Career Statement

Noah A. Smith (nasmith@cs.cmu.edu)

Abstract

Modulo formatting, this document constitutes a portion of my application for tenure ("section 5"). It details my research, teaching, and service efforts and goals. It is not exhaustive, and it reflects my view of my activities in late May 2013.

1 Research

My research goal is to automate inference from natural language text, including:

- algorithms that interpret text into abstract linguistic structures (§1.1);
- learning algorithms that infer the parameters of such models from text annotated by linguistic experts or non-experts, unannotated text, and text observed in its social context (§1.2);
- task-specific inferences in the form of translation into other languages or real-world predictions (§1.3).

The following elaborates some important aspects of each in turn, emphasizing developments since my review for promotion in 2010 and plans for the future. Some research threads that are not being actively pursued are suppressed for space (e.g., work on text-to-text transformations [Das and Smith, 2009; Heilman and Smith, 2010], including my former Ph.D. student Michael Heilman's dissertation project; and dynamic programming extensions [Cohen et al., 2008b; Gimpel and Smith, 2009a]).

1.1 Computational Linguistics and Natural Language Processing

Making inferences from a text might require analyzing its meaning; this conjecture underlies much current research in translation, question answering, and information extraction. Operationalizing "meaning" remains a central challenge in computational linguistics. Many agree that syntactic parsing is an important first step toward mapping strings to meanings (see Fig. 1, top, for an example). Further, representations of the propositional semantics of strings have attracted a great deal of attention (i.e., "who did what to whom?" and illustrated in Fig. 1, bottom). My NLP research seeks algorithms for text analysis, subject to the often-competing desiderata listed in Tab. 1. I call this research program **linguistic structure prediction**, which is also the title of my monograph [Smith, 2011].

Much of my early work used exact dynamic programming algorithms that find the model-optimal[1] linguistic analysis of a piece of text [Cohen et al., 2008b, 2011b; Dreyer et al., 2006; Eisner and Smith, 2011; Eisner et al., 2004a,b, 2005; Smith and Smith, 2004, 2007; Smith and Johnson, 2007; Smith et al., 2005]. Classic examples are the Viterbi algorithm [Viterbi, 1967] and the weighted version of Earley's algorithm [Earley, 1970]. These algorithms remain foundational in NLP, but using them requires us to make strong assumptions about the mappings of words to deeper linguistic structures; hence, they can be seen as weak on desideratum (iii).

1.1.1 Approximating Structured Inference: AD^3

Unfortunately, inference with guarantees of model-optimality quickly becomes intractable as our models become more linguistically expressive. For example, we might wish to capture non-local interactions between two or more syntactic or semantic arguments of a word [Das et al., 2012], different parts of a translation [Gimpel and Smith, 2009a], or between morphological (word-internal) and syntactic analyses [Cohen and Smith, 2007]. My former Ph.D. student André Martins, his co-advisors Mário Figueiredo (IST), Pedro Aguiar (IST), and Eric Xing (CMU), and I developed a range of general *approximate* inference techniques

[1] Briefly, such algorithms imply the use of a statistical model, typically a weighted grammar. By "model-optimal," we mean that the algorithm is guaranteed to find the analysis that the model deems most probable. In statistical terms, this is a kind of maximum *a posteriori* inference: $\arg\max_{\boldsymbol{y}\in\mathcal{Y}_{\boldsymbol{x}}} score(\boldsymbol{x},\boldsymbol{y})$, where $\mathcal{Y}_{\boldsymbol{x}}$ is the set of well-formed analyses of an input string $\boldsymbol{x}$ and $score$ is a function learned from data, often a probability (see §1.2).

i.	generality across different languages, genres, and linguistic representations
ii.	computational efficiency, in application and in construction
iii.	sensitivity to a wide range of the phenomena and constraints in human language
iv.	strong performance guarantees
v.	high accuracy when judged against expert human annotations

Table 1: Desiderata for algorithms (in arbitrary order).

[Martins et al., 2009a]; the most important of these is alternating directions dual decomposition (**AD**^{3}) [Martins, 2012; Martins et al., 2011a, in review]. In terms of desiderata (i–v) in Table 1, AD^3 is a general-purpose maximum *a posteriori* inference algorithm for graphical models (i). It divvies up work among "worker" nodes, each of which iteratively solve subproblems with closed-form solutions or to which efficient dynamic programming machinery can be applied (ii). AD^3 can take advantage of rich non-local features and constraints represented in first-order logic (iii). Formally (iv), AD^3 is robust to approximate solutions by worker nodes and produces a certificate when it finds an exact solution. It enjoys $O(1/\epsilon)$ convergence, an improvement over previous methods such as projected subgradient [Komodakis et al., 2007]. It produces primal and dual residuals, which allow for a straightforward stopping criterion. Finally, on a range of datasets, AD^3 outperforms strong existing techniques [Globerson and Jaakkola, 2008; Jojic et al., 2010; Komodakis et al., 2007; Sontag et al., 2011] (v); its use in NLP is discussed below. André has released an open-source implementation of AD^3 [Martins, 2012].

1.1.2 Syntactic Analysis: TurboParser

We have used AD^3 to build state-of-the-art natural language parsers. **TurboParser** is André Martins' open-source, trainable dependency parser built on AD^3 [Kong and Smith, 2013; Martins, 2009; Martins et al., 2010, 2011b, 2013]. TurboParser offers a range of options trading accuracy for speed. At one extreme (first-order "arc-factored" models) its speeds are comparable to optimized, specialized parsing methods like [Rush and Petrov, 2012]; at the other, it parses *Wall Street Journal* English at around 700 tokens per second. For nine out of fourteen languages in the CoNLL 2006–7 suite of benchmarks[2] as well as English dependency parsing with Yamada-Matsumoto and Stanford dependency representations, some variation of TurboParser achieves the best published unlabeled attachment performance. Fig. 1 (top) shows an example of TurboParser's output.

1.1.3 Semantic Analysis: SEMAFOR

Frame-semantic parsing is a kind of general-purpose text analysis that maps words to abstract *frames* in the FrameNet lexicon [Fillmore et al., 2003] and labels spans of text corresponding to the fillers of those frames' roles (see Fig. 1, bottom). Led by my former Ph.D. student Dipanjan Das, my group has developed the open-source **SEMAFOR** frame-semantic analyzer, which achieves the best-published performance on the FrameNet 1.5 benchmark corpus [Das et al., 2010a,b, 2014]. Frame-semantic analysis is quite hard: on the standard aggregate evaluation measure, we achieve 68% precision and 61% recall; subjectively, it is easy to see room for improvement.[3] Nonetheless, SEMAFOR has been downloaded hundreds of times and is being used actively by several research groups at CMU and within the FrameNet project at Berkeley. SEMAFOR also serves as an excellent testbed for algorithmic research; for example, we used AD^3 to incorporate pairwise requirement and exclusion constraints on a frame's arguments (specified by the FrameNet lexicon) [Das et al., 2012, 2014]. Ongoing projects in collaboration with Tom Mitchell's and Ed Hovy's research teams at CMU seek to integrate semantic analysis with knowledge bases and with other algorithms

[2]TurboParser is within 1% of state of the art on 3 other languages; on a fourth, it does not beat our earlier "stacked" parser [Martins et al., 2008].

[3]An online demo of both TurboParser and SEMAFOR is available at http://demo.ark.cs.cmu.edu/parse.

that extract structured information from text. Chris Dyer and I were recently awarded funding from the NSF to explore multilingual distributed semantics on very large corpora in ten languages.

1.1.4 Other Projects: Lightweight Annotation, Sociolinguistics, Social Media, Extraction

Above, I discussed *algorithms* that map a string to its best linguistic analysis, sidestepping the scoring function that defines "best." Typically this is learned from data (more discussion in §1.2). Such data tend to be expensive and require a great deal of annotation expertise; one thread of my research has sought to obtain data more cheaply, especially for languages and genres that have as-yet received less attention. With CMU–Qatar collaborators Kemal Oflazer and Behrang Mohit and my student Nathan Schneider, I have explored the viability of WordNet supersenses [Ciaramita and Johnson, 2003] as a representation for direct annotation by humans with relatively little training, producing a small corpus of supersense-tagged articles from Arabic Wikipedia [Schneider et al., 2012]. Nathan's ongoing thesis research considers annotation and models for identifying multiword expressions and disambiguating prepositions, "the bastard children of lexicon and grammar" (in Nathan's words). Another ongoing effort explores a syntactic and semantic annotation framework, called **graph fragment language**, that allows underspecification, so that an less-expert annotator may sidestep decisions about which he is uncertain or focus on a specific subtask [Schneider et al., 2013b].

In a long-running collaboration with Jacob Eisenstein, my Ph.D. student Brendan O'Connor, and Eric Xing, I have conducted **sociolinguistic** studies on very large corpora of social media messages with their associated metadata. Our studies have modeled geographic variation in lexical usage and automatic geolocation of authors [Eisenstein et al., 2010], demographic variation [Eisenstein et al., 2011; O'Connor et al., 2010b], and geographical diffusion of words over time and space [Eisenstein et al., 2012]. This work was featured on NPR's *All Things Considered*[4] and in the *New York Times*,[5] among others.

Driven by my group's growing interest in social media as a source of linguistic data for studies like these and for applications (§1.3), my group has developed an open-source part-of-speech tagger for social media messages [Gimpel et al., 2011]. This project required adjustments to standard part of speech annotation conventions (e.g., handling of emoticons and discourse markers such as at-mentions). Apart from its widespread use (over 1,300 downloads of version 0.3 at this writing), this project notably included Tobi Owoputi, a CMU CS sophomore, in improving accuracy and efficiency [Owoputi et al., 2013].

Two new projects on extracting structured information from text are worthy of note. One, a collaboration with Norman Sadeh and his team, seeks to extract key features from privacy policies for online services [Ammar et al., 2012]. Humans rarely read these documents but may benefit from automatic inferences about the implications of privacy policies before clicking "I agree." In the other project, recently funded by the Sloan Foundation, Bryan Routledge and I are applying NLP to infer relationships among American corporations from a large corpus of publically available press releases. The resulting database is intended to support studies in corporate demography (see also §1.3.3).

1.2 Machine Learning for Language

The best source of information that we have about language comes to us as *data*. Extracting useful representations and algorithms from data is known as **machine learning**; I have made leading contributions to unsupervised, latent-variable, semi-supervised, and supervised machine learning from text, for NLP. My research seeks learning algorithms that meet the desiderata in Tab. 1, with the overarching theme of marrying of linguistic domain knowledge (iii) with principled statistical reasoning that gives rise to (iv). Indeed, the title of my NSF CAREER grant is "flexible learning for natural language processing."

[4] `http://www.npr.org/2011/01/18/133024500/you-have-an-accent-even-on-twitter`

[5] `http://www.nytimes.com/2011/10/30/opinion/sunday/twitterology-a-new-science.html`

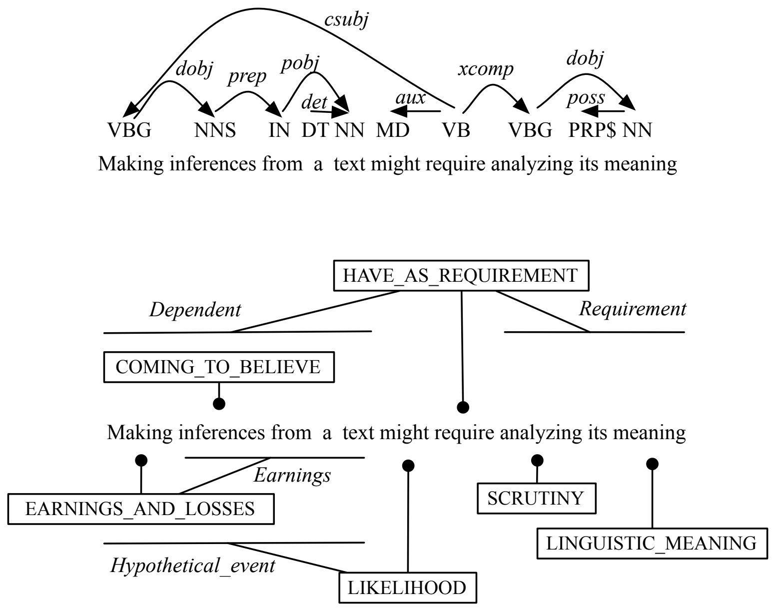

Figure 1: An example sentence (from §1.1), analyzed by TurboParser (top, discussed in §1.1.2) and SEMAFOR (bottom, discussed in §1.1.3). Here, TurboParser has been called to provide Stanford dependencies, a popular syntactic representation. TurboParser has made one obvious error, misattaching the prepositional phrase *from a text* to *inferences* rather than *making*. SEMAFOR provides FrameNet-style predicates evoked by words in the sentence (called *frames*, shown in capitals, boxed) and their arguments. Arguably, the frame for *making* should be different (*making inferences* might better be analyzed as a multiword expression), and the *Hypothetical_event* argument for *might* is the requirement, not the inference-making.

1.2.1 Unsupervised Language Learning

My dissertation research focused on unsupervised dependency parsing [Smith, 2006; Smith and Eisner, 2005b], introducing new variations on maximum likelihood estimation for probabilistic grammars [Smith and Eisner, 2005a]. Estimation corresponds to a non-convex optimization problem, so I also sought optimization techniques to find local optima corresponding to accurate models [Smith and Eisner, 2004, 2006]. Later, my former Ph.D. student Shay Cohen and I were the first to demonstrate the benefits of empirical Bayesian learning for this problem, if the prior encodes some basic knowledge about linguistic categories [Cohen and Smith, 2009, 2010b; Cohen et al., 2008a]. My former Ph.D. student Kevin Gimpel also returned to the problem of initialization, and we developed improved data-driven techniques for finding a good initializer [Gimpel and Smith, 2012b]. I am continuing work on unsupervised syntax learning in collaboration with Jason Baldridge and Dan Garrette at the University of Texas, and Chris Dyer, focusing on complex categories used by combinatory categorial grammar; Chris, our Ph.D. student Waleed Ammar, and I are also developing a new formulation for unsupervised structure learning inspired by autoencoders.

While it is well-known that unsupervised grammar learning is "hard" (in lay terms), Shay Cohen's dissertation provides a computational complexity analysis. Even given the grammar, maximum likelihood estimation for probabilistic context-free grammars from just strings is NP-hard [Cohen and Smith, 2010a]. We then turned to *sample* complexity and found polynomial bounds in the amount of data needed for learning PCFG parameters [Cohen and Smith, 2010c, 2012]. This includes a comparison between supervised and unsupervised learning.

A core language technology that has been remarkably stable for over a decade is **language modeling**, or estimating a probability distribution over sequences of words. We have explored cases where "words" may not be the best level of analysis. The first of these is scenarios where multiword expressions are commonplace; Kevin Gimpel and I applied a nonparametric Bayesian model to identify multiword units in monolingual and bilingual texts, even noncontiguous units, and showed benefit for translation [Gimpel and Smith, 2011b]. The second scenario is when rich intra-word effects (e.g., inflectional affixes) create more apparent words than we have data to estimate probabilities for. Chris Dyer, our student Victor Chahuneau, and I developed a nonparametric Bayesian approach that incorporates a (possibly incomplete) morphological analyzer into the prior distribution and learns improved language models for languages with rich morphology (e.g., Russian, Turkish, and Czech) [Chahuneau et al., 2013].

1.2.2 Supervised Learning for Linguistic Structure

Why is even *supervised* learning different for NLP than for other problems? The widespread use of approximate statistical inference (see, e.g., §1.1.1) in NLP opens up theoretical and practical questions about effects on learning algorithms (which typically make many calls to inference as a subroutine). I have explored a range of techniques that seek to understand this interaction [Martins et al., 2009b], more closely integrate inference and learning [Martins et al., 2008], avoid inference during learning [Smith et al., 2007], and parallelize across many processors [Gimpel et al., 2010].

My recent contributions have focused on learning algorithms that exploit better knowledge of the domain. In particular, we have explored regularizers that induce **structured sparsity**, such as the group lasso [Yuan and Lin, 2007]. This is a way of capturing a belief that the usefulness of features in a predictive model is best understood in groups, with the goal being to eliminate entire predefined groups of features deemed irrelevant. The result is a more compact model and a high-level inference about which groups matter. We have developed efficient, scalable online learning algorithms for a wide range of composite regularizers and loss functions that induce group sparsity [Martins et al., 2011d] and applied them to linguistic structured prediction problems including named entity recognition and dependency parsing [Martins et al., 2011c] (TurboParser served as a testbed; see §1.1.2). These methods are known as **online proximal gradient** algorithms. We have also used group sparsity in work on modeling demographic associations with lexical usage [Eisenstein et al., 2011] (see §1.1.4) and in sparse lexicon expansion to improve our frame-semantic analyzer, SEMAFOR [Das and Smith, 2012] (see §1.1.3). My collaborators and I gave a tutorial about structured sparsity at NAACL 2012 and published a general CS audience article about it [Smith and Martins, 2013].

1.2.3 Between Supervised and Unsupervised Learning

Latent discrete variables arise in many NLP problems; a simple variant can be found in topic models such as latent Dirichlet allocation [Blei et al., 2003]. We have explored how the basic idea of topic modeling can be applied to various ends by selecting certain kinds of words in text and informing the prior distribution. Examples include the discovery of character personas found across many films, using character mentions and syntactic contexts in film summaries [Bamman et al., 2013b], and the discovery of verbal frames expressing various states of international conflict or cooperation from a large news corpus about state-political actors [O'Connor et al., 2013].

A classic example of necessary latent structured variables is found in the word and phrase alignments between input and output strings in machine translation. Chris Dyer and I have developed state-of-the art word alignment methods, including a discriminative model that learns from unannotated parallel corpora [Dyer et al., 2011a] and a newly parameterized variant of IBM Model 2 that leads to improved translations [Dyer et al., 2013] compared to the strong baseline of IBM Model 4 [Brown et al., 1993]. These are available as open-source software within Chris' cdec library for machine translation.[6]

Translation learning is made even harder because, while the correct output is available to the learner, it may not be reachable by the inference algorithm, either due to the model family's limited expressive power or the approximate nature of inference. Kevin Gimpel and I developed **Rampion**, a learner that copes explicitly and robustly with these difficulties, and also builds on previous practical and theoretical work on the ramp loss [Gimpel and Smith, 2012a]. Rampion has been released as open-source software [Gimpel, 2012] and also integrated into cdec.

My group has encountered a range of problems requiring learning from incomplete data, developing algorithmic enhancements as required. Recent examples include:

- Arabic named entity recognition for Wikipedia data, a rather different genre than the news domain for which labeled data are available. We developed a new recall-oriented variant of self-training that exploits the ability of margin-based learning to use a task-specific cost function [Mohit et al., 2012].

[6] `http://www.cdec-decoder.org`

- Scenarios where we have strong performing NLP in one or a small number of languages and wish to improve performance for other languages that have no labeled data. One example uses helper languages' models through initialization of the new language's model, in unsupervised parsing [Cohen et al., 2011a]. Another uses imperfect machine translation into English and English WordNet-supersense tagging to improve supersense tagging for the other language, capitalizing on reasonably good content-word translation [Schneider et al., 2013a].
- Scenarios where we have reliable knowledge of the semantics of some words, but not all. We have used distributional similarity and graph-based learning to expand knowledge from the FrameNet lexicon to many more word types found in unannotated corpora, improving SEMAFOR's performance on unknown words by more than 13% absolute F_1 [Das and Smith, 2011].

1.3 Intelligent Applications

AI research must be motivated by concrete applications or risk falling into the cracks between theory and engineering. This does not mean that AI researchers need to *produce* commercial-strength implementations; I think that industry is usually far better suited to that role. Academics may be able to identify *new* applications and to discover new algorithms with a role to play in current and future applications; these are where my focus has been. My primary contributions to applications have been in machine translation (an old challenge) and text-driven forecasting (a new one).

1.3.1 Machine Translation

Since translation systems are complex, open-source toolkits have played a major role in enabling advances. I helped develop the first of these, Egypt/Giza [Al-Onaizan et al., 1999a,b], and members of my research group have continued this trend [Chahuneau et al., 2012b; Gimpel, 2009]. Data is also central; my early work contributed systems for finding translation data on the web [Resnik and Smith, 2003; Smith, 2002], and my group continues to contribute such data [Chahuneau, 2012].

My former Ph.D. student Kevin Gimpel developed a series of translation systems based on "loose" syntactic transformations, identifying an effective tradeoff between the use of linguistic analysis as a guide (§1.1) while still allowing flexibility [Gimpel and Smith, 2008, 2009b, 2011a, to appear]. Systems developed within my group, led by then-post-doctoral researcher Chris Dyer (now a faculty colleague), have won competitions: the 2011 Workshop on Statistical Machine Translation German-English human evaluation [Dyer et al., 2011b], and 2012 NIST Open Machine Translation Competition in Korean-English [Dyer et al., 2012]. We have also contributed to translation evaluation methodology, in collaboration with Alon Lavie and Jonathan Clark [Clark et al., 2011].

1.3.2 Text-Driven Forecasting

In text-driven *forecasting*, text inputs are used to make concrete predictions about future real-world events or measurements. My collaborators and I have developed regression models that take shallow text statistics as input, and shown that they can use text to make accurate predictions about financial risk [Kogan et al., 2009], opening-weekend revenues of films [Joshi et al., 2010a], scientific community response to articles [Yogatama et al., 2011], and food prices [Chahuneau et al., 2012a]. This work has led me to collaborations with social scientists, most notably financial economist Bryan Routledge, who participates in weekly meetings with my research group. In current work, we are exploring scalable models for temporal variation, led by my Ph.D. student Dani Yogatama [Yogatama et al., 2013], and new decision-theoretic language models that incorporate notions of utility in text production.

1.3.3 Inferences about Social Context

Much of my group's current research seeks to relate text content to the social context in which that text was created. An early example of this was a very simple text analysis method, developed primarily by Brendan O'Connor, for measuring public opinion about the economy and the president from Twitter messages, with high correlation to traditional telephone polls [O'Connor et al., 2010a]. More recently:

- My Ph.D. students David Bamman and Brendan O'Connor and I analyzed social media messages from China to characterize message deletion, finding evidence of political censorship [Bamman et al., 2012];
- David, Assyriologist Adam Anderson (Harvard), and I inferred social rank among merchants in a Middle Bronze Age trade colony in Old Assyria, from cuneiform tablets [Bamman et al., 2013a];
- Brendan, political scientist Brandon Stewart (Harvard), and I made inferences about international relations events from automatically parsed newstext [O'Connor et al., 2013];
- My Ph.D. students Tae Yano, Dani Yogatama, and I analyzed social media messages from U.S. Congressmen and related message content to campaign donations [Yano et al., 2013];
- Tae, political scientist John Wilkerson (UW), and I built predictive models of bill survival in Congressional committees [Yano et al., 2012];
- My Ph.D. student Yanchuan Sim, political scientists Justin Gross and Brice Acree (UNC Chapel Hill), and I developed a text analysis model to measure the ideological content in political speeches, and applied it to 2008 and 2012 U.S. presidential election speeches [Gross et al., 2013].

These projects vary greatly in the genre of data being explored, the sophistication in the statistical modeling techniques, and the underlying theoretical frameworks. My students, my collaborators, and I have been openly opportunistic about finding interesting contextualized text datasets to model. In doing so, we have learned a great deal about the aspirations of researchers in other disciplines to use text data, and we have established ourselves as leaders in applying NLP and machine learning to those ends.

The theme emerging from these efforts is that texts can answer substantive questions about their authors and those authors' communities. By allowing such questions to drive the development of computational models, we are beginning to identify building blocks that can be assembled to quickly answer new questions. In the 20th century, the field of statistics gave rise to such building blocks for quantitative data; we seek to do the same for text data.

1.4 Summary

My research interests are broad, but they have now proven to be mutually reinforcing. For example, André Martins', Dipanjan Das', and Kevin Gimpel's dissertations bring together principled machine learning with high-performance parsing, semantic analysis, and machine translation, respectively. Brendan O'Connor's latest work uses syntactic representations to make inferences about geopolitics [O'Connor et al., 2013]. My group is using its knowledge about building effective and efficient NLP to create useful tools for social media genres attracting intense commercial interest [Owoputi et al., 2013]. Our long-running investment in unsupervised NLP is starting to bear fruit in applications like machine translation [Gimpel and Smith, to appear]. What ties all of this work together is the use of text to make inferences: about what a particular text means, about the linguistic system and social context of its author, and about human societies more broadly.

2 Education

My educational goals are to lead a world-class research group with diverse skills and interests and whose alumni are recognized as leaders in the field; to help my departments lead in providing top-notch graduate and undergraduate instruction; and to promote scholarly exchange by regularly teaching outside CMU.

2.1 Advising

Of my twelve Ph.D. advisees to-date, five have completed the degree and gone on to research positions in academia (University of Edinburgh, Toyota Technological Institute at Chicago), non-profit (Educational Testing Service), and industry (Google Research, Priberam Labs); a sixth will finish this summer, and the other six are making steady progress. I meet with every student face-to-face virtually every week for technical discussion. I encourage active collaboration within the group and outside, as the publication record shows. My students have interned in research labs at Google, Microsoft, Amazon, and Facebook, and have made semester or summer-long visits to research groups at Princeton, Harvard, and the University of Southern California.

In addition to Ph.D.s, I have actively mentored more than twenty additional student researchers (Ph.D independent study students, visiting Ph.D. students, masters, undergraduates, recent undergraduates employed as research programmers, and a high school student). My mentees have gone on to Ph.D. programs at Stanford, MIT, CMU, and JHU. My weekly group meetings are lively, technical, constructively critical, collaborative, and designed to give students both depth and breadth. Many collaborations among students start there, and I work hard to balance intellectual stimulation with an atmosphere of collectivism and mutual support. Senior students typically mentor junior ones, and undergraduates and visitors are welcomed. I have aimed to build a diverse group, notably mentoring female researchers at Ph.D., masters, undergraduate, and high school levels.

2.2 Teaching

Undergraduate education. In 2008, on my own initiative, I designed an undergraduate course in NLP, giving a broad overview of the field. The course has been offered every spring since (five times by me, once by other faculty), and has grown to about 35 students (the most successful undergraduate offering in the LTI). Its most innovative aspect is an adversarial question answering project in which students are incentivized to transfer ideas from the lecture into working systems that are evaluated competitively in blind tests at the end of the course. In 2008 and 2010, the competition included teams from Rebecca Hwa's NLP course at the University of Pittsburgh. The project is described in an NSF workshop paper [Smith et al., 2008]. My former undergraduate students have reported back that this course project was unique in forcing them to work as a team to creatively solve an open-ended, hard problem. The course is crafted for undergraduates, with short-term, low-stakes assignments to encourage exploration and de-emphasize grades. Lectures are modular and come with one-page summaries to encourage active participation rather than rushing to take down notes.

Graduate education. A great deal of my teaching effort has been in keeping LTI graduate courses in NLP up to date with the latest research developments. I designed "Language and Statistics II" from scratch, and taught it for four years, then synthesized into a monograph [Smith, 2011]. In 2011, William Cohen and I worked together to merge that course with his "Information Extraction" course, since the area had matured and grown suitable for a wider audience; the result is "Structured Prediction," which I will teach for a second time this year with Chris Dyer. I have also taught or co-taught pre-existing graduate courses in the LTI and the Machine Learning Department ("Algorithms for NLP" twice and "Probabilistic Graphical Models" once). In 2009 I started an advanced reading seminar in NLP and have offered it every year since. In 2009 I offered an interdisciplinary special topics seminar ("Text-Driven Forecasting," both projects from which led to high-profile publications [Joshi et al., 2010a; O'Connor et al., 2010a]), and in 2013 revamped the LTI's hands-on NLP lab course into projects to develop open-source finite-state morphology tools for a range of languages.[7] My advanced courses typically have a term paper component with multiple rounds of feedback to help students improve their technical writing skills.

Tutorials outside CMU. Twice I have given tutorials at conferences (ICML 2009, NAACL 2012), and I regularly offer short courses or lectures at international summer schools. On invitation, I taught a short course at IBM's T.J. Watson Research Center.

Looking ahead. Apart from annual updates to content, my educational offerings at CMU have stabilized. A strong sign of this is that most of the courses I teach are mature enough to be taught by different instructors

[7] `http://www.cs.cmu.edu/~nasmith/NLPLab`

or teams of co-instructors. I continue to seek new, relevant material to incorporate (in existing classes and in classes I have not taught yet), and enjoy the intellectual growth that comes with the experience of teaching while broadening my expertise. I also aspire to opportunistically integrate research and teaching in ways that are mutually enhancing (e.g., course assignments where students grapple with the state of the art and its shortcomings, and generating data, as a by-product of teaching, that can support research efforts). One question on the horizon for me and for my institutions is whether and how to expand the reach of our teaching, through online offerings, MOOCs, and other vehicles. This is not a question I can answer on my own, though I do have immediate plans to publically release video of lectures from "Structured Prediction" this fall. I am open to greater disruption, though I have not yet decided whether to take on a visible leadership role in imagining the future of the university CS classroom. At the very least, I plan to continue to lead by innovating within my own courses and sharing my materials and experiences openly.

3 Service

My service goal is to maintain and incrementally improve the environments—local and global—that have contributed to my success and that of my students.

CMU. I have served (or am serving) on graduate admissions committees (seven years), faculty hiring committees (three times, once as chair), the LTI curriculum committee (four years), the university-wide undergraduate research fellowship committee (three years), and seventeen Ph.D. thesis committees (not including those of my own advisees). Since 2010 I have served as the faculty liaison to the LTI student body; in 2012 I organized the LTI faculty retreat. I served as the organizer for an existing seminar series in AI for two years, and recently started a new seminar series on Machine Learning and the Social Sciences. In addition to regularly hosting external speakers in various SCS colloquia, I have hosted representatives from companies seeking to hire in language technologies and machine learning (e.g., Thomson-Reuters and Amazon).

Research community. I have served (or am serving) on the editorial boards of several top journals in my field (*Computational Linguistics* (2009–2011), *Journal of Artificial Intelligence Research* (2011–present), and *Transactions of the Association for Computational Linguistics*) (2012–present); I review for others including *Journal of Machine Learning Research*, *IEEE Intelligent Systems*, and *Proceedings of the National Academy of Sciences*. I serve regularly on the program committees of the major NLP and machine learning conferences (ACL, NAACL, EMNLP, ICML, NIPS, and others), often as a senior member or in an organizational role, and on National Science Foundation review panels and advice-giving panels for aspiring NSF awardees. I am currently serving a two-year term as the secretary-treasurer of the SIGDAT special interest group within ACL, which sponsors the EMNLP conference. I have served or am serving as an external member of three Ph.D. thesis committees.

Data and software releases. NLP research relies heavily on sharing of datasets and implemented tools. When my research leads to reusable datasets, I have consistently made these available to the research community. Examples include forecasting datasets that link text documents to real-world measurements, like film reviews and revenue [Joshi et al., 2010b] and financial statements and stock market statistics [Routledge et al., 2009]. Data produced by small-scale annotation efforts conducted in my group are always made freely available to the research community. We have also released numerous open-source software tools; nearly every student in my group has released open-source datasets and/or tools, some of which have been downloaded hundreds of times and used in many others' research projects at CMU and elsewhere.

Outreach. I have frequently given presentations about my work to non-academic, non-research audiences. Recent examples include a panel organized by CMU's Center for Innovation and Entrepreneurship, another at the World Economic Forum in Davos, and invited presentations at technology conferences including South by Southwest, Tech@State (at the U.S. State Department), and Computer Assisted Reporting. I have contributed problems to the Computational Linguistics Olympiad, a high-school outreach program.

References

Y. Al-Onaizan, J. Cuřin, M. Jahr, K. Knight, J. Lafferty, I. D. Melamed, N. A. Smith, F.-J. Och, D. Purdy, and D. Yarowsky. Statistical machine translation. CLSP Research Notes 42, Johns Hopkins University, Baltimore, MD, 1999a. URL http://www.cs.cmu.edu/~nasmith/papers/alonaizan+etal.ws99.pdf.

Y. Al-Onaizan, J. Cuřin, M. Jahr, K. Knight, J. Lafferty, I. D. Melamed, N. A. Smith, F.-J. Och, D. Purdy, and D. Yarowsky. Egypt, 1999b. URL http://www.clsp.jhu.edu/ws99/projects/mt/toolkit.

W. Ammar, S. Wilson, N. Sadeh, and N. A. Smith. Automatic categorization of privacy policies: A pilot study. Technical Report CMU-LTI-12-019, Carnegie Mellon University, Pittsburgh, PA, December 2012. URL http://www.cs.cmu.edu/~nasmith/papers/ammar+wilson+sadeh+smith.tr12.pdf.

D. Bamman, B. O'Connor, and N. A. Smith. Censorship and content deletion in Chinese social media. *First Monday*, 17(3), March 2012. URL http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3943/3169.

D. Bamman, A. Anderson, and N. A. Smith. Inferring social rank in an Old Assyrian trade network, July 2013a. URL http://arxiv.org/abs/1303.2873.

D. Bamman, B. O'Connor, and N. A. Smith. Learning latent personas of film characters. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013b. URL http://www.cs.cmu.edu/~nasmith/papers/bamman+oconnor+smith.acl13.pdf.

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

P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19(2):263–311, 1993.

V. Chahuneau. Global Voices Malagasy-English parallel corpus, 2012. URL http://www.ark.cs.cmu.edu/global-voices.

V. Chahuneau, K. Gimpel, B. R. Routledge, L. Scherlis, and N. A. Smith. Word salad: Relating food prices and descriptions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*, Jeju, Korea, July 2012a. URL http://www.cs.cmu.edu/~nasmith/papers/chahuneau+gimpel+routledge+scherlis+smith.emnlp12.pdf.

V. Chahuneau, N. A. Smith, and C. Dyer. pycdec: A Python interface to cdec. *Prague Bulletin of Mathematical Linguistics*, 98:51–61, October 2012b. URL http://ufal.mff.cuni.cz/pbml/98/art-chahuneau-smith-dyer.pdf.

V. Chahuneau, N. A. Smith, and C. Dyer. Knowledge-rich morphological priors for Bayesian language models. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. URL http://www.cs.cmu.edu/~nasmith/papers/chahuneau+smith+dyer.naacl13.pdf.

M. Ciaramita and M. Johnson. Supersense tagging of unknown nouns in WordNet. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2003.

J. H. Clark, C. Dyer, A. Lavie, and N. A. Smith. Better hypothesis testing for statistical machine translation: Controlling for optimizer instability. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, Portland, OR, June 2011. URL http://www.cs.cmu.edu/~nasmith/papers/clark+dyer+lavie+smith.acl11.pdf.

S. B. Cohen and N. A. Smith. Joint morphological and syntactic disambiguation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 208–217, Prague, Czech Republic, June 2007. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+smith.emnlp07.pdf.

S. B. Cohen and N. A. Smith. Shared logistic normal distributions for soft parameter tying in unsupervised grammar induction. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 74–82, Boulder, CO, May/June 2009. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+smith.naacl09.pdf.

S. B. Cohen and N. A. Smith. Viterbi training for PCFGs: Hardness results and competitiveness of uniform initialization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1502–1511, Uppsala, Sweden, July 2010a. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+smith.acl10.pdf.

S. B. Cohen and N. A. Smith. Covariance in unsupervised learning of probabilistic grammars. *Journal of Machine*

Learning Research, 11:3017–3051, November 2010b. URL http://www.jmlr.org/papers/volume11/cohen10a/cohen10a.pdf.

S. B. Cohen and N. A. Smith. Empirical risk minimization with approximations of probabilistic grammars. In *Advances in Neural Information Processing Systems 23*, Vancouver, BC, December 2010c. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+smith.nips10.pdf.

S. B. Cohen and N. A. Smith. Empirical risk minimization for probabilistic grammars: Sample complexity and hardness of learning. *Computational Linguistics*, 38(3), September 2012. URL http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00092.

S. B. Cohen, K. Gimpel, and N. A. Smith. Logistic normal priors for unsupervised probabilistic grammar induction. In *Advances in Neural Information Processing Systems 21*, pages 321–328, Vancouver, BC, December 2008a. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+gimpel+smith.nips08.pdf.

S. B. Cohen, R. J. Simmons, and N. A. Smith. Dynamic programming algorithms as products of weighted logic programs. In *Proceedings of the International Conference on Logic Programming*, Udine, Italy, December 2008b. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+simmons+smith.iclp08.pdf. **Best student paper award.**

S. B. Cohen, D. Das, and N. A. Smith. Unsupervised structure prediction with non-parallel multilingual guidance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011a. URL http://www.cs.cmu.edu/~nasmith/papers/cohen+das+smith.emnlp11.pdf.

S. B. Cohen, R. J. Simmons, and N. A. Smith. Products of weighted logic programs. *Theory and Practice of Logic Programming*, 11(2–3):263–296, January 2011b. URL http://journals.cambridge.org/repo_A818QGnU.

D. Das and N. A. Smith. Paraphrase identification as probabilistic quasi-synchronous recognition. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 468–476, Singapore, August 2009. URL http://www.cs.cmu.edu/~nasmith/papers/das+smith.acl09.pdf.

D. Das and N. A. Smith. Semi-supervised frame-semantic parsing for unknown predicates. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011. URL http://www.cs.cmu.edu/~nasmith/papers/das+smith.acl11.pdf.

D. Das and N. A. Smith. Graph-based lexicon expansion with sparsity-inducing penalties. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012. URL http://www.cs.cmu.edu/~nasmith/papers/das+smith.naacl12.pdf.

D. Das, N. Schneider, D. Chen, and N. A. Smith. Probabilistic frame-semantic parsing. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010a. URL http://www.cs.cmu.edu/~nasmith/papers/das+schneider+chen+smith.naacl10.pdf.

D. Das, N. Schneider, D. Chen, and N. A. Smith. Semafor, 2010b. URL http://www.ark.cs.cmu.edu/SEMAFOR.

D. Das, A. F. T. Martins, and N. A. Smith. An exact dual decomposition algorithm for shallow semantic parsing with constraints. In *Proceedings of the Joint Conference on Lexical and Computational Semantics*, Montréal, Québec, June 2012. URL http://www.cs.cmu.edu/~nasmith/papers/das+martins+smith.starsem12.pdf.

D. Das, D. Chen, A. F. T. Martins, N. Schneider, and N. A. Smith. Frame-semantic parsing. *Computational Linguistics*, 40(1), March 2014. URL http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00163.

M. Dreyer, D. A. Smith, and N. A. Smith. Vine parsing and minimum risk reranking for speed and precision. In *Proceedings of the Conference on Natural Language Learning*, pages 201–205, New York, NY, June 2006. URL http://www.cs.cmu.edu/~nasmith/papers/dreyer+smith+smith.conll06.pdf.

C. Dyer, J. H. Clark, A. Lavie, and N. A. Smith. Unsupervised word alignment with arbitrary features. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011a. URL http://www.cs.cmu.edu/~nasmith/papers/dyer+clark+lavie+smith.acl11.pdf.

C. Dyer, K. Gimpel, J. H. Clark, and N. A. Smith. The CMU-ARK German-English translation system. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*, Edinburgh, UK, July 2011b. URL http://www.cs.cmu.edu/~nasmith/papers/dyer+gimpel+clark+smith.smt11.pdf.

C. Dyer, N. A. Smith, G. Morehead, P. Blunsom, and A. Levenberg. The CMU-Oxford translation system for the

NIST open machine translation 2012 evaluation, May 2012. URL http://www.cs.cmu.edu/~nasmith/papers/dyer+smith+morehead+blunsom+levenberg.nist12.pdf.

C. Dyer, V. Chahuneau, and N. A. Smith. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. URL http://www.cs.cmu.edu/~nasmith/papers/dyer+chahuneau+smith.naacl13.pdf.

J. Earley. An efficient context-free parsing algorithm. *Communications of the ACM*, 13(2):94–102, 1970.

J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing. A latent variable model for geographic lexical variation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, October 2010. URL http://www.cs.cmu.edu/~nasmith/papers/eisenstein+oconnor+smith+xing.emnlp10.pdf.

J. Eisenstein, N. A. Smith, and E. P. Xing. Discovering sociolinguistic associations with structured sparsity. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011. URL http://www.cs.cmu.edu/~nasmith/papers/eisenstein+smith+xing.acl11.pdf.

J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing. Mapping the geographical diffusion of new words. In *Proceedings of the NIPS Workshop on Social Network and Social Media Analysis: Methods, Models and Applications*, Lake Tahoe, NV, December 2012. URL http://arxiv.org/abs/1210.5268.

J. Eisner and N. A. Smith. Favor short dependencies: Parsing with soft and hard constraints on dependency length. In H. Bunt, P. Merlo, and J. Nivre, editors, *Trends in Parsing Technology: Dependency Parsing, Domain Adaptation, and Deep Parsing*, volume 43 of *Text, Speech, and Language Technology*, chapter 8, pages 121–150. Springer, January 2011. URL http://springerlink.metapress.com/content/k13158105hk43963.

J. Eisner, E. Goldlust, D. A. Smith, M. Dreyer, and N. A. Smith. Dyna, 2004a. URL http://www.dyna.org.

J. Eisner, E. Goldlust, and N. A. Smith. Dyna: A declarative language for implementing dynamic programs. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, pages 218–221, Barcelona, Spain, July 2004b. URL http://acl.ldc.upenn.edu/P/P04/P04-3032.pdf.

J. Eisner, E. Goldlust, and N. A. Smith. Compiling Comp Ling: Practical weighted dynamic programming and the Dyna language. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 281–290, Vancouver, BC, October 2005. URL http://acl.ldc.upenn.edu/H/H05/H05-1036.pdf.

C. J. Fillmore, C. R. Johnson, and M. R. Petruck. Background to FrameNet. *International Journal of Lexicography*, 16(3), 2003.

K. Gimpel. Quipu, 2009. URL http://www.ark.cs.cmu.edu/Quipu.

K. Gimpel. Rampion, 2012. URL http://www.ark.cs.cmu.edu/MT.

K. Gimpel and N. A. Smith. Rich source-side context for statistical machine translation. In *Proceedings of the ACL Workshop on Statistical Machine Translation*, pages 9–17, Columbus, OH, June 2008. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.smt08.pdf.

K. Gimpel and N. A. Smith. Cube summing, approximate inference with non-local features, and dynamic programming without semirings. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 157–166, Athens, Greece, March/April 2009a. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.eacl09.pdf.

K. Gimpel and N. A. Smith. Feature-rich translation by quasi-synchronous lattice parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 219–228, Singapore, August 2009b. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.emnlp09.pdf.

K. Gimpel and N. A. Smith. Quasi-synchronous phrase dependency grammars for machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011a. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.emnlp11.pdf.

K. Gimpel and N. A. Smith. Generative models of monolingual and bilingual gappy patterns. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*, Edinburgh, UK, July 2011b. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.smt11.pdf.

K. Gimpel and N. A. Smith. Structured ramp loss minimization for machine translation. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012a. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.naacl12.pdf.

K. Gimpel and N. A. Smith. Concavity and initialization for unsupervised dependency parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012b. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+smith.naacl12b.pdf.

K. Gimpel and N. A. Smith. Phrase dependency machine translation with quasi-synchronous tree-to-tree features. *Computational Linguistics*, to appear.

K. Gimpel, D. Das, and N. A. Smith. Distributed asynchronous online learning for natural language processing. In *Proceedings of the Conference on Computational Natural Language Learning*, Uppsala, Sweden, July 2010. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+das+smith.conll10.pdf.

K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, Portland, OR, June 2011. URL http://www.cs.cmu.edu/~nasmith/papers/gimpel+etal.acl11.pdf.

A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *Neural Information Processing Systems 20*, 2008.

J. H. Gross, B. Acree, Y. Sim, and N. A. Smith. Testing the etch-a-sketch hypothesis: A computational analysis of Mitt Romney's ideological makeover during the 2012 primary vs. general elections, September/August 2013.

M. Heilman and N. A. Smith. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. URL http://www.cs.cmu.edu/~nasmith/papers/heilman+smith.naacl10b.pdf.

V. Jojic, S. Gould, and D. Koller. Accelerated dual decomposition for MAP inference. In *Proceedings of the International Conference of Machine Learning*, 2010.

M. Joshi, D. Das, K. Gimpel, and N. A. Smith. Movie reviews and revenues: An experiment in text regression. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010a. URL http://www.cs.cmu.edu/~nasmith/papers/joshi+das+gimpel+smith.naacl10.pdf.

M. Joshi, D. Das, K. Gimpel, and N. A. Smith. Movie$ corpus, 2010b. URL http://www.ark.cs.cmu.edu/movie$-data.

S. Kogan, D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith. Predicting risk from financial reports with regression. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 272–280, Boulder, CO, May/June 2009. URL http://www.cs.cmu.edu/~nasmith/papers/kogan+levin+routledge+sagi+smith.naacl09.pdf.

N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *International Conference on Computer Vision*, 2007.

L. Kong and N. A. Smith. Stanford dependencies without constituent parsing. 2013.

A. F. T. Martins. TurboParser, 2009. URL http://www.ark.cs.cmu.edu/TurboParser.

A. F. T. Martins. AD^3, 2012. URL http://www.ark.cs.cmu.edu/AD3.

A. F. T. Martins, D. Das, N. A. Smith, and E. P. Xing. Stacking dependency parsers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 157–166, Waikiki, HI, October 2008. URL http://www.cs.cmu.edu/~nasmith/papers/martins+das+smith+xing.emnlp08.pdf.

A. F. T. Martins, N. A. Smith, and E. P. Xing. Concise integer linear programming formulations for dependency parsing. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 342–350, Singapore, August 2009a. URL http://www.cs.cmu.edu/~nasmith/papers/martins+smith+xing.acl09.pdf. **Best paper award.**

A. F. T. Martins, N. A. Smith, and E. P. Xing. Polyhedral outer approximations with application to natural language parsing. In *Proceedings of the International Conference on Machine Learning*, pages 713–720, Montréal, Québec, June 2009b. URL http://www.cs.cmu.edu/~nasmith/papers/martins+smith+xing.icml09.pdf.

A. F. T. Martins, N. A. Smith, E. P. Xing, P. M. Q. Aguiar, and M. A. T. Figueiredo. Turbo parsers: Dependency parsing by approximate variational inference. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, October 2010. URL http://www.cs.cmu.edu/~nasmith/papers/martins+etal.emnlp10.pdf.

A. F. T. Martins, P. M. Q. Aguiar, M. A. T. Figueiredo, N. A. Smith, and E. P. Xing. An augmented Lagrangian approach to constrained MAP inference. In *Proceedings of the International Conference on Machine Learning*, Bellevue, WA, June/July 2011a. URL http://www.cs.cmu.edu/~nasmith/papers/martins+figueiredo+

aguiar+smith+xing.icml11.pdf.

A. F. T. Martins, N. A. Smith, P. M. Q. Aguiar, and M. A. T. Figueiredo. Dual decomposition with many overlapping components. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011b. URL http://www.cs.cmu.edu/~nasmith/papers/martins+smith+aguiar+figueiredo.emnlp11.pdf.

A. F. T. Martins, N. A. Smith, P. M. Q. Aguiar, and M. A. T. Figueiredo. Structured sparsity in structured prediction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011c. URL http://www.cs.cmu.edu/~nasmith/papers/martins+smith+aguiar+figueiredo.emnlp11b.pdf.

A. F. T. Martins, N. A. Smith, E. P. Xing, P. M. Q. Aguiar, and M. A. T. Figueiredo. Online learning of structured predictors with multiple kernels. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, April 2011d. URL http://www.cs.cmu.edu/~nasmith/papers/martins+smith+xing+aguiar+figueiredo.aistats11.pdf.

A. F. T. Martins, M. Almeida, and N. A. Smith. Turning on the turbo: Fast third-order non-projective turbo parsers. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013. URL http://www.cs.cmu.edu/~nasmith/papers/martins+almeida+smith.acl13.pdf.

A. F. T. Martins, M. A. T. Figueiredo, P. M. Q. Aguiar, N. A. Smith, and E. P. Xing. AD^3: Alternating directions dual decomposition for MAP inference in graphical models. *Journal of Machine Learning Research*, in review. URL http://arxiv.org/abs/1212.6550.

B. Mohit, N. Schneider, R. Bhowmick, K. Oflazer, and N. A. Smith. Recall-oriented learning of named entities in Arabic Wikipedia. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France, April 2012. URL http://www.cs.cmu.edu/~nasmith/papers/mohit+schneider+bhowmick+oflazer+smith.eacl12.pdf.

B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 122–129, Washington, DC, May 2010a. URL http://www.cs.cmu.edu/~nasmith/papers/oconnor+balasubramanyan+routledge+smith.icwsm10.pdf.

B. O'Connor, J. Eisenstein, E. P. Xing, and N. A. Smith. Discovering demographic language variation. In *Proceedings of the NIPS Workshop on Machine Learning for Social Computing*, Whistler, BC, December 2010b. URL http://www.cs.cmu.edu/~nasmith/papers/oconnor+eisenstein+xing+smith.nips-ws10.pdf.

B. O'Connor, B. Stewart, and N. A. Smith. Learning to extract international relations from political context. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013. URL http://brenocon.com/oconnor+stewart+smith.irevents.acl2013.pdf.

O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. URL http://www.cs.cmu.edu/~nasmith/papers/owoputi+oconnor+dyer+gimpel+schneider+smith.naacl13.pdf.

P. Resnik and N. A. Smith. The Web as a parallel corpus. *Computational Linguistics*, 29(3):349–380, September 2003. URL http://acl.ldc.upenn.edu/J/J03/J03-3002.pdf.

B. R. Routledge, S. Kogan, J. S. Sagi, and N. A. Smith. 10-K corpus, 2009. URL http://www.ark.cs.cmu.edu/10K.

A. M. Rush and S. Petrov. Vine pruning for efficient multi-pass dependency parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, 2012.

N. Schneider, B. Mohit, K. Oflazer, and N. A. Smith. Coarse lexical semantic annotation with supersenses: An Arabic case study. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Jeju, Korea, July 2012. URL http://www.cs.cmu.edu/~nasmith/papers/schneider+mohit+oflazer+smith.acl12.pdf.

N. Schneider, B. Mohit, C. Dyer, K. Oflazer, and N. A. Smith. Supersense tagging for Arabic: the MT-in-the-middle attack. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013a. URL http://www.cs.cmu.edu/~nasmith/papers/schneider+mohit+dyer+oflazer+smith.naacl13.pdf.

N. Schneider, B. O'Connor, N. Saphra, D. Bamman, M. Faruqui, J. Baldridge, N. A. Smith, and C. Dyer. A framework for (under)specifying dependency syntax without overloading annotators. In *Proceedings of the ACL Linguistic Annotation Workshop*, Sofia, Bulgaria, August 2013b. URL http://www.cs.cmu.edu/~nasmith/papers/

schneider+oconnor+saphra+bamman+faruqui+smith+dyer+baldridge.law13.pdf.

D. A. Smith and N. A. Smith. Bilingual parsing with factored estimation: Using English to parse Korean. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 49–56, Barcelona, Spain, July 2004. URL http://www.cs.cmu.edu/~nasmith/papers/smith+smith.emnlp04.pdf.

D. A. Smith and N. A. Smith. Probabilistic models of nonprojective dependency trees. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 132–140, Prague, Czech Republic, June 2007. URL http://www.cs.cmu.edu/~nasmith/papers/smith+smith.emnlp07.pdf.

N. A. Smith. From words to corpora: Recognizing translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 95–102, Philadelphia, PA, July 2002. URL http://acl.ldc.upenn.edu/W/W02/W02-1013.pdf.

N. A. Smith. *Novel Estimation Methods for Unsupervised Discovery of Latent Structure in Natural Language Text*. PhD thesis, Department of Computer Science, Johns Hopkins University, Baltimore, MD, October 2006. URL http://www.cs.cmu.edu/~nasmith/papers/smith.thesis06.pdf. Supervised by Jason Eisner.

N. A. Smith. *Linguistic Structure Prediction*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool, May 2011. URL http://www.cs.cmu.edu/~nasmith/LSP.

N. A. Smith and J. Eisner. Annealing techniques for unsupervised statistical language learning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 487–494, Barcelona, Spain, July 2004. URL http://acl.ldc.upenn.edu/P/P04/P04-1062.pdf.

N. A. Smith and J. Eisner. Contrastive estimation: Training log-linear models on unlabeled data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 354–362, Ann Arbor, MI, June 2005a. URL http://acl.ldc.upenn.edu/P/P05/P05-1044.pdf.
Nominated for best paper award.

N. A. Smith and J. Eisner. Guiding unsupervised grammar induction using contrastive estimation. In *Proceedings of the IJCAI Workshop on Grammatical Inference Applications*, pages 73–82, Edinburgh, UK, July 2005b. URL http://www.cs.cmu.edu/~nasmith/papers/smith+eisner.ijcaigia05.pdf.

N. A. Smith and J. Eisner. Annealing structural bias in multilingual weighted grammar induction. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics*, pages 569–576, Sydney, Australia, July 2006. URL http://acl.ldc.upenn.edu/P/P06/P06-1072.pdf.

N. A. Smith and M. Johnson. Weighted and probabilistic context-free grammars are equally expressive. *Computational Linguistics*, 33(4):477–491, December 2007. URL http://www.cs.cmu.edu/~nasmith/papers/smith+johnson.cl07.pdf.

N. A. Smith and A. F. T. Martins. Linguistic structure prediction with the sparseptron. *ACM Crossroads*, 19(3):44–48, April 2013. URL http://www.cs.cmu.edu/~nasmith/papers/smith+martins.xrds13.pdf.

N. A. Smith, D. A. Smith, and R. W. Tromble. Context-based morphological disambiguation with random fields. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 475–482, Vancouver, BC, October 2005. URL http://acl.ldc.upenn.edu/H/H05/H05-1060.pdf.

N. A. Smith, D. L. Vail, and J. D. Lafferty. Computationally efficient M-estimation of log-linear structure models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 752–759, Prague, Czech Republic, June 2007. URL http://www.cs.cmu.edu/~nasmith/papers/smith+vail+lafferty.acl07.pdf.

N. A. Smith, M. Heilman, and R. Hwa. Question generation as a competitive undergraduate course project. In *Proceedings of the NSF Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington, VA, September 2008. URL http://www.cs.cmu.edu/~nasmith/papers/smith+heilman+hwa.nsf08.pdf.

D. Sontag, A. Globerson, and T. Jaakkola. Introduction to dual decomposition for inference. In *Optimization for Machine Learning*. MIT Press, 2011.

A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967.

T. Yano, N. A. Smith, and J. D. Wilkerson. Textual predictors of bill survival in Congressional committees. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 793–802, Montréal, Québec, June 2012. URL http://www.cs.cmu.edu/~nasmith/papers/yano+smith+

wilkerson.naacl12.pdf.

T. Yano, D. Yogatama, and N. A. Smith. A penny for your tweets: Campaign contributions and Capitol Hill microblogs. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Boston, MA, July 2013. URL http://www.cs.cmu.edu/~nasmith/papers/yano+yogatama+smith.icwsm13.pdf.

D. Yogatama, M. Heilman, B. O'Connor, C. Dyer, B. R. Routledge, and N. A. Smith. Predicting a scientific community's response to an article. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. URL http://www.cs.cmu.edu/~nasmith/papers/yogatama+heilman+oconnor+dyer+routledge+smith.emnlp11.pdf.

D. Yogatama, B. R. Routledge, and N. A. Smith. A sparse and adaptive prior for time-dependent model parameters. 2013.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68(1):49–67, 2007.