL is in line with acceptance rates with the most recent editions of the ACL. ACL 2019 also had an acceptance rate of 22.7% with similar rates for long and short papers. For a history of acceptance rates at the ACL, please visit the ACL Wiki.



4.3 Track Statistics

Next, we break up the acceptance rate by track. Machine Learning for NLP, Dialogue and Interactive Technologies, Machine Translation, Information Extraction and NLP Applications were the top five most popular tracks, with each having over 200 submissions. Machine Learning for NLP had nearly 300 submissions (296). To show how much our field has grown, ACL 2002 received 258 submissions *total* across all tracks.



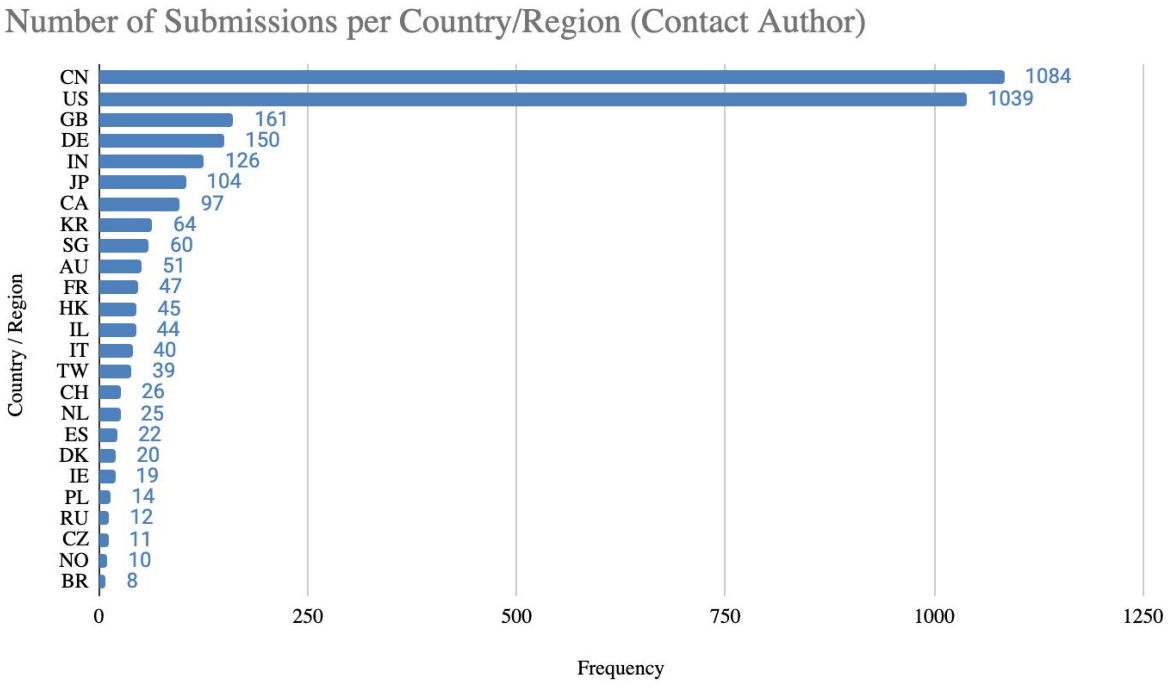
Acceptance rates for each track ranged from 17.9% to 41.7% as follows.

Track	Submissions	Accepted	% Accepted
Cognitive Modeling and Psycholinguistics	62	13	21
Computational Social Science and Social Media	108	23	21.3
Desk Reject or Withdrawn	341	0	0
Dialogue and Interactive Systems	250	62	24.8
Discourse and Pragmatics	56	10	17.9
Ethics and NLP	44	13	29.5
Generation	198	49	24.7
Information Extraction	227	52	22.9
Information Retrieval and Text Mining	86	20	23.3
Interpretability and Analysis of Models for NLP	95	29	30.5
Language Grounding to Vision, Robotics and Beyond	77	22	28.6
Machine Learning for NLP	296	67	22.6
Machine Translation	245	68	27.8
Multidisciplinary and Area Chair COI	58	19	37.9
NLP Applications	213	48	22.5
Phonology, Morphology and Word Segmentation	49	15	30.6
Question Answering	150	33	22
Resources and Evaluation	120	42	35
Semantics: Lexical	95	17	17.9
Semantics: Sentence Level	103	29	28.2
Semantics: Textual Inference and Other Areas of Semantics	81	24	29.6

Sentiment Analysis, Stylistic Analysis, and Argument Mining	161	33	20.5
Speech and Multimodality	62	16	25.8
Summarization	115	30	26.1
Syntax: Tagging, Chunking and Parsing	60	16	26.7
Theme	65	24	36.9
Theory and Formalism in NLP (Linguistic and Mathematical)	12	5	41.7
Total	**3429**	**779**	**22.7**

4.4 Country/Region Statistics

We follow ACL 2019 in extracting the country/region listed in Softconf by the contact author and calculating statistics on that set. Please note that the country/region data is self-reported by each author. There are 57 countries/regions represented in this subset. Below we list the 25 countries/regions with the most submissions. China led with 1,084, followed closely by the United States. In ACL 2019, the United States had 820 submissions and China had 817. Germany had the third most submissions then with 136.



Finally, we analyze the acceptance rates for each country/region, again based on contact author. The following table lists all 57 countries/regions in alphabetical order. In the end, 37 countries/regions have papers in the conference. The five with the most accepted papers were the US (305), China (185), Great Britain (50), Germany (44), and Japan (24). As observed by the ACL2019 Program Chairs, the distribution is too skewed for a fair comparison of acceptance rates. As in their analysis, if we consider the top 15 countries/regions in terms of number of submissions, those with the highest acceptance rates were Israel (40.9%), Great Britain (31.1%), United States (29.4%), Germany (29.3%), and a tie between Hong Kong and Singapore, both with 26.7%. This list is almost the same as last year, except Great Britain was not in the top five.

Country / Region	Code	Count	Accepted	% Accepted
Argentina	AR	1	0	0
Australia	AU	51	12	23.5
Austria	AT	4	0	0
Bangladesh	BD	4	0	0
Belgium	BE	6	1	16.7
Brazil	BR	8	1	12.5
Bulgaria	BG	2	1	50
Canada	CA	97	15	15.5
China	CN	1084	185	17.1
Croatia	HR	3	1	33.3

Czech Republic	CZ	11	1	9.1
Denmark	DK	20	6	30
Egypt	EG	4	0	0
Estonia	EE	1	0	0
Finland	FI	2	0	0
France	FR	47	10	21.3
Germany	DE	150	44	29.3
Great Britain	GB	161	50	31.1
Greece	GR	5	2	40
Hong Kong	HK	45	12	26.7
Hungary	HU	2	0	0
India	IN	126	15	11.9
Iran	IR	8	1	12.5
Ireland	IE	19	6	31.6
Israel	IL	44	18	40.9
Italy	IT	40	7	17.5
Japan	JP	104	24	23.1
Kazakhstan	KZ	5	0	0
Macao	MO	6	2	33.3
Mexico	MX	3	0	0
Netherlands	NL	25	5	20
New Zealand	NZ	3	1	33.3
Norway	NO	10	1	10
Pakistan	PK	2	0	0
Peru	PE	2	1	50
Poland	PL	14	1	7.1
Portugal	PT	8	1	12.5
Qatar	QA	7	1	14.3
Republic of Korea	KR	64	9	14.1
Romania	RO	6	0	0
Russian Federation	RU	12	1	8.3
Rwanda	RW	1	0	0
Saudi Arabia	SA	2	0	0
Singapore	SG	60	16	26.7
Slovakia	SK	1	0	0
Slovenia	SI	1	0	0
South Africa	ZA	2	0	0
Spain	ES	22	9	40.9
Sri Lanka	LK	2	0	0
Sweden	SE	6	2	33.3
Switzerland	CH	26	5	19.2
Taiwan	TW	39	6	15.4

Turkey	TR	4	0	0
United Arab Emirates	AE	2	1	50
United States	US	1039	305	29.4
Uruguay	UY	2	0	0
Venezuela	VN	3	0	0