This is an exciting time in clinical natural language processing (NLP). Many of the ideas discussed in the NLP community over the last decade have finally come to fruition: we have had shared tasks in the clinical arena [1, 2]; several NLP pipelines have been released as open source resources [3, 4]; the community is working together to develop standards that will increase interoperability of NLP modules; and de-identified repositories of clinical notes are being made available for NLP research [5, 6]. These advances will enable quicker development of NLP tools, greater collaboration among NLP researchers through shared resources and standards, and expanded opportunities to implement NLP in real clinical, public health, and research settings.

In spite of published success in limited domains and the promise that lies ahead, to this point NLP has had modest impact on patient care. Congruent to the question Ball and colleagues [7] asked in Failure to Provide Clinicians Useful IT Systems (“What is responsible for the low adoption of HIT in physician practices?”) we should ask ourselves “What is responsible for the nominal integration of NLP in clinical informatics applications?” Over the next decade, I believe we need to expand our focus from the goal of achieving human-like interpretation of text to the goal of providing value to users. Rather than aiming to replace humans at a task that is expensive or tedious to perform, our goal should be to assist humans in performing their work more efficiently or more accurately. To do this, we need to involve end users earlier in the development process, to identify the types of annotations that are most effective for supporting them in their tasks, and to direct more of our research toward usability of NLP applications.

Five articles in this journal investigate the interface of humans and applications that integrate NLP or structured knowledge representation. In the paper Integration of Relational and Textual Biomedical Sources: A Pilot Experiment Using a Semi-automated Method for Logical Schema Acquisition, Garcia-Remesal and colleagues [8] tackle the problem of representing knowledge in a way that enables users to formulate queries over databases containing structured and textual data without requiring them to have deep domain knowledge or specialized understanding of how the knowledge is represented in the database.

Interface terminologies support interactions between humans and structured medical information systems. Creating an interface terminology that assists users in direct entry of detailed structured information requires balancing the terminology’s comprehensiveness with its usability. In Construction of an Interface Terminology on SNOMED-CT: Generic Approach and its Application in Intensive Care, Bakhshi-Raiez and colleagues [9] describe a methodology for creating a terminology by extracting a relevant subset of an existing terminology in the domain of interest and extending the subset with local concepts, relationships, and terms. They illustrate their method by creating an interface terminology in Dutch.

Hasegawa and colleagues [10] developed an interface terminology in Japanese, which they employ in an application for capturing a radiologist’s observations and interpretations of a chest x-ray in struc-
tured form in the article Development of a System that Generates Structured Reports for Chest X-ray Radiography. They showed that radiologists maintained comprehensive-ness without increasing the time they normally spend on dictation.

In Chi-square-based Scoring Function for Categorization of MEDLINE Citations, Kastrin and colleagues [11] addressed the usability challenge of integrating a classifier into an existing framework without requiring long computational times and tedious manual preparation of datasets, both of which can discourage implementation of text classifiers.

McKinlay and colleagues [12] explored the interface of humans and the output of a natural language generation system in Design Issues for Socially Intelligent User Interfaces: A Discourse Analysis of a Data-to-text System for Summarizing Clinical Data. Through discourse analysis of automatically and manually generated text summaries of detailed clinical data, they identified characteristics of coherency found in human-generated text that can guide creation of more natural output of the summarization system.

I hope the next decade will be ripe with studies like these – studies that integrate NLP and related methodologies within existing applications with the aim of helping enter, summarize, and interpret the volumes of data available, studies that lead towards clinical application rather than toward a small increase in technical performance. Like attempts at translating evidence into practice, we should consider how best to integrate the methodologies and tools we develop into clinical care, and it is encouraging to see that integration of NLP and related methodologies in clinical care is becoming a global question.

Creating applications of value will take more than expertise in NLP development and more than getting end users involved in the development process. Just as there is a gap between the accumulation of evidence and the translation of that evidence in clinical practice, there is a gap between development of NLP-driven informatics tools and translation of those tools in clinical practice.

One potential reason for this gap is that in today’s environment of academic capitalism it can be challenging to find a business model for the sustained funding that is necessary to successfully translate methodologies into applications that are both motivated by and adaptive to the needs of clinical users. With the expanding attention to translating evidence into practice that is backed by large initiatives from funding agencies, one could argue that in theory, at least, the business model is emerging. However, research funding agencies may never provide the support for software development and maintenance needed to integrate NLP in the clinical environment. To fill the gap between NLP research and translation, we should actively pursue more collaboration with companies developing electronic medical record systems or performing text mining and coding. NLP researchers can provide awareness of existing standards, insight into the linguistic challenges of modeling clinical information, and innovation of new methodologies; industry can provide excellence in software development, success in deploying tools, and a pulse on the market. Building bridges between NLP research and industry could create the type of symbiotic relationship that will contribute to translation of NLP in the clinical environment. We need more thought and discussion about working models of collaborative relationships between industry and academics in the context of NLP and knowledge representation.

Another potential reason for the adoption gap is that much of our research is designed and carried out without full consideration of users’ needs and of the environment in which the NLP applications will be applied. It may be wise for researchers in our domain to consider the advice of Sarin [16] in Inverse planning for the T1 – T2 conundrum in translational research: Let’s reconsider our methodological approaches in the light of their ability to add value and translate in the clinical setting. Careful study of the clinical setting may alter the directions we pursue. For example, many research studies have focused on creating text classifiers that categorize documents into a single category, such as “consistent with acute bacterial pneumonia” or “symptoms consistent with an influenza-like illness”. Document-level classification can be useful for many tasks, including hospital or community surveillance and finding patients for epidemiological or comparative effectiveness studies. Further investigation of the clinical setting in which these tasks are performed reveals, though, that the document-level classification may not add substantive value to users. In a chart review, for instance, the abstractor wading through the medical record to identify patients meeting a case definition may review narrative documents from a variety of sources. Many of the documents may be accessible for text classification, but some of the documents may be inaccessible, such as reports stored in a proprietary system or notes that are recorded on paper. For chart review, document-level classifications of a portion of the patient record would provide minimal value to the chart abstractor. Instead, NLP techniques for organizing and ultimately visualizing the electronic information for more efficient review could provide great value. Inverse planning – considering our methodological approaches in light of their ability to add value and translate in the clinical setting – can guide our future efforts so that we develop applications that truly provide value to consumers.

References

*Expanded on by Robinson [13], academic capitalism has been defined by Slaughter and Leslie [14] as “institutional and professorial market or mar-
ket: ‘like efforts to secure external money’” (p 8). In the twenty-first century, the term was modified to include “the new economy” [15] and took on the meaning of “a regime that entails colleges and universities engaging in market and market-like behaviors” (p 37).