Towards a Computational History of the ACL: 1980–2008

Ashton Anderson
Stanford University

Dan McFarland
Stanford University

Dan Jurafsky
Stanford University

`ashtona@stanford.edu dmcfarla@stanford.edu jurafsky@stanford.edu`

Abstract

We develop a people-centered computational history of science that tracks authors over topics and apply it to the history of computational linguistics. We present four findings in this paper. First, we identify the topical subfields authors work on by assigning automatically generated topics to each paper in the ACL Anthology from 1980 to 2008. Next, we identify four distinct research epochs where the pattern of topical overlaps are stable and different from other eras: an early NLP period from 1980 to 1988, the period of US government-sponsored MUC and ATIS evaluations from 1989 to 1994, a transitory period until 2001, and a modern integration period from 2002 onwards. Third, we analyze the flow of authors across topics to discern how some subfields flow into the next, forming different stages of ACL research. We find that the government-sponsored bakeoffs brought new researchers to the field, and bridged early topics to modern probabilistic approaches. Last, we identify steep increases in author retention during the bakeoff era and the modern era, suggesting two points at which the field became more integrated.

1 Introduction

The rise of vast on-line collections of scholarly papers has made it possible to develop a computational history of science. Methods from natural language processing and other areas of computer science can be naturally applied to study the ways a field and its ideas develop and expand (Au Yeung and Jatowt, 2011; Gerrish and Blei, 2010; Tu et al., 2010; Aris et al., 2009). One particular direction in computational

history has been the use of topic models (Blei et al., 2003) to analyze the rise and fall of research topics to study the progress of science, both in general (Griffiths and Steyvers, 2004) and more specifically in the ACL Anthology (Hall et al., 2008).

We extend this work with a more people-centered view of computational history. In this framework, we examine the trajectories of individual authors across research topics in the field of computational linguistics. By examining a single author's paper topics over time, we can trace the evolution of her academic efforts; by superimposing these individual traces over each other, we can learn how the entire field progressed over time. One goal is to investigate the use of these techniques for computational history in general. A second goal is to use the ACL Anthology Network Corpus (Radev et al., 2009) and the incorporated ACL Anthology Reference Corpus (Bird et al., 2008) to answer specific questions about the history of computational linguistics. What is the path that the ACL has taken throughout its 50-year history? What roles did various research topics play in the ACL's development? What have been the pivotal turning points?

Our method consists of four steps. We first run topic models over the corpus to classify papers into topics and identify the topics that people author in.

We then use these topics to identify epochs by correlating over time the number of persons that topics share in common. From this, we identify epochs as sustained patterns of topical overlap.

Our third step is to look at the flow of authors between topics over time to detect patterns in how authors move between areas in the different epochs. We group topics into clusters based on when authors move in and out of them, and visualize the flow

of people across these clusters to identify how one topic leads to another.

Finally, in order to understand how the field grows and declines, we examine patterns of entry and exit within each epoch, studying how author retention (the extent to which authors keep publishing in the ACL) varies across epochs.

2 Identifying Topics

Our first task is to identify research topics within computational linguistics. We use the ACL Anthology Network Corpus and the incorporated ACL Anthology Reference Corpus, with around 13,000 papers by approximately 11,000 distinct authors from 1965 to 2008. Due to data sparsity in early years, we drop all papers published prior to 1980.

We ran LDA on the corpus to produce 100 generative topics (Blei et al., 2003). Two senior researchers in the field (the third author and Chris Manning) then collaboratively assigned a label to each of the 100 topics, which included marking those topics which were non-substantive (lists of function words or affixes) to be eliminated. They produced a consensus labeling with 73 final topics, shown in Table 1 (27 non-substantive topics were eliminated, e.g. a pronoun topic, a suffix topic, etc.).

Each paper is associated with a probability distribution over the 100 original topics describing how much of the paper is generated from each topic. All of this information is represented by a matrix P, where the entry P_{ij} is simply the loading of topic j on paper i (since each row is a probability distribution, $\sum_j P_{ij} = 1$). For ease of interpretation, we sparsify the matrix P by assigning papers to topics and thus set all entries to either 0 or 1. We do this by choosing a threshold T and setting entries to 1 if they exceed this threshold. If we call the new matrix Q, $Q_{ij} = 1 \iff P_{ij} \geq T$. Throughout all our analyses we use $T = 0.1$. This value is approximately two standard deviations above $\overline{P}$, the mean of the entries in P. Most papers are assigned to 1 or 2 topics; some are assigned to none and some are assigned to more.

This assignment of papers to topics also induces an assignment of authors to topics: an author is assigned to a topic if she authored a paper assigned to that topic. Furthermore, this assignment is natu-

rally *dynamic:* since every paper is published in a particular year, authors' topic memberships change over time. This fact is at the heart of our methodology — by assigning authors to topics in this principled way, we can track the topics that authors move through. Analyzing the flow of authors through topics enables us to learn which topics beget other topics, and which topics are related to others by the people that author across them.

3 Identifying Epochs

What are the major epochs of the ACL's history? In this section, we seek to partition the years spanned by the ACL's history into clear, distinct periods of topical cohesion, which we refer to as *epochs*. If the dominant research topics people are working on suddenly change from one set of topics to another, we view this as a transition between epochs.

To identify epochs that satisfy this definition, we generate a set of matrices (one for each year) describing the number of people that author in every pair of topics during that year. For year y, let N^y be a matrix such that N_{ij}^y is the number of people that author in both topics i and j in year y (where authoring in topic j means being an author on a paper p such that $Q_{pj} = 1$). We don't normalize by the total number of people in each topic, thus proportionally representing bigger topics since they account for more research effort than smaller topics. Each matrix is a signature of which topic pairs have overlapping author sets in that year.

From these matrices, we compute a final matrix C of year-year correlations. C_{ij} is the Pearson correlation coefficient between N^i and N^j. C captures the degree to which years have similar patterns of topic authorship overlap, or the extent to which a consistent pattern of topical research is formed. We visualize C as a thermal in Figure 1.

To identify epochs in ACL's history, we ran hierarchical complete link clustering on C. This resulted in a set of four distinct epochs: 1980–1988, 1989–1994, 1995–2001, and 2002–2008. For three of these periods (all except 1995–2001), years within each of these ranges are much more similar to each other than they are to other years. During the third period (1995–2001), none of the years are highly similar to any other years. This is indicative of a

14

Number	Name	Topics
1	Big Data NLP	**Statistical Machine Translation (Phrase-Based)**: bleu, statistical, source, target, phrases, smt, reordering **Dependency Parsing**: dependency/ies, head, czech, depen, dependent, treebank **MultiLingual Resources**: languages, spanish, russian, multilingual, lan, hindi, swedish **Relation Extraction**: pattern/s, relation, extraction, instances, pairs, seed **Collocations/Compounds**: compound/s, collocation/s, adjectives, nouns, entailment, expressions, MWEs **Graph Theory + BioNLP**: graph/s, medical, edge/s, patient, clinical, vertex, text, report, disease **Sentiment Analysis**: question/s, answer/s, answering, opinion, sentiment, negative, positive, polarity
2	Probabilistic Methods	**Discriminative Sequence Models**: label/s, conditional, sequence, random, discriminative, inference **Metrics + Human Evaluation**: human, measure/s, metric/s, score/s, quality, reference, automatic, correlation, judges **Statistical Parsing**: parse/s, treebank, trees, Penn, Collins, parsers, Charniak, accuracy, WSJ **ngram Language Models**: n-gram/s, bigram/s, prediction, trigram/s, unigram/s, trigger, show, baseline **Algorithmic Efficiency**: search, length, size, space, cost, algorithms, large, complexity, pruning **Bilingual Word Alignment**: alignment/s, align/ed, pair/s, statistical, source, target, links, Brown **ReRanking**: score/s, candidate/s, list, best, correct, hypothesis, selection, rank/ranking, scoring, top, confidence **Evaluation Metrics**: precision, recall, extraction, threshold, methods, filtering, extract, high, phrases, filter, f-measure **Methods (Experimental/Evaluation)**: experiments, accuracy, experiment, average, size, 100, baseline, better, per, sets **Machine Learning Optimization**: function, value/s, parameter/s, local, weight, optimal, solution, criterion, variables
3	Linguistic Supervision	**Biomedical Named Entity Recognition**: biomedical, gene, term, protein, abstracts, extraction, biological **Word Segmentation**: segment/ation, character/s, segment/s, boundary/ies, token/ization **Document Retrieval**: document/s, retrieval, query/ies, term, relevant/ance, collection, indexing, search **SRL/Framenet**: argument/s, role/s, predicate, frame, FrameNet, predicates, labeling, PropBank **Wordnet/Multilingual Ontologies**: ontology/ies, italian, domain/s, resource/s, i.e, ontological, concepts **WebSearch + Wikipedia**: web, search, page, xml, http, engine, document, wikipedia, content, html, query, Google **Clustering + Distributional Similarity**: similar/ity, cluster/s/ing, vector/s, distance, matrix, measure, pair, cosine, LSA **Word Sense Disambiguation**: WordNet, senses, disambiguation, WSD, nouns, target, synsets, Yarowsky **Machine Learning Classification**: classification, classifier/s, examples, kernel, class, SVM, accuracy, decision **Linguistic Annotation**: annotation/s/ed, agreement, scheme/s, annotators, corpora, tools, guidelines **Tutoring Systems**: student/s, reading, course, computer, tutoring, teaching, writing, essay **Chunking/Memory Based Models**: chunk/s/ing, pos, accuracy, best, memory-based, Daelemans **Named Entity Recognition**: entity/ies, name/s/d, person, proper, recognition, location, organization, mention **Dialog**: dialogue, utterance/s, spoken, dialog/ues, act, interaction, conversation, initiative, meeting, state, agent **Summarization**: topic/s, summarization, summary/ies, document/s, news, articles, content, automatic, stories
4	Discourse	**Multimodal (Mainly Generation)**: object/s, multimodal, image, referring, visual, spatial, gesture, reference, description **Text Categorization**: category/ies, group/s, classification, texts, categorization, style, genre, author **Morphology**: morphological, arabic, morphology, forms, stem, morpheme/s, root, suffix, lexicon **Coherence Relations**: relation, rhetorical, unit/s, coherence, texts, chains **Spell Correction**: error/s, correct/ion, spelling, detection, rate **Anaphora Resolution**: resolution, pronoun, anaphora, antecedent, pronouns, coreference, anaphoric **Question Answering Dialog System**: response/s, you, expert, request, yes, users, query, question, call, database **UI/Natural Language Interface**: users, database, interface, a71, message/s, interactive, access, display **Computational Phonology**: phonological, vowel, syllable, stress, phonetic, phoneme, pronunciation **Neural Networks/Human Cognition**: network/s, memory, acquisition, neural, cognitive, units, activation, layer **Temporal IE/Aspect**: event/s, temporal, tense, aspect, past, reference, before, state **Prosody**: prosody/ic, pitch, boundary/ies, accent, cues, repairs, phrases, spoken, intonation, tone, duration
5	Early Probability	**Lexical Acquisition Of Verb Subcategorization**: class/es, verb/s, paraphrase/s, subcategorization, frames **Probability Theory**: probability/ies, distribution, probabilistic, estimate/tion, entropy, statistical, likelihood, parameters **Collocations Measures**: frequency/ies, corpora, statistical, distribution, association, statistics, mutual, co-occurrences **POS Tagging**: tag/ging, POS, tags, tagger/s, part-of-speech, tagged, accuracy, Brill, corpora, tagset **Machine Translation (Non Statistical + Bitexts)**: target, source, bilingual, translations, transfer, parallel, corpora
6	Automata	**Automata Theory**: string/s, sequence/s, left, right, transformation, match **Tree Adjoining Grammars** : trees, derivation, grammars, TAG, elementary, auxiliary, adjoining **Finite State Models (Automata)**: state/s, finite, finite-state, regular, transition, transducer **Classic Parsing**: grammars, parse, chart, context-free, edge/s, production, CFG, symbol, terminal **Syntactic Trees**: node/s, constraints, trees, path/s, root, constraint, label, arcs, graph, leaf, parent
7	Classic Linguistics	**Planning/BDI**: plan/s/ning, action/s, goal/s, agent/s, explanation, reasoning **Dictionary Lexicons**: dictionary/ies, lexicon, entry/ies, definition/s, LDOCE, **Linguistic Example Sentences**: John, Mary, man, book, examples, Bill, who, dog, boy, coordination, clause **Syntactic Theory**: grammatical, theory, functional, constituent/s, constraints, LFG **Formal Computational Semantics**: semantics, logic/al, scope, interpretation, meaning, representation, predicate **Speech Acts + BDI**: speaker, utterance, act/s, hearer, belief, proposition, focus, utterance **PP Attachment**: ambiguity/ies/ous, disambiguation, attachment, preference, preposition **Natural Language Generation**: generation/ing, generator, choice, generated, realization, content **Lexical Semantics**: meaning/s, semantics, metaphor, interpretation, object, role **Categorial Grammar/Logic**: proof, logic, definition, let, formula, theorem, every, iff, calculus **Syntax**: clause/s, head, subject, phrases, object, verbs, relative, nouns, modifier **Unification Based Grammars**: unification, constraints, structures, value, HPSG, default, head **Concept Ontologies / Knowledge Rep**: concept/s, conceptual, attribute/s, relation, base
8	Government	**MUC-Era Information Extraction**: template/s, message, slot/s, extraction, key, event, MUC, fill/s **Speech Recognition**: recognition, acoustic, error, speaker, rate, adaptation, recognizer, phone, ASR **ATIS dialog**: spoken, atis, flight, darpa, understanding, class, database, workshop, utterances
9	Early NLU	**1970s-80s NLU Work**: 1975-9, 1980-6, computer, understanding, syntax, semantics, ATN, Winograd, Schank, Wilks, lisp **Code Examples**: list/s, program/s, item/s, file/s, code/s, computer, line, output, index, field, data, format **Speech Parsing And Understanding**: frame/s, slot/s, fragment/s, parse, representation, meaning

Table 1: Results of topic clustering, showing some high-probability representative words for each cluster.
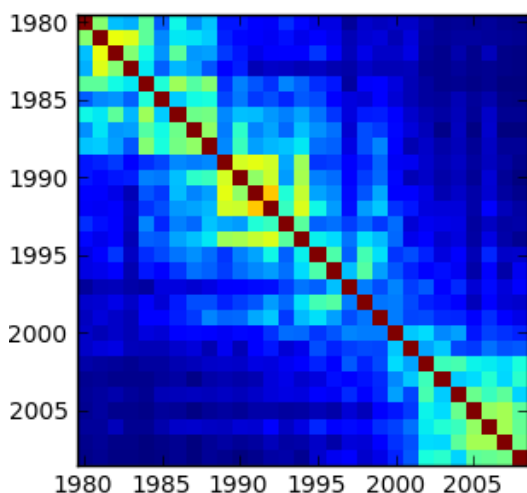
Figure 1: Year-year correlation in topic authoring patterns. Hotter colors indicate high correlation, colder colors denote low correlation.

state of flux in which authors are constantly changing the topics they are in. As such, we refer to this period as a transitory epoch. Thus our analysis has identified four main epochs in the ACL corpus between 1980 and 2008: three focused periods of work, and one transitory phase.

These epochs correspond to natural eras in the ACL's history. During the 1980's, there were coherent communities of research on natural language understanding and parsing, generation, dialog, unification and other grammar formalizations, and lexicons and ontologies.

The 1989–1994 era corresponds to a number of important US government initiatives: MUC, ATIS, and the DARPA workshops. The Message Understanding Conferences (MUC) were an early initiative in information extraction, set up by the United States Naval Oceans Systems Center with the support of DARPA, the Defense Advanced Research Projects Agency. A condition of attending the MUC workshops was participation in a required evaluation (bakeoff) task of filling slots in templates about events, and began (after an exploratory MUC-1 in 1987) with MUC-2 in 1989, followed by MUC-3 (1991), MUC-4 (1992), MUC-5 (1993) and MUC-6 (1995) (Grishman and Sundheim, 1996). The Air Travel Information System (ATIS) was a task for measuring progress in spoken language under-

standing, sponsored by DARPA (Hemphill et al., 1990; Price, 1990). Subjects talked with a system to answer questions about flight schedules and airline fares from a database; there were evaluations in 1990, 1991, 1992, 1993, and 1994 (Dahl et al., 1994). The ATIS systems were described in papers at the DARPA Speech and Natural Language Workshops, a series of DARPA-sponsored workshsop held from 1989–1994 to which DARPA grantees were strongly encouraged to participate, with the goal of bringing together the speech and natural language processing communities.

After the MUC and ATIS bakeoffs and the DARPA workshops ended, the field largely stopped publishing in the bakeoff topics and transitioned to other topics; participation by researchers in speech recognition also dropped off significantly. From 2002 onward, the field settled into the modern era characterized by broad multilingual work and specific areas like dependency parsing, statistical machine translation, information extraction, and sentiment analysis.

In summary, our methods identify four major epochs in the ACL's history: an early NLP period, the "government" period, a transitory period, and a modern integration period. The first, second, and fourth epochs are periods of sustained topical coherence, whereas the third is a transitory phase during which the field moved from the bakeoff work to modern-day topics.

4 Identifying Participant Flows

In the previous section, we used topic co-membership to identify four coherent epochs in the ACL's history. Now we turn our attention to a finer-grained question: How do scientific areas or movements arise? How does one research area develop out of another as authors transition from a previous research topic to a new one? We address this question by tracing the paths of authors through topics over time, in aggregate.

4.1 Topic Clustering

We first group topics into clusters based on how authors move through them. To do this, we group years into 3-year time windows and consider adjacent time periods. We aggregate into 3-year windows because

the flow across adjacent single years is noisy and often does not accurately reflect shifts in topical focus. For each adjacent pair of time periods (for example, 1980–1982 and 1983–1985), we construct a matrix S capturing author flow between each topic pair, where the S_{ij} entry is the number of authors who authored in topic i during the first time period and authored in topic j during the second time period. These matrices capture people flow between topics over time.

Next we compute similarity between topics. We represent each topic by its flow profile, which is simply the concatenation of all its in- and out-flows in all of the S matrices. More formally, let F_i be the resulting vector after concatenating the i-th row (transposed into a column) and i-th column of every S matrix. We compute a topic-topic similarity matrix T where T_{ij} is the Pearson correlation coefficient between F_i and F_j. Two topics are then similar if they have similar flow profiles. Note that topics don't need to share authors to be similar — authors just need to move in and out of them at roughly the same times. Through this approach, we identify topics that play similar roles in the ACL's history.

To find a grouping of topics that play similar roles, we perform hierarchical complete link clustering on the T matrix. The goal is to identify clusters of topics that are highly similar to each other but are dissimilar from those in other clusters. Hierarchical clustering begins with every topic forming a singleton cluster, then iteratively merges the two most similar clusters at every step until there is only one cluster of all topics remaining. Every step gives a different clustering solution, so we assess cluster fitness using Krackhard and Stern's E-I index, which measures the sum of external ties minus the sum of internal ties divided by the sum of all ties. Given T as an input, the E-I index optimizes identical profiles as clusters (i.e., topic stages), not discrete groups. The optimal solution we picked using the E-I index entails 9 clusters (shown in Table 1), numbered roughly backwards from the present to the past. We'll discuss the names of the clusters in the next section.

4.2 Flows Between Topic Clusters

Now that we have grouped topics into clusters by how authors flow in and out of them, we can com-

pute the flow between topics or between topic clusters over time. First we define what a flow between topics is. We use the same flow matrix used in the above topic clustering: the flow between topic i in one time period and topic j in the following time period is simply the number of authors present in both at the respective times. Again we avoid normalizing because the volume of people moving between topics is relevant.

Now we can define flow between clusters. Let $\mathbf{A}$ be the set of topics in cluster C_1 and let $\mathbf{B}$ be the set of topics in cluster C_2. We define the flow between C_1 and C_2 to be the average flow between topics in $\mathbf{A}$ and $\mathbf{B}$:

$$f(C_1, C_2) = \frac{\sum_{A \in \mathbf{A}, B \in \mathbf{B}} f(A, B)}{|A| \cdot |B|}$$

(where $f(A, B)$ represents the topic-topic flow defined above). We also tried defining cluster-cluster flow as the maximum over all topic-topic flows between the clusters, and the results were qualitatively the same.

Figure 2 shows the resulting flows between clusters. Figure 2a shows the earliest period in our (post-1980) dataset, where we see reflections of earlier natural language understanding work by Schank, Woods, Winograd, and others, quickly leading into a predominance of what we've called "Classic Linguistic Topics". Research in this period is characterized by a more linguistically-oriented focus, including syntactic topics like unification and categorial grammars, formal syntactic theory, and prepositional phrase attachments, linguistic semantics (both lexical semantics and formal semantics), and BDI dialog models. Separately we see the beginnings of a movement of people into phonology and discourse and also into the cluster we've called "Automata", which at this stage includes (pre-statistical) Parsing and Tree Adjoining Grammars.

In Figure 2b we see the movement of people into the cluster of government-sponsored topics: the ATIS and MUC bakeoffs, and speech.

In Figure 2c bakeoff research is the dominant theme, but people are also beginning to move in and out of two new clusters. One is Early Probabilistic Models, in which people focused on tasks like Part of Speech tagging, Collocations, and Lexical Acqui-
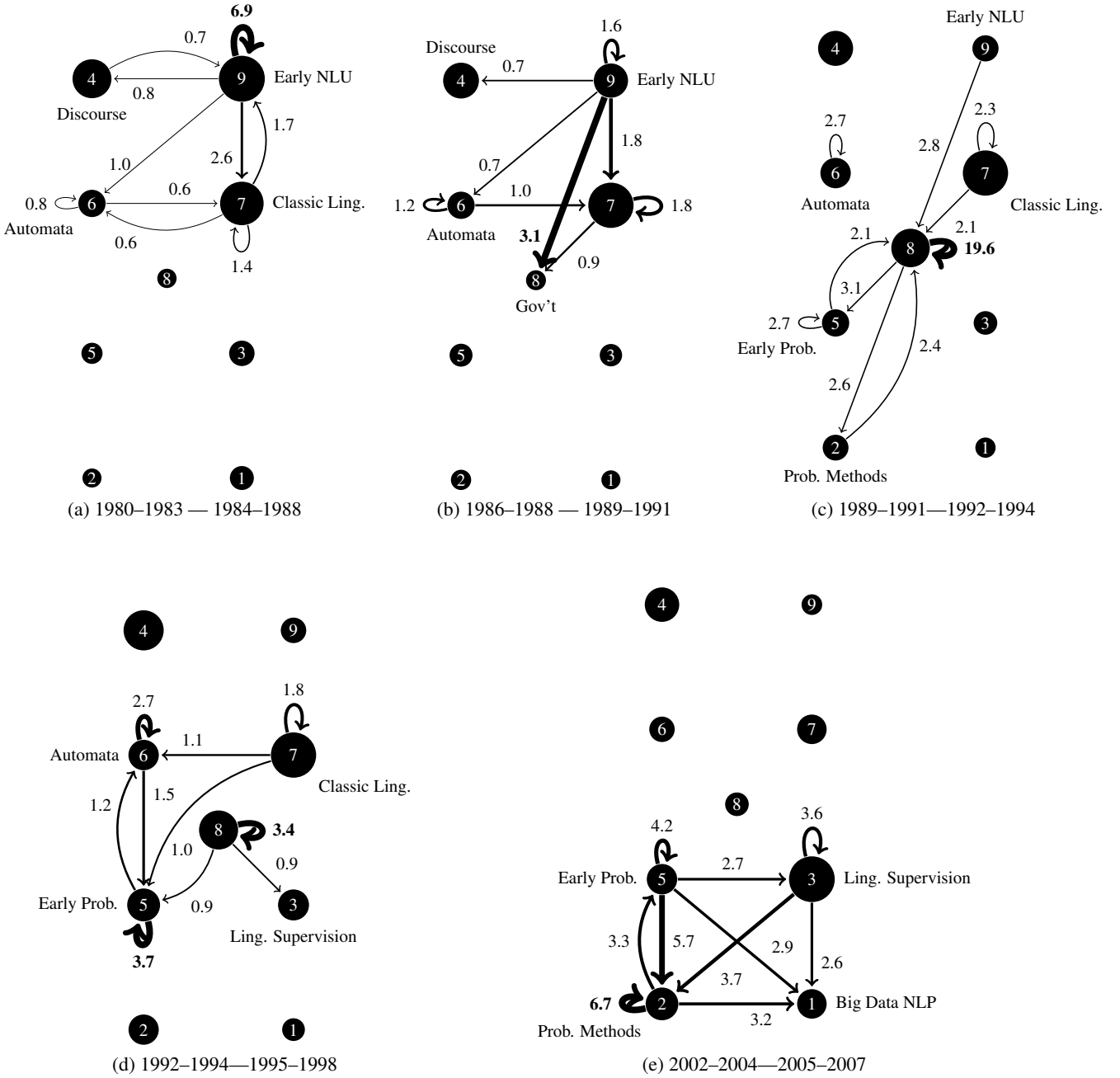
Figure 2: Author flow between topic clusters in five key time periods. Clusters are sized according to how many authors are in those topics in the first time period of each diagram. Edge thickness is proportional to volume of author flow between nodes, relative to biggest flow in that diagram (i.e. edge thicknesses in are not comparable across diagrams).

sition of Verb Subcategorization. People also begin to move specifically from the MUC Bakeoffs into a second cluster we call Probabilistic Methods, which in this very early stage focused on Evaluations Metrics and Experimental/Evaluation Methods. People working in the "Automata" cluster (Tree Adjoining Grammar, Parsing, and by this point Finite State Methods) continue working in these topics.

By Figure 2d, the Early Probability topics are very central, and probabilistic terminology and early tasks (tagging, collocations, and verb subcategorization) are quite popular. People are now moving into a new cluster we call "Linguistic Supervised", a set of tasks that apply supervised machine learning (usually classification) to tasks for which the gold labels are created by linguists. The first task to appear in this area was Named Entity Recognition, populated by authors who had worked on MUC, and the core methods topics of Machine Learning Classification and Linguistic Annotation. Other tasks like Word Sense Disambiguation soon followed.

By Figure 2e, people are leaving Early Probability topics like part of speech tagging, collocations, and non-statistical MT and moving into the Linguistic Supervised (e.g., Semantic Role Labeling) and Probabilistic Methods topics, which are now very central. In Probabilistic Methods, there are large groups of people in Statistical Parsing and N-grams. By the end of this period, Prob Methods is sending authors to new topics in Big Data NLP, the biggest of which are Statistical Machine Translation and Sentiment Analysis.

In sum, the patterns of participant flows reveal how sets of topics assume similar roles in the history of the ACL. In the initial period, authors move mostly between early NLP and classic linguistics topics. This period of exchange is then transformed by the arrival of government bakeoffs that draw authors into supervised linguistics and probabilistic topics. Only in the 2000's did the field mature and begin a new period of cohesive exchange across a variety of topics with shared statistical methods.

5 Member Retention and Field Integration

How does the ACL grow or decline? Do authors come and go, or do they stay for long periods? How much churn is there in the author set? How do these
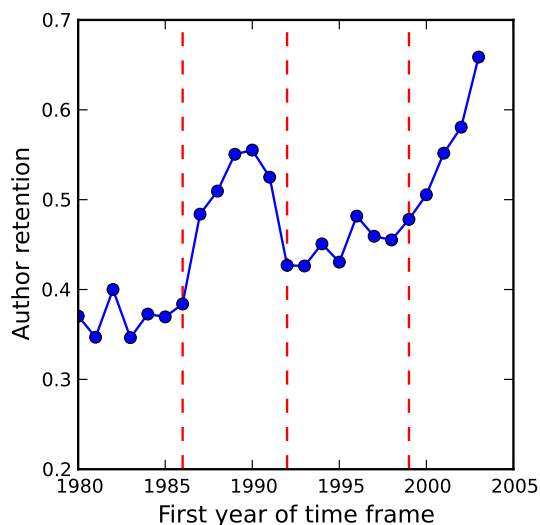

Figure 3: Overlap of authors in successive 3-year time periods over time. The x-axis indicates the first year of the 6-year time window being considered. Vertical dotted lines indicate epoch boundaries, where a year is a boundary if the first time period is entirely in one epoch and the second is entirely in the next.

trends align with the epochs we identified? To address these questions, we examine author retention over time — how many authors stay in the field versus how many enter or exit.

In order to calculate membership churn, we calculate the Jaccard overlap in the sets of people that author in adjacent 3-year time periods. This metric reflects the author retention from the first period to the second, and is inherently normalized by the number of authors (so the growing number of authors over time doesn't bias the trend). We use 3-year time windows since it's not unusual for authors to not publish in some years while still remaining active. We also remove the bulk of one-time authors by restricting the authors under consideration to those who have published at least 10 papers, but the observed trend is similar for any threshold (including no threshold). The first computation is the Jaccard overlap between those who authored in 1980–1982 and those who authored in 1983–1985; the last is between the author sets of the 2003–2005 and 2006–2008 time windows. The trend is shown in Figure 3.

The author retention curve shows a clear alignment with the epochs we identified. In the first

epoch, the field is in its infancy: authors are working in a stable set of topics, but author retention is relatively low. Once the bakeoff epoch starts, author retention jumps significantly — people stay in the field as they continue to work on bakeoff papers. As soon as the bakeoffs end, the overlap in authors drops again. The fact that author retention rocketed upwards during the bakeoff epoch is presumably caused by the strong external funding incentive attracting external authors to enter and repeatedly publish in these conferences.

To understand whether this drop in overlap of authors was indeed indicative of authors who entered the field mainly for the bakeoffs, we examined authors who first published in the database in 1989. Of the 50 most prolific such authors (those with more than 8 publications in the database), 25 (exactly half) were speech recognition researchers. Of those 25 speech researchers, 16 exited (never published again in the ACL conferences) after the bakeoffs. But 9 (36%) of them remained, mainly by adapting their (formerly speech-focused) research areas toward natural language processing topics. Together, these facts suggest that the government-sponsored period led to a large influx of speech recognition researchers coming to ACL conferences, and that some fraction of them remained, continuing with natural language processing topics.

Despite the loss of the majority of the speech recognition researchers at the end of the bakeoff period, the author retention curve doesn't descend to pre-bakeoff levels: it stabilizes at a consistently higher value during the transitory epoch. This may partly be due to these new researchers colonizing and remaining in the field. Or it may be due to the increased number of topics and methods that were developed during the government-sponsored period. Whichever it is, the fact that retention didn't return to its previous levels suggests that the government sponsorship that dominated the second epoch had a lasting positive effect on the field.

In the final epoch, author retention monotonically increases to its highest-ever levels; every year the rate of authors publishing continuously rises, as does the total number of members, suggesting that the ACL community is coalescing as a field. It is plausible that this final uptick is due to funding — governmental, industrial, or otherwise — and it is an in-teresting direction for further research to investigate this possibility.

In sum, we observe two epochs where member retention increases: the era of government bakeoffs (1989–1994) and the more recent era where NLP has received significantly increased industry interest as well as government funding (2002–2008). These eras may thus both be ones where greater external demand increased retention and cohesion.

6 Conclusion

We offer a new people-centric methodology for computational history and apply it to the AAN to produce a number of insights about the field of computational linguistics.

Our major result is to elucidate the many ways in which the government-sponsored bakeoffs and workshops had a transformative effect on the field in the early 1990's. It has long been understood that the government played an important role in the field, from the early support of machine translation to the ALPAC report. Our work extends this understanding, showing that the government-supported bakeoffs and workshops from 1989 to 1994 caused an influx of speech scientists, a large percentage of whom remained after the bakeoffs ended. The bakeoffs and workshops acted as a major bridge from early linguistic topics to modern probabilistic topics, and catalyzed a sharp increase in author retention.

The significant recent increase in author overlap also suggests that computational linguistics is integrating into a mature field. This integration has drawn on modern shared methodologies of statistical methods and their application to large scale corpora, and may have been supported by industry demands as well as by government funding. Future work will be needed to see whether the current era is one much like the bakeoff era with an outflux of persons once funding dries up, or if it has reached a level of maturity reflective of a well-established discipline.

Acknowledgments

References

A. Aris, B. Shneiderman, V. Qazvinian, and D. Radev. 2009. Visual overviews for discovering key papers and influences across research fronts. *Journal of the American Society for Information Science and Technology*, 60(11):2219–2228.

C. Au Yeung and A. Jatowt. 2011. Studying how the past is remembered: towards computational history through large scale text mining. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1231–1240. ACM.

S. Bird, R. Dale, B.J. Dorr, B. Gibson, M. Joseph, M.Y. Kan, D. Lee, B. Powley, D.R. Radev, and Y.F. Tan. 2008. The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proc. of the 6th International Conference on Language Resources and Evaluation Conference (LREC'08)*, pages 1755–1759.

D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(5):993–1022.

D.A. Dahl, M. Bates, M. Brown, W. Fisher, K. Hunicke-Smith, D. Pallett, C. Pao, A. Rudnicky, and E. Shriberg. 1994. Expanding the scope of the atis task: The atis-3 corpus. In *Proceedings of the workshop on Human Language Technology*, pages 43–48. Association for Computational Linguistics.

S. Gerrish and D.M. Blei. 2010. A language-based approach to measuring scholarly impact. In *Proceedings of the 26th International Conference on Machine Learning*.

T.L. Griffiths and M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5228.

R. Grishman and B. Sundheim. 1996. Message understanding conference-6: A brief history. In *Proceedings of COLING*, volume 96, pages 466–471.

David Hall, Daniel Jurafsky, and Christopher D. Manning. 2008. Studying the history of ideas using topic models. In *Proceedings of EMNLP 2008*.

C.T. Hemphill, J.J. Godfrey, and G.R. Doddington. 1990. The atis spoken language systems pilot corpus. In *Proceedings of the DARPA speech and natural language workshop*, pages 96–101.

P. Price. 1990. Evaluation of spoken language systems: The atis domain. In *Proceedings of the Third DARPA Speech and Natural Language Workshop*, pages 91–95. Morgan Kaufmann.

D.R. Radev, P. Muthukrishnan, and V. Qazvinian. 2009. The acl anthology network corpus. In *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, pages 54–61. Association for Computational Linguistics.

Y. Tu, N. Johri, D. Roth, and J. Hockenmaier. 2010. Citation author topic model in expert search. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1265–1273. Association for Computational Linguistics.