Inaugural meeting of the Society for Computation in Linguistics

To be held concurrently with the Annual meeting of the LSA in Salt Lake City January 4-7 2018

Invited talks form a special session on "Perceptrons and Syntactic Structures at 60" funded by NSF conference grant BCS-1651142 to the University of Massachusetts Amherst (views expressed are not necessarily those of the NSF). These talks are indicated as PSS@60 below.

This year's Cognitive Modeling and Computational Linguistics meeting is also being held with the SCiL meeting. The talks are on Sunday, and the CMCL posters will be presented in the Friday SCiL poster session (indicated as CMCL alongside the abstracts), and in a special CMCL poster session Sunday. All of the poster session abstracts can be found below following the talk abstracts, which are given in order of presentation.

Thursday

4:00
Jordan Kodner (University of Pennsylvania)
Part-of-Speech Learning as Iterative Prototype-Driven Clustering

We lay out a computational model for syntactic category acquisition which combines psychologically plausible concepts from minimally supervised part-of-speech tagging applications with simple distributional statistics. The model assumes a small set of seed words (Haghighi & Klein 2006), an approach with motivation in Pinker (1984)'s semantic bootstrapping hypothesis, and iteratively constructs a hierarchical clustering over a growing lexicon. Clustering is performed on the basis of word-adjacent syntactic frames alone (Mintz 2003) which has been shown to yield qualitatively coherent POS clusters (Reddington 1998), with no reference to morphological features. A prototype-driven labelling process based on tree-distance yields results comparable to computationally complex unsupervised algorithms while maintaining its cognitive underpinnings.

4:30
Spencer Caplan (University of Pennsylvania)
Word Learning as Category Formation

A fundamental question in word learning is how, given only evidence about what objects a word has previously referred to, children are able to generalize the total class (Smith, 1979; Xu and Tenenbaum, 2007). E.g. how a child ends up knowing that `poodle' only picks out a specific subset of dogs rather than the whole class and vice versa. Here we present a tractable computational model of word learning as category formation. Our model accounts for a wide range of previously conflicting experimental findings, including the 'Suspicious Coincidence Effect' and its sensitivity to stimulus presentation style (Spencer et al, 2011).

5:00
Ryan Cotterell (Johns Hopkins University)
Christo Kirov (Johns Hopkins University)
Mans Hulden (University of Colorado)
Jason Eisner (Johns Hopkins University)
On the Complexity and Typology of Inflectional Morphological Systems

We quantify the linguistic complexity of different languages' morphological systems. We find a trade-off between paradigm size and irregularity: a morphological system may be either large in size or highly irregular, but never both. Our methodology quantifies paradigm size as the entropy over inflectional slots (thus, rarely used slots count for less). We quantify paradigm irregularity as the entropy of the distribution over paradigms -- how hard it is to jointly predict all the surface forms of a paradigm -- and derive a variational approximation. Our measurements are taken on large morphological paradigms from 31 typologically diverse languages.

5:30 - 7:00 **Poster Session 1**

Friday

9:00 (1 hour plenary, PSS@60)
Paul Smolensky (Microsoft Research and Johns Hopkins)
Vertical Integration of Neural and Symbolic Computation: Theory and Experiment

I will present results of a research program that 'vertically' integrates symbolic and neural computation: one and the same computational system looks like neural computation at a lower, fine-grained level of description, and like a symbolic system at a higher, more abstract level. I will summarize current theoretical results on what symbolic functions can be computed in such systems. Then I will present recent experimental results on how learning in such systems can create structural roles in a data-driven way to solve problems in domains such as question-answering and image captioning. The symbolic level enhances the interpretability of these networks.

10:00 (PSS@60)
Matt Goldrick (Northwestern University)
Pyeong Whan Cho (Johns Hopkins University)
Gradient Symbolic Computation: Integrating continuous computation and discrete linguistic knowledge

A fundamental issue in theories of linguistic cognition is the need to reconcile the dynamic, continuous nature of neural processing with the discrete, symbolic nature of linguistic knowledge. Gradient Symbolic Computation (GSC) addresses this challenge by instantiating linguistic representations within a continuous, abstract symbolic space, organized around the dimensions that define linguistic structure. Structure-governed linguistic computations occur via connectionist computation: dynamic evolution of mental representations within this continuous space. Case studies show how this theoretical perspective provides new insights into fundamental problems in language processing and the structure of grammatical knowledge.

10:30 (PSS@60)
Robert Frank (Yale University)
Tal Linzen (Johns Hopkins University)
Tom McCoy (Johns Hopkins University)
Joe Pater (University of Massachusetts Amherst)
Neural network syntax in the age of deep learning: the case of question formation

Elman (1990) famously demonstrated that recurrent neural networks could model simple syntax: trained to perform word prediction, Simple Recurrent Networks learn to enforce subject-verb agreement across unbounded distances. While subsequent work pointed to SRNs' difficulties with more complex grammatical patterns, alternative recurrent network architectures have been behind recent advances in NLP technologies. We might ask then whether these new architectures could have relevance for language science as well. Here, we build on Frank and Mathis's (2007) SRN experiments, exploring whether modern sequence to sequence models used in machine translation overcome poverty of the stimulus issues in question formation.

11:00 (PSS@60)
Perceptrons and Syntactic Structures at 60 Discussion Session

11:30
Thomas Graf (Stony Brook University)
Grammar Size and Quantitative Restrictions on Movement

Every Minimalist grammar can be converted into a normal form such that every phrase moves at most once in a derivation. While the normal form reduces the complexity of movement dependencies, it also runs the risk of massively increasing the size of the grammar. I show that no lexical blow-up obtains with linguistically plausible grammars that respect common constraints on movement. This establishes not only the cost-free nature of this normal form for realistic grammars, but also that the known restrictions on movement greatly reduce the range of licit movement configurations relative to what unconstrained Minimalist grammars are capable of.

2:00 - 3:30 **Poster Session 2**

3:30
Zoey Liu (University of California, Davis)
Kenji Sagae (University of California, Davis)
Dependency Length Minimization and Lexical Frequency in Prepositional Phrase Ordering in English

Previous research has shown cross-linguistically that the human language parser prefers constituent orders that minimize the distance between syntactic heads and their dependents, but the interaction between dependency length minimization (DLM) and other factors governing linear word ordering is still unknown. We examine the effects of DLM, lexical frequency, and the traditional rule of Manner before Place before Time (MPT) in ordering of prepositional phrase (PP) adjuncts in English using corpora in different language genres annotated with syntactic structure. While MPT and DLM were consistently predictive of PP ordering in our analysis, lexical frequency information was sensitive to language genre.

4:00
Liwen Hou (Northeastern University)
David Smith (Northeastern University)
Modeling the Decline in English Passivization

Evidence from the Hansard corpus shows that the passive voice in British English has declined in relative frequency over the last two centuries. We investigate which factors are predictive of whether a given two-argument verb phrase is passivized. We show the increasing importance of the person-hierarchy effects noted by Bresnan et al. (2001), with increasing strength of the constraint against passivizing clauses with local agents, as well as the rising prevalence of such agents. Moreover, our ablation experiments on the Wall Street Journal corpus show strong support for the importance of structural parallelism noted by Weiner & Labov (1983).

4:30
Bonnie Webber (University of Edinburgh)
Hannah Rohde (University of Edinburgh)
Anna Dickinson (University of Edinburgh)
Annie Louis (University of Edinburgh)
Nathan Schneider (Georgetown University)
Explicit Discourse Connectives / Implicit Discourse Relations

While explicit discourse connectives can signal coherence relations, a common assumption is that only their absence or ambiguity necessitates relation inference. Using a crowdsourced conjunction completion task to collect 40K+ judgments on 50 discourse adverbials, we find this common assumption to be false. Instead, naive subjects systematically infer an implicit connective alongside an explicit discourse adverbial, but sometimes different subsets of subjects may each endorse different connectives. The size of these subsets means that such differences cannot be written off as error. Rather, they demonstrate how the coherence associated with explicit adverbials relates to coherence inferred between the clauses themselves.

5:00
Anne Marie Crinnion (Harvard University; Dept. of Psychological and Brain Sciences, Villanova University)
Beth Malmskog (Dept. of Mathematics & Computer Science, Colorado College; Dept. of Mathematics & Statistics, Villanova University)
Joseph C. Toscano (Dept. of Psychological and Brain Sciences, Villanova University)
A graph theoretic approach for generating hypotheses about phonetic cues in speech

Current models of speech perception suggest that combining acoustic cues and factoring out contextual variability allows listeners to recognize speech across different talkers. However, it remains unclear which specific cues are necessary and how their use varies between individual talkers. We use graph theoretic techniques to address these problems by constructing networks connecting talkers and possible cues. We identify subgraphs (Steiner trees) that connect talkers via cues consistently used to indicate specific phonemes. Classifiers trained on these cues match listeners' data better than those trained on all cues, suggesting that Steiner trees can identify the cues necessary for speech recognition.

5:30
Aleksei Nazarov (University of Huddersfield)
Learning both variability and exceptionality in probabilistic OT grammars

The co-existence of variability and exceptionality in the same language, like in Modern Hebrew (Temkin-Martínez 2010), challenges OT-style learners. Probabilistic OT (e.g., Boersma 1998) captures variability,

while exceptional words can be identified (e.g., Becker 2009) by inconsistency detection (Tesar 1995) in non-probabilistic OT; no previous proposal can do both. I propose a "soft inconsistency" criterion that identifies exceptional words in the probabilistic Expectation Driven Learning framework (Jarosz 2015), allowing learning of both variability and exceptionality. Tested on simplified Hebrew data, this model learns both the variable default pattern (>=95% accuracy) and the pattern of exceptions (>=95% overall accuracy on data).

6:00 - 7:00 Business meeting

Saturday

9:00
Émile Enguehard (École Normale Supérieure)
Edward Flemming (MIT)
Giorgio Magri (CNRS and University of Paris 8)
Statistical learning theory explains linguistic typology: a learnability perspective on OT's strict domination

This paper develops a learnability argument for Optimality Theory's assumption of strict domination. The gist of the argument is that strict domination allows a margin-based learner to achieve a smaller generalization error. The argument is twofold. On the one hand, we review error bounds in the recent statistical learning literature and bring them to bear on the assumption of strict domination. On the other hand, we report simulation results on realistic phonological test cases, comparing the generalization error achieved on target grammars which satisfy strict domination to that achieved on target grammars which flout it.

9:30
Coral Hughto (University of Massachusetts, Amherst)
Investigating the consequences of iterated learning in phonological typology

This work builds on previous investigations of the effects of learning biases on gradient typological predictions in phonology. Hughto et al (2016; in prep), using an interactive, agent-based learning model found robust biases against cumulativity effects in weighted-constraint grammars, and towards more deterministic grammars, where one output accumulates majority probability. This work compares the results of using an iterated learning model, in which "parent" agents teach "child" agents in a generational chain, and finds that the deterministic bias was not present, and the anti-cumulativity bias only emerged if child agents' initial weights were set to zero (rather than randomly sampled).

10:00
Giorgio Magri (CNRS)
Arto Anttila (Stanford University)
T-orders across categorical and probabilistic constraint-based phonology

Consider a typology T of phonological grammars, construed as mappings from underlying to surface representations. The implicational universal (x, y) --> (x', y') holds provided each grammar in T that maps x to y also maps x' to y', or statistically, assigns a probability to (x', y') that is at least as large as the probability assigned to (x, y). We develop a formal theory of implicational universals (T-orders) in OT, HG, Maxent, and their stochastic variants, and show how to efficiently compute them. Maxent T-orders turn out to behave counterintuitively in being sensitive to the number of candidates.

10:30
Adam Goodkind (Northwestern University)
Michelle Lee (Northwestern University)
Gary E. Martin (St. John's University)
Molly Losh (Northwestern University)
Klinton Bicknell (Northwestern University)
Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics

Social language deficits are universally observed in individuals with autism spectrum disorder (ASD). Quantifying the social-linguistic features of ASD has the potential to both improve clinical treatment and help identify gene-behavior relationships in ASD. Here, we extend the application of vector semantics to transcripts of semi-structured interviews with children with ASD. We find that groups of children with ASD are more semantically variable and less semantically similar relative to typically developing controls, even after accounting for differences in transcript length. These findings suggest that linguistic signatures of ASD pervade child speech broadly, and can be observed even in semi-structured conversations.

11:00 (PSS@60)
Jacob Andreas (University of California Berkeley)
Dan Klein (University of California Berkeley
Formal Semantics for Informal Worlds

Formal representations of meaning have played a key role in models for automated question answering and instruction following. But their application is fundamentally limited by the difficulty of constructing an expressive-enough formal representation of the world. Deep learning approaches promise to solve the representation problem by mapping directly between language and perceptual or behavioral primitives. But these approaches are also limited---existing models fail to generalize in ways that suggest they don't adequately capture the compositional structure of language. This talk presents two ways of injecting the flexible compositionality that formal meaning representations provide back into deep models: first as a scaffold for building utterance-specific computations, and second as a probe for understanding the structure of learned representations.

11:30 (PSS@60)
Samuel R. Bowman (NYU)
Teaching neural networks compositional semantics

Artificial neural networks now represent the state of the art in most large-scale applied language understanding tasks. This talk presents a few methods and results, organized around the task of recognizing textual entailment, which measure the degree to which these models can or do learn something resembling compositional semantics. I discuss experiments on artificial data and on a hand-built million-example corpus of natural data, and report encouraging results. I close with a brief discussion of the role of syntax and tree structure in these models.

2:00 (PSS@60)
Chris Dyer (Google DeepMind)
Recurrent Neural Networks and Bias in Learning Natural Languages

As universal function approximators, recurrent neural networks are capable of representing any distribution over sequences, given sufficient capacity. Although empirically impressive at learning distributions over natural language sentences, they have a bias toward learning to represent sentences in terms of sequential recency— a poor match for the structural dependency that is characteristic of natural language. I introduce recurrent neural network grammars (RNNGs), which add a latent structural component to sequential RNNs.

RNNGs have a bias that is more suitable for representing natural language, and results show that RNNGs are both excellent models of language, as well as predicting syntactic structure.

2:30 (PSS@60)
Perceptrons and Syntactic Structures at 60 Discussion Session

3:00 (PSS@60)
Sharon Goldwater (University of Edinburgh)
Learning more from less: can neural networks incorporate human-like learning biases?

I will discuss recent work on developing a fully unsupervised speech recognition system that segments speech audio into word-like units and clusters similar units together into lexical items. The model combines insights from Bayesian modeling (how to incorporate useful learning biases) with neural network technology for learning better representations of the audio (more speaker-independent) in a fully unsupervised way. Although cognitive plausibility was not a primary goal of the model, certain design aspects are inspired by infant perceptual learning. I will discuss these connections and more generally, how thinking about human language learning can inspire better technology and vice versa.

3:30 (PSS@60)
Jason Eisner (Johns Hopkins University)
Probabilistically Modeling Surface Patterns Using Latent Structure

A language's lexicon of surface forms and constructions includes many systematic regularities, as well as semi-regular and irregular exceptions. Generative linguists often explain regularities using shared latent representations and regular derivational processes. A probabilistic model with those elements can naturally allow for deviations from regularity and model the fact that some deviations are improbable. The probability of a derivational change can be sensitive to subtle properties of the context. I will outline several probabilistic models of the morphophonological and syntactic lexicons, which can extrapolate predictions based on their reconstruction of latent structure: e.g., underlying forms, cyclic derivations, and input-output alignments.

4:00 (PSS@60)

Emily M. Bender (University of Washington)
The Role of Linguistic Structure in Computer Natural Language Understanding

Much recent work in research on computer natural language understanding (NLU) aspires to extract meaning from text or speech inputs for a variety of applications, and typically involves machine learning systems that learn to map from domain-typical inputs to task-specific meaning representations. I will address what is meant by "meaning" in that context and the relationship between "meaning" and (linguistic) "semantics". This will lead to a discussion of the role of morphology and syntax in meaning-targeting NLP and how a grammar-based compositional approach to constructing linguistic semantic representations can increase the domain portability and precision of NLU applications.

4:30 (PSS@60)
Perceptrons and Syntactic Structures at 60 Discussion Session

Sunday (Cognitive Modeling in Computational Linguistics)

9:00 (CMCL)
Evan Jaffe (The Ohio State University)
Cory Shain (The Ohio State University)
William Schuler (The Ohio State University)
Coreference and Focus in Reading Times

This paper presents evidence of an attentional focus effect on coreference resolution in broad-coverage human sentence processing. While previous work has explored the role of attentional focus in coreference resolution (Almor, 1999; Foraker and McElree, 2007), these studies use constructed stimuli and estimate attentional focus using specific syntactic patterns (e.g. cleft constructions). This paper explores the generalizability of this focus effect on coreference resolution to the broad-coverage setting. In particular, the current work proposes several new estimators of attentional focus appropriate for broad-coverage sentence processing and evaluates them as predictors of reading behavior in the Natural Stories corpus.

9:30 (CMCL)
Adam Goodkind (Northwestern University)
Klinton Bicknell (Northwestern University)
Predictive power of word surprisal for reading times is a linear function of language model quality

Words with low probability in context take longer to read. This relationship has been quantified using information-theoretic surprisal, the amount of information a word conveys. Here, we compare surprisal estimates derived from a range of language models including n-gram models and state-of-the-art deep learning models. We show that the predictive power of surprisal for reading times improves as a tight linear function of the linguistic quality of the language model used to derive it. Further, the size of the surprisal effect is estimated consistently across all language models, pointing toward a lack of bias and striking robustness of surprisal estimates.

10:00 (CMCL)
Pyeong Whan Cho (Johns Hopkins University)
Matthew Goldrick (Northwestern University)
Richard L. Lewis (University of Michigan)
Paul Smolensky (Johns Hopkins University)
Dynamic encoding of structural uncertainty in gradient symbols

A key insight into language processing is the discovery of the relationship between processing difficulty and surprisal. We provide a mechanistic account of this effect, bridging symbolic and subsymbolic connectionist models. Gradient Symbolic Computation is a continuous-time, continuous-state stochastic dynamical systems framework that computes the representation of a discrete structure gradually. We apply this to incremental parsing and show it can dynamically encode and update structural uncertainty via the gradient activation of symbolic constituents. We show that in this model surprisal is closely related to the amount of change in the optimal activation state driven by a new word input.

10:30 (CMCL)
Laura Gwilliams (New York University)
Tal Linzen (Johns Hopkins University)
David Poeppel (New York University)
Alec Marantz (New York University)
Phonological (un)certainty weights lexical activation

Spoken word recognition involves: i) matching acoustic input to phonological categories (e.g. /b/, /p/), ii) activating words consistent with those phonological categories. Here we test the hypothesis that activation of a lexical candidate is weighted both by certainty of phonological discretisation and word frequency. Neural responses were recorded from auditory cortex using magneto-encephalography, and modelled as a function of the size and relative activation of lexical candidates. Our findings indicate that towards the beginning of a word, the processing system weights lexical candidates by both phonological certainty and lexical frequency; later into the word, activation is weighted by frequency alone.

11:00 (CMCL)
Filip Miscevic (Indiana University, Bloomington)
Aida Nematzadeh (University of California, Berkeley)
Suzanne Stevenson (University of Toronto)
Predicting and Explaining Human Semantic Search in a Cognitive Model

Recent work has attempted to characterize the structure of semantic memory and the search algorithms which, together, best approximate human patterns of search revealed in a semantic fluency task. However, these models vary in the degree of their cognitive plausibility and neglect the constraints that the incremental process of language acquisition place on the structure of semantic memory. We present a model that incrementally updates a semantic network with limited computational steps, and replicates patterns found in human semantic fluency using a random walk. We also show that both structural and semantic features are requisite for replicating human performance patterns.

11:30 (CMCL)
Yevgen Matusevych (University of Toronto)
Amir Ardalan Kalantari Dehaghi (University of Toronto)
Suzanne Stevenson (University of Toronto)
Modeling bilingual word associations as connected monolingual networks

Word associations are a common tool in research on the mental lexicon. Bilinguals tend to produce different associations in their non-native language than monolinguals do, and three mechanisms have been proposed for this difference: relying on native associations (through translation), on collocational patterns, and on phonological similarity between words. We show that the observed difference is significant, and present a computational model of bilingual word associations, implemented as a semantic network with a retrieval mechanism. Our model predicts bilingual responses better than monolingual baselines. Its success is mainly explained by translation; collocational and phonological associations do not improve the model.

12:00 - 12:30 **CMCL Poster Session**

Poster Session 1 (Thursday 5:30 - 7)

Steven Abney (University of Michigan)
A bidirectional mapping between English and CNF-based reasoners

Language transduces between sound and meaning: it is an input-output device for reasoning. Among fully-explicit models of reasoning, the commonest manipulate a normal form of predicate calculus known as Clause Normal Form (CNF). The textbook parser/interpreter (sound to meaning) produces a restricted range R of first-order predicate calculus (FOPC) expressions. There is a well-known algorithm for converting arbitrary FOPC expressions to CNF, but no algorithm for converting CNF to an expression in R. This paper addresses that lack and gives a method of inverting the interpreter to define a generator

Galia Barsever (University of California, Irvine)
Rachael Lee (University of California, Irvine)
Gregory Scontras (University of California, Irvine)
Lisa Pearl (University of California, Irvine)
Quantitatively assessing the development of adjective ordering preferences using child-directed and child-produced speech corpora

The relative ordering of adjectives is cross-linguistically robust, with multi-adjective strings like “small gray kitten” preferable to “gray small kitten”. While it has been shown that adjective subjectivity predicts adult preferences, with less subjective adjectives preferred closer to the modified noun, it remains unknown when and how this preference develops. We assess English corpora of child-produced and child-directed speech data, using quantitative metrics to determine the underlying representation most likely to yield the observable data. We find strong support for a subjectivity-based representation, with qualitative similarity between adult-to-adult and child-directed data and development of adult-like subjectivity-based preferences by age two.

Shohini Bhattasali (Cornell University, USA)
John Hale (Cornell University, USA)
Christophe Pallier (INSERM-CEA Cognitive Neuroimaging Unit, France)
Jonathan R. Brennan (University of Michigan, USA)
Wen-Ming Luh (Cornell University, USA)
R. Nathan Spreng (Cornell University, USA)
Differentiating Phrase Structure Parsing and Memory Retrieval in the Brain

On some level, human sentence comprehension must involve both memory retrieval and structural composition. This study differentiates these two processes using neuroimaging data collected during naturalistic listening. Retrieval is formalized in terms of "multiword expressions" while structure-building is formalized in terms of bottom-up parsing. The results most strongly implicate Anterior Temporal regions for structure-building and Precuneus Cortex for memory retrieval.

Emmanuele Chersoni (Aix-Marseille University)
Alessandro Lenci (University of Pisa)
Philippe Blache (Aix-Marseille University)
Logical Metonymy in a Distributional Model of Sentence Comprehension

Logical metonymy is defined as the combination of an event-subcategorizing verb with an entity-denoting object (The author began the book), so that its interpretation requires the retrieval of a covert event (writing). Psycholinguistic studies revealed extra processing costs for logical metonymy, a phenomenon generally explained with the introduction of new semantic structure.
In this paper, we present a distributional model for sentence comprehension inspired by the Memory,

Unification and Control model by Hagoort (2016). We show that our framework can account for the extra costs of logical metonymy and can identify the covert event in a classification task.

Bruce Hayes (UCLA) *(Talk alternate)*
Allomorph discovery as a basis for learning alternations

I describe an implemented system for learning the allomorphs of morphemes. It inputs glossed paradigms and outputs the allomorphs of each morpheme. The system takes the form of a maxent grammar, with constraints preferring minimal alternation and segmental contiguity. Learning the allomorphs first greatly simplifies the learning of underlying forms: string alignment of allomorphs reveals the full set of alternations, which can be used to construct a highly informative candidate set from which the correct UR's and constraint rankings are readily discovered. To illustrate, I show how the system learns UR's and rankings for several well-known phonology problems.

Andres Karjus (University of Edinburgh)
Richard A. Blythe (University of Edinburgh)
Simon Kirby (University of Edinburgh)
Kenny Smith (University of Edinburgh)
Topical advection as a baseline model for corpus-based lexical dynamics

An important question in the field of corpus-based evolutionary language dynamics research is concerned with distinguishing genuine linguistic change (selection) from neutral evolution, and from changes stemming from language-external factors (cultural drift). A commonly used proxy to the popularity or selective fitness of an element over time is its frequency in a representative corpus. However, it has been pointed out recently that raw frequencies can often be misleading due to shifting discourse topics and societal trends. We propose a computationally simple model to control for topical drift and demonstrate its capacity to account for variability in word frequency changes over time.

Anna Mai (UC San Diego)
Eric Baković (UC San Diego)
Matt Goldrick (Northwestern University)
Phonological opacity as local optimization in Gradient Symbolic Computation

We present a novel approach to counterbleeding rule interactions in Yokuts (Californian) using Gradient Symbolic Computation (GSC). GSC, a dynamical systems model, optimizes two constraint sets: a set specifying a Harmonic Grammar (HG) and a set of *quantization* constraints preferring discrete symbolic states. During optimization, quantization strength gradually increases, increasing the relative harmony of discrete symbolic vs. intermediate blend states. The output of the system therefore reflects the dynamics of optimization, not simply grammatical harmony. With appropriate dynamics, relatively high harmony intermediate states can trap optimization near less globally harmonic but locally optimal symbolic candidates; this can model Yokuts counterbleeding.

Elliott Moreton (University of North Carolina, Chapel Hill) *(Talk alternate)*
Conditions on abruptness in a gradient-ascent Maximum Entropy learner

When does a gradual learning rule translate into gradual learning performance? This paper studies a gradient-ascent Maximum Entropy learner in a two-alternative forced-choice task. The main result is that if all initial weights are zero, then 2AFC performance improves fastest at the outset, making later abrupt acceleration impossible. The 2AFC learning curve of a learner whose initial state is near zero converges to that of one whose initial state is exactly zero. Large nonzero initial weights can produce slow initial learning that accelerates later. Abrupt learning is thus an effect of transfer from Universal Grammar or from a previously-learned weighting.

Brandon Prickett (University of Massachusetts, Amherst)
Similarity-based Phonological Generalization

Halle (1978) suggested that simple, feature-bundle-based representations could explain the generalization seen in the phonology of languages. An alternative explanation for this is that the similarity of different sounds causes speakers to treat them similarly (Cristia et al. 2013). To test which explanation better predicts the experimental results of Cristia et al. (2013), I created a MaxEnt learner that uses similarity in its learning update to encourage generalization to similar segments. My learner's predictions match the results found by Cristia et al. (2013) more accurately than a previously proposed MaxEnt learner (Moreton et al. 2017) that relies on feature-bundle-based generalization.

Natalie M. Schrimpf (Yale University)
Using Rhetorical Topics for Automatic Summarization

Summarization involves finding the most important information to convey a document's meaning. I present a method for using topic information to influence which content is selected for a summary. Texts are divided into topics using rhetorical information that creates a partition into a sequence of topics. I compare the output of summarizing a text without topics to summarizing individual topics and combining them into a summary. The results show that the use of rhetorical topics improves summarization performance compared to a summarization system that incorporates no topic information, demonstrating the utility of topic structure and rhetorical information for automatic summarization.

Zachary Stone (Department of Linguistics, University of Maryland, College Park)
A structural theory of syntactic derivations

We describe a category of structured sets and show how to use it to model syntactic derivations. Its objects are (derived) trees (as partial orders) connected by order-preserving maps. This generalization allows for feature-sharing (Pesetsky & Torrego 2007; Frampton & Gutmann 2000) and feature geometry (Harley and Ritter 2002; Bye & Svenonius 2011). This category induces good definitions for isomorphisms (which keep track of the dependency structure of each derived object, and relations between the dependency structures between steps) and substructures (describing constituency). The category admits many "good" constructions such as products and coproducts. We then give applications of these constructions for formalizing grammatical operations as pushouts (Ehrig et al. 1973; Ehrig et al. 1997; Van den Broek 1991), including showing consequences for feature-sharing models of agreement.

Mai Ha Vu (University of Delaware)
Toward a formal description of NPI-licensing patterns

This paper is a model-theoretic study of a simplified version of Negative Polarity Item (NPI) licensing requirements in two languages, English and Hungarian. Using subregular logical formalisms defined over tree-languages, I show that neither pattern can be described with Tier-based Strictly Local (TSL) constraints, and suggest that they need more complex logical formula. In particular, Hungarian patterns can be described using Tier-based Locally Testable (TLT) constraints, whereas for English, even that is not sufficient. I also give a definition of two subregular tree-languages, TSL and TLT, previously only defined for String-languages.

Shiying Yang (Brown University)
Chelsea Sanker (Brown University)
Uriel Cohen Priva (Brown University)
The Organization of Lexicons: A cross-linguistic analysis of monosyllabic words

Lexicons utilize a fraction of licit structures. Different theories predict either that lexicons prioritize contrastiveness or structural economy. Study 1 finds that the lexicon of Mandarin is less distinctive for CVX syllables than a randomly sampled baseline using the phonological inventory, as predicted in

Dautriche et al. 2017: lexicons are more regular rather than distinctive. Study 2 finds that the lexicons of Mandarin and American English have fewer phonotactically complex words than the random baseline: Words tend not to have multiple infrequent components. This suggests that phonological constraints can have superadditive penalties for combined violations, consistent with e.g. Albright (ms.).

Poster Session 2 (Friday 2 - 3:30)

Samira Abnar (University of Tehran)
Max Mijnheer (University of Amsterdam)
Rasyan Ahmed (University of Amsterdam)
Willem Zuidema (University of Amsterdam)
Distributional and Dependency-based Word Embeddings have Complementary Roles in Decoding Brain Activity

We evaluate different word embeddings on their usefulness for predicting the neural activation patterns associated with concrete nouns. Our goal is to assess the cognitive plausibility of these models, and understand how we can improve the methods for interpreting brain imaging data. We show that neural word embeddings exhibit superior performance beating experiential word representations. Interestingly, the error patterns of these models are markedly different. This may support the idea that the brain uses different systems for processing different kinds of words. We suggest that taking the relative strengths of different embedding models into account will lead to better models. **(CMCL)**

Alëna Aksënova (Stony Brook University)
Sanket Deshmukh (Stony Brook University)
Formal restrictions on multiple tiers

The class of tier-based strictly local (TSL) languages has shown itself as a good fit for natural language patterns. Although there are some cases when one TSL grammar is not enough, there have never been proposed any limitations on tier alphabets of several cooperating TSL grammars. Here, we use harmonic systems with multiple feature spreadings as the litmus test for the possible configurations of tier alphabets. While theoretically possible relations among them are containment, disjunction and intersection, the latter one is unattested and we show why the absence of such configuration might simplify the system in whole.

Abigail Benecke (Villanova University)
Joseph Toscano (Villanova University)
How far can VOT take us? Voicing categorization with and without the use of VOT

Voice-onset time (VOT) is an extremely reliable cue to word-initial stop voicing, such that VOT alone may be sufficient as a voicing cue. To test this, 35 potential cues were measured and used to train logistic regression classifiers, asking whether VOT is sufficient, whether other cues increase categorization accuracy, and whether, without VOT, other cues produce listener-level accuracy. Results show that human-like performance was never achieved without VOT or with VOT alone. Models using a cue-integration approach (additively combining multiple cues) offered the closest performance to human listeners. Thus, VOT appears to be necessary, but not sufficient, for voicing judgments.

Aniello De Santo (Stony Brook University)
Extending TSL to Account for Interactions of Local and Non-Local Constraints

Recent research in computational linguistics suggests that unbounded dependencies in phonotactics, morphology, and maybe even syntax can all be captured by the class of Tier-based Strictly Local languages (TSL). Here, I explore the consequences of relaxing a particular constraint on the tier-projection mechanism of TSL grammars. I show how a more general definition of tier-projection naturally extends

TSL while preserving all its formal properties, and easily captures patterns in which local and non-local dependencies interact, that are unaccountable for in standard TSL accounts. This results support subregularity as a good computational hypothesis for phonotactic complexity.

Jonathan Dunn (Illinois Institute of Technology)
Modeling the Complexity and Descriptive Adequacy of Construction Grammars

This paper uses the Minimum Description Length paradigm to model both the complexity of a CxG itself and the descriptive adequacy of that CxG against an unannotated corpus. These two quantities are combined to create a discovery-device CxG that searches for the optimum grammar to describe a corpus, with optimality defined using the MDL metric. Results for English, Spanish, French, German, and Italian show (i) that these grammars provide significant generalizations as measured using the MDL metric and (ii) that more complex CxGs with access to multiple levels of representation provide greater generalizations than single-representation CxGs.

Marina Ermolaeva (University of Chicago)
Daniel Edmiston (University of Chicago)
Distributed Morphology as a regular relation

This research reorganizes the Distributed Morphology (DM) framework to work over strings. Typically, DM operates on binary trees, with the syntax-morphology interface implicitly treated as a tree-transducer. We contend that using (binary) trees is overpowered, predicting patterns unattested in natural language. Assuming the standard Y-model, DM operating on trees presumes that the flattening of the derivation for PF takes place post-morphology. We however flatten the structure above the morphological module, between the syntax and morphology. Restricting the morphological component to working on strings, we correctly predict that morphology can be modeled with regular string languages.

Jena D. Hwang (Institute of Human and Machine Cognition)
Archna Bhatia (Institute of Human and Machine Cognition)
Na-Rae Han (University of Pittsburgh)
Tim O'Gorman (University of Colorado Boulder)
Vivek Srikumar (University of Utah)
Nathan Schneider (Georgetown University)
Double Trouble: The Problem of Construal in Semantic Annotation of Adpositions

We consider the semantics of prepositions, revisiting a broad-coverage annotation scheme used for annotating all preposition tokens in a 55,000-word corpus of English. In an attempt to resolve problematic cases in English and apply the scheme to adpositions and case markers in other languages, we reconsider the assumption that an adposition's lexical contribution is equivalent to the role/relation that it mediates, embracing the potential for construal to manage complexity and avoid sense proliferation. We suggest a framework to represent both the scene role and the adposition's lexical function, and discuss how it would allow for a simpler inventory of labels.

Andrew Lamont (University of Massachusetts, Amherst)
Subsequential steps to unbounded tonal plateauing

Recent work has shown that most phonological maps from underlying representation to surface form can be modeled by subsequential functions (see Heinz (forthcoming) for an overview). Jardine (2016) demonstrates that certain tonal phenomena are more computationally complex, exceeding the subsequential class. This paper argues that while these results hold under a parallel phonological derivation in which all transformations occur simultaneously, they are weakened under a serial derivation, in which each step is

limited to a single transformation. Specifically, it is shown that the tonal phenomena of interest can be decomposed into a series of subsequential steps.

R. Thomas McCoy (Johns Hopkins University)
Robert Frank (Yale University)
Phonologically Informed Edit Distance Algorithms for Word Alignment with Low-Resource Languages

We present three methods for weighting edit distance algorithms based on linguistic information. These methods base their penalties on (i) phonological features, (ii) distributional character embeddings, or (iii) differences between cognate words. We also introduce a novel method for evaluating edit distance through the task of low-resource word alignment by using edit-distance neighbors in a high-resource pivot language to inform alignments from the low-resource language. At this task, the cognate-based scheme outperforms our other methods and the Levenshtein edit distance baseline, showing that NLP applications can benefit from information about cross-linguistic phonological patterns.

Jeff Parker (Brigham Young University)
Robert Reynolds (Brigham Young University)
Andrea D. Sims (The Ohio State University)
A Bayesian investigation of factors shaping the network structure of inflection class systems

Some inflection class systems exhibit Marginal Detraction (MD) -- classes with fewer lexemes contribute more to a system's complexity than larger classes, where complexity is the average uncertainty associated with one form of a lexeme given knowledge of another (Stump & Finkel 2013; Sims & Parker 2016). In this paper we model the emergence of MD with a multi-generational agent-based Bayesian learning model. By treating inflection class systems as networks of classes (nodes) without overlapping exponents (edges), we show that the emergence of MD in the model depends on the network properties of the ten artificial systems used as input.

Jon Rawski (Stony Brook University)
Subregular Complexity Across Speech and Sign

Do the computational properties of phonology hold independently of modality, or are they inextricably dependent on the physical articulators which externalize it? I evaluate these claims with respect to the recent Subregular Hypothesis, which states that all phonological systems obey strict computational bounds, falling within the weakest sub-classes of the Regular acceptor/transducers. I show that in sign languages, metathesis and final syllable reduplication are Strictly Local functions, just like the parallel spoken language processes, as is compound reduction via evaluating its sequential nature. Subregularity, then, supports an algebraic phonology independent of modality.

K.J. Savinelli (University of California, Irvine)
Greg Scontras (University of California, Irvine)
Lisa Pearl (University of California, Irvine)
Exactly two things to learn from modeling scope ambiguity resolution: Developmental continuity and numeral semantics

Behavioral data suggest that both children and adults struggle to access the inverse interpretation of scopally-ambiguous utterances in certain contexts. To determine whether the causes of both child and adult difficulty are similar, we extend an existing computational model of children's scope ambiguity resolution in context. We find that the same utterance-disambiguation mechanism is active in both children and adults, supporting the theory of developmental continuity. Moreover, because adult behavior requires an exact semantics for numerals, we also provide empirical support for this theory of linguistic representation.
(CMCL)

Miikka Silfverberg (University of Colorado)
Lingshuang Jack Mao (University of Colorado)
Mans Hulden (University of Colorado)
Sound Analogies with Phoneme Embeddings

In computational linguistics, vector space models of words learned from unannotated data (word embeddings) have been shown to reliably encode subtle semantic information, offering capabilities such as solving proportional word analogy tasks of the format man:woman::king:X; answer: X = queen. We study how well such purely distributional properties carry over to similarly learned phoneme embeddings, and whether phoneme vector spaces align with articulatory distinctive features. We demonstrate a statistically significant correlation between distinctive feature spaces and phoneme vector spaces learned from raw data with different techniques. Furthermore, these distributed representations yield coherent models of proportional phoneme analogies such as p:b::t:d.

Kristina Strother-Garcia (University of Delaware)
Imdlawn Tashlhiyt Berber Syllabification is Quantifier-Free

Imdlawn Tashlhiyt Berber (ITB) is unusual due to its tolerance of non-sonorant syllabic nuclei. There are successful Rule-based and Optimality-Theoretic accounts of ITB syllabification, but they do not address the question of how complex this process is. Model theory and formal logic allow for comparison of complexity across theories by identifying the expressivity of linguistic formalisms in a grammar-independent way. I develop a mathematical formalism for representing ITB syllabification using Quantifier-Free logic, showing that ITB syllabification is relatively simple from a computational standpoint and that grammatical formalisms could succeed with even less powerful mechanisms than are currently accepted.

Dingquan Wang (Johns Hopkins University)
Jason Eisner (Johns Hopkins University)
Predicting Fine-Grained Syntactic Typology from Surface Features

We show how to predict the basic word-order facts of a novel language given only a corpus of its part-of-speech (POS) sequences. We predict how often direct objects follow their verbs, how often adjectives follow their nouns, and in general the directionalities of all dependency relations. Although recovering syntactic structure is usually regarded as unsupervised learning, we train our predictor on languages of known structure. It outperforms the state-of-the-art unsupervised learning by a large margin, especially when we augment the training data with many synthetic languages. Full details can be found in http://www.cs.jhu.edu/~jason/papers/#wang-eisner-2017.