Preface

APPLIED NATURAL LANGUAGE PROCESSING

Applied Natural Language Processing (ANLP) is an emerging field of study concerned with how computational approaches can assist with the identification, investigation, and resolution of real-life language-related issues. The NLP part of ANLP is predominantly (but not exclusively) the domain of computer scientists. It is they who are responsible for most (but not all) of the advancements in textual analysis tools and approaches. The A part of ANLP is predominantly (but not exclusively) the domain of cognitive psychologists and linguists. It is they who predominantly (but not exclusively) apply NLP to linguistic data with the goal of increasing our knowledge of how the mind represents and retrieves knowledge, increasing our ability to mimic human intelligence, and/or increasing our ability to assess and describe how language impacts the world and the individuals and groups within it.

We label ANLP an "emerging" field because it is not yet clear whether it is sufficiently focused to draw in researchers under its own gravity. Thus, ANLP could be described as a "field" because, like other fields, it produces knowledge and establishes practices that can be taught and researched. But at the same time, ANLP may equally be described as simply a convenient bucket into which many pieces of otherwise homeless research are dropped. That is, a great many studies simply *end up* as ANLP, while the studies' researchers would not label themselves as members of the field of ANLP. Perhaps this scenario is to be expected from interdisciplinary studies of real world problems, and therefore ANLP will always be (largely) a field in which we graze rather than sow.

ANLP may well be an emerging field, but if it is to lose its modifier, it has to begin forming a recognized identity. With this in mind, we can admit that there is clearly much to be done: There is terminology to be agreed; there are prototypical topics to be established; there are seminal works to be sanctified, and there is a form, a voice, and discourse move that need to coalesce. Of course, all these aspects of any field come largely as a result of convention, and conventions take time. We hope that this book represents a suitable point of departure for such conventions, and that it at least provides current researchers with some guidelines from which to begin, some framework within which to work, and some goals for which to strive.

NLP AND ANLP

The amount of information that humans have gathered and made available to other humans is, of course, phenomenal. And however large this repository of knowledge is, we know that by this time tomorrow, it will be larger still. But perhaps what is most relevant to us about this information is that most of it ap-

pears in textual form, and that if we are ever to manage it, understand it, assess it, evaluate it, summarize it, or even find it, then a broad range of natural language processing tools, systems, algorithms, models, theories, and techniques will be needed. The fields of Natural Language Processing (NLP) and Applied Natural Language Processing (ANLP) are both dedicated to this venture. But while their goals are highly overlapping (as is much of their research), their contribution to those goals is quite distinguishable.

The field of NLP is concerned with the development of natural language processing approaches (i.e., tools, systems, algorithms, models, theories, and techniques). More specifically, it is concerned with how these approaches are applied to a fairly well established set of tasks (e.g., summarization, part-of-speech tagging, named entity recognition, co-reference resolution, natural language understanding, text-type disambiguation, and so forth) that have fairly well established methods of appraisal (e.g., compare recall, precision, F1 to previously tested systems) and fairly well established sets of data (e.g., corpora such as the Wall Street Journal or the Microsoft Paraphrase Corpus). This having been said, we should not think of such tasks as trivial or esoteric. Instead, NLP might be thought of as the laboratory, the prototypes, and the testing ground.

NLP can be said to have become ANLP when the focus of the research shifts away from honing the accuracy and validity of the NLP approach to adapting the technology wholesale to a real world situation. Thus, a prototypical example of NLP might be described as Vasile Rus' development of a lexico-syntactic approach to entailment assessment (Rus et al., 2008); whereas a prototypical example of *A*NLP might be described as Vasile Rus' using that approach to assess paraphrase evaluation in an Intelligent Tutoring System (Rus et al., in press).

But of course, research is more than simply *using* something. ANLP is concerned with how those approaches stack up against new problems, issues, identified knowledge gaps, or real world based data sets. In many ways then, ANLP can be distinguished from NLP not so much by its content, form, or span, but by its focus. This change in focus results in research where less time and attention is spent concerned with the approach, which has presumably been described elsewhere, as it is spent concerned with the issue, the investigation, and the resolution. This is not to say that the mechanics of the approach can be ignored (they cannot), but it is to say that the mechanics are relegated to being, as it were, a guest at the party, as opposed to the host.

Given the nature of ANLP, it is often an *X-solution* applied to a *Y-problem*. As such, ANLP can often be a quick and "sufficient" answer, even while it may be a far from perfect one. For example, latent semantic analysis (see Kintsch and Kintsch, this volume) was not designed to assess feedback for paraphrase evaluations anymore than it was designed to be the foundation stone of dialogue management in intelligent tutoring systems. Yet McCarthy, Guess, and McNamara (2009) identified feedback for paraphrase evaluation problems and successfully used latent semantic analysis (LSA) to resolve them. Similarly, researchers of intelligent systems (see AutoTutor: this volume; iSTART: this volume) have also implemented LSA to verify dialogue. To be sure, LSA results have ranged from extremely encouraging (McNamara et al. 2007) to quite problematic (McCarthy et al., 2007, 2008). Thus, one key element of ANLP research is establishing the degree to which an approach works, and the identification of which elements in that research need to be addressed to make the approach more than merely "sufficient." This identification of a partial solution along with its limitations may often result in the later development of hybrid approaches, as with paraphrase evaluation through a combination of LSA and syntactic assessment (McCarthy et al., 2009) or as with introducing *entailment evaluations* to dialogue assessment in combination with LSA (Rus et al., 2008). Thus, ANLP research does not *have to* be viewed as solely a solution; it is often a journey, often a treatment, often a diagnostic, often a finger in the damn till help arrives.

Although the focus of ANLP might contrast with NLP, the areas of interest do not: anywhere NLP goes, *A*NLP must surely follow (and often arrive first). Thus, the topics of interest for ANLP include (but, by definition, are not limited to) summarization, text mining, categorization, authorship recognition, genre recognition, word sense disambiguation, first/second language acquisition, text and discourse analysis, paraphrasing, entailment, anaphora resolution, co-reference resolution, text cohesion and coherence, dialogue management and systems, language generation, language models, human computer interfaces, multilingual processing, standardization issues, language resources, corpora, learning environments, semantics, ontologies, machine translation, intelligent tutoring, question answering, parsing, tagging, annotating, tokenization, morphology, stemming, information extraction, syntax, English for specific purposes, humor analysis, user language understanding and assessment, web assessment, blog analysis, grammar checking, speech recognition, speech production, data mining, and any and all other areas that involve computation and text.

A BRIEF HISTORY OF ANLP

If ANLP is an emerging field, then we must describe from where it has emerged, and from where it continues to be emerging. Not surprisingly, we find that the history of ANLP is closely tied to NLP and, more specifically, at various times, in various ways, to offshoots from NLP into real world issues. These offshoots have met with varying degrees of success (in some ways, perhaps too much success), but as we will see from the history described below, it is hard to ignore the fact that there is no shortage of interest in the activities of ANLP.

In 1983, a series of ANLP conferences began and continued until 2000. The ANLP series grew out of the Association for Computational Linguistics (ACL) conference. More specifically, the series was an outgrowth of a 1981 ACL workshop. The ACL conference at that time was small, met yearly, and was somewhat similar to the International Conference on Computational Linguistics (COLING), which met bi-yearly. An ANLP conference was proposed as an alternate venue for papers where there could be discussions of the role of natural language processing in solving real world problems.

The first ANLP conference was held in Santa Monica in 1983. It featured 26 papers across six tracks: 1) domain-independent natural language interfaces, 2) knowledge-based approaches, 3) handling ill-formed input, 4) text analysis, 5) machine translation, and 6) speech interfaces. The third conference went international, being held in Trento, Italy. Although the conferences would never grow much larger than their initial numbers, they did expand each meeting so that for the sixth and final conference in 2000, the program committee chose from 131 submissions received from 24 different countries.

The proceedings from these conferences display an increasing focus on the interaction of technology and the market. As technology advanced, business and government enterprises were better able to use NLP techniques to resolve, or at least investigate, their problems. The advent of digital communication expanded both NLP technologies and the ability of the conference organizers to get submissions and attendees. By 2000, nearly a third of the submissions represented business, private interests, or government, rather than academic sources. This success would prove problematic, however, with increasing amounts of ANLP research and tools becoming proprietary.

The conferences' success, in terms of interest, would eventually lead to its early retirement. Research that had been brought to light from the previous six ANLP conferences now meant that more established conferences were welcoming, and even expecting papers with direct applications to real world problems.

With ACL and COLING conferences being held regularly, and ANLP conferences with gaps as long as four years, researchers could hardly be blamed for looking elsewhere for venues. But in order to accommodate the requirements of these NLP conferences, ANLP research needed to become more empirical, which ultimately led to ANLP blurring back into NLP.

Although the ANLP conferences were now a thing of the past, interest in ANLP certainly was not. In 2006, Vasile Rus introduced a track to the International Florida Artificial Intelligence Research Society (FLAIRS) focused on research and tools concerned with the understanding, organizing, and mining of text based information. Despite being a new track, it received a lot of attention, gaining 19 submissions (about 10% of the over-all track submissions), 8 of which were accepted as papers. Vasile had correctly detected a revived interest in ANLP. He also saw that this interest was in step with new developments in intelligent tutoring systems, such as AutoTutor and iSTART (see Chapters 11 and 16 this book). These systems require the development of specialized algorithms and assessment approaches in order that they can provide suitable feedback to users. In other words, NLP was needed to solve real world problems. In yet other words, *A*NLP was needed.

In 2007, Christian Hempelman and Phil McCarthy took over Vasile's track and renamed it Applied Natural Language Processing (ANLP), the name it still has today. The interest from 2006 was maintained in 2007, and grew steadily through 2008 and 2009 under the direction of Phil McCarthy and Scott Crossly. In 2010, Phil McCarthy was joined by Chutima Boonthum-Denecke. By this time, Phil McCarthy was chairing the FLAIRS program itself, and Chutima Boonthum-Denecke was the special tracks chair of FLAIRS, so Vasile Rus stepped back into the leadership role of the ANLP track together with Mihai Lintean. Their promotion of the track led to the most successful year for ANLP to date: 19 accepted papers, a workshop, a demonstration session, and a special track invited guest. In fact, by this time, ANLP was receiving more submissions and producing more talks than the conference main track.

The role of FLAIRS as the stage for the emerging field of ANLP is undeniable. But if it could be said that there was any single person or place that was the driving force being the products that FLAIRS put on show, then that person would be Danielle McNamara, and that place would be the Institute for Intelligent Systems (IIS) at the University of Memphis. All of the researchers that chaired the ANLP track at FLAIRS passed through the IIS at some stage of their careers, and each of them have also worked with Danielle McNamara on at least one of her projects. As of 2011, Danielle (now director of the IIS) had co-authored 27 FLAIRS publications, and in 2007, she was the first invited speaker of the track. Her main contribution to the field was the Coh-Metrix text analysis tool (see Chapter 11). Coh-Metrix was the first free, widely available software of its kind, allowing researchers to process large numbers of text to assess such metrics as cohesion, readability, lexical diversity, frequency, semantic overlap, and numerous others. In short, Coh-Metrix was (and arguably still is) the ultimate ANLP tool. Danielle would also contribute to FLAIRS and science in general (especially cognitive science) with her intelligent tutoring systems (iSTART and Wpal: see chapters 15 and 17). She also plays a part in the development of other systems such as AutoTutor (see Chapter 10) and other assessment approaches (see entailment in Chapter 7). Each of these projects has also appeared at FLAIRS. In sum, the field of ANLP owes a great debt of gratitude to Danielle McNamara and the Institute of Intelligent Systems.

The burgeoning interest in ANLP led directly to this book. The design of the book was such that the leading names and most notable achievements in ANLP could be brought together so that the *emerging* field might sooner become *emerged*. But the book's purview was not simply to compile what existed; it was also to draw in new researchers, especially those whose work had often been seen as merely straddling the boundaries of conventional fields. The book was also designed with students in mind. Thus, it

had to be accessible enough to be integrated into courses as a main or supplemental course book, relevant to graduate students and advanced under-graduates. Because ANLP is inherently inter-disciplinary, the book also had to be sufficiently diverse to accommodate departments of Computer Science, Cognitive Science, and Linguistics, and yet at the same time to be cohesive enough to bring researchers and students from these departments together.

To what degree this book has successfully achieved its goals will be determined further along the road. However, that its goals are realistic is evidenced by the breadth of researchers who have made contributions to it. Indeed, one of the editors is a linguist, the other is a computer scientist, and the researcher whose name appears most often in this book (Danielle McNamara) is a cognitive scientist. As for the book being embraced in the classroom, we point to Hearst (2005), who argued that there is much that can and needs to be taught in ANLP, but that there is no suitable text for such a course. This problem, at least, we hope we have addressed here.

ORGANIZATION OF THE BOOK

Although NLP might seem to be able to get along without ANLP, the reverse is a more difficult case to make. For this reason, Section 1 of this book focuses on foundational sub-fields of NLP. Of course, it is impossible to cover all sub-fields of NLP (even if such a list were possible), therefore, we offer in Section 1 seven chapters that perhaps speak most closely to issues that arise in ANLP. An eighth chapter in Section 1 directly addresses an issue highlighted by Hearst (2005) in her paper on teaching ANLP: the need for a guide to practical programming.

Section 2 focuses on successful systems and approach in ANLP. By successful, we mean that the systems and approaches have become established, generated a large amount of research, and/or become seminal works in ANLP. The eight chapters range from multiple text processing tools (e.g., Coh-Metrix, LIWC, DocuScope) through semantic assessment tools (e.g., LSA), to intelligent tutoring systems (e.g., AutoTutor, Summary Street) that incorporate numerous NLP approaches.

For any field to fully emerge, it has to be constantly and consistently producing high quality research. Section 3 features 16 such examples. The studies cover all aspects of ANLP including developing intelligent tutoring systems, text processing tools, algorithms, methods, techniques, and approaches.

Section 1

Following this introduction, Chapter 1 features Arthur C. Graesser, Vasile Rus, Zhiqiang Cai, and Xiangen Hu, who provide an overview of recent developments in question answering and generation. They define automated question answering as the task of providing answers automatically to questions asked in natural language, and they explain the flip process of question asking, which is automated question asking or generation, as the task of supplying answers automatically to questions by the use of various forms of input (e.g., text, meaning representation, databases). The authors also speculate on the future of these pursuits, arguing that question asking/generation will revolutionize learning and dialogue systems.

In Chapter 2, Martin Hassel and Hercules Dalianis discuss the development of automatic summarization systems. The authors' focus is on systems that use methods that are more or less directly transferable from one language to another.

In Chapter 3, Alexandra Kent and Philip McCarthey outline the basic theoretical assumptions that underpin the many different methodological approaches within Discourse Analysis. The chapter then considers these approaches in terms of the major themes of their research, the ongoing and future directions for study, and the scope of contribution to scientific knowledge that discourse analytic research can make.

In Chapter 4, Christian Hempelmann presents an account of key NLP issues in *search*. More specifically, he gives a general overview on NLP and *search* to show the advantages of ontological semantic technology (OST) and ways in which it can be implemented.

In Chapter 5, Martin Atzmueller gives an overview on data mining, focusing on approaches for pattern mining, cluster analysis, and predictive model construction. For each of these approaches, the author describes exemplary techniques that are especially useful in the context of applied natural language processing.

In Chapter 6, T. Daniel Midgley discusses historical and recent work in dialogue act tagging and dialogue structure inference. He explains that dialogue act tagging is a classification task in which utterances in dialogue are marked with the intentions of the speaker. The chapter argues that the structure of dialogue can be represented by dialogue grammar, segmentation, or with a hierarchical structure.

In Chapter 7, Vasile Rus, Mihai Lintean, Arthur C. Graesser, and Danielle S. McNamara discuss measuring semantic similarity between texts. According to the authors, semantic similarity can be defined quantitatively, e.g. in the form of a normalized value between 0 and 1, and qualitatively in the form of semantic relations such as elaboration, entailment, or paraphrase. The authors present a generic approach that relies on word-to-word similarity measures as well as experiments and results obtained with various instantiations of the approach.

In Chapter 8, Patrick Jeuniaux, Andrew M. Olney, and Sidney D'Mello address students and researchers who are eager to learn about practical programmatic solutions to natural language processing (NLP) problems. They discuss the role of programming and specifically the Python programming language. They then give a step by step approach in illustrating the development of a program to solve a NLP problem. The authors also provide some hints to help readers initiate their own NLP programming projects.

Section 2

In Chapter 9, Walter and Eileen Kintsch describe an educational application of latent semantic analysis (LSA) that provides immediate, individualized content feedback to middle school students writing summaries. The authors describe LSA as a machine learning method that constructs a map of meaning that permits researchers to calculate the semantic similarity between words and texts.

In Chapter 10, Arthur C. Graesser, Sidney D'Mello, Xiangen Hu, Zhiqiang Cai, Andrew Olney, and Brent Morgan describe AutoTutor, an intelligent tutoring system that helps students learn science, technology, and other technical subject matters. The authors also describe some ways that AutoTutor has been evaluated with respect to learning gains, conversation quality, and learner impressions.

In Chapter 11, Danielle S. McNamara and Arthur C. Graesser describe Coh-Metrix and studies that have been conducted validating the Coh-Metrix indices. Coh-Metrix provides indices for the characteristics of texts on multiple levels of analysis, including word characteristics, sentence characteristics, and the discourse relationships between ideas in text. They also describe the Coh-Metrix text easability component scores, which provide a picture of text ease (and hence potential challenges).

In Chapter 12, Cindy K. Chung and James W. Pennebaker examine the ANLP role of the linguistic inquiry and word count (LIWC) program. The authors explain that LIWC is a word counting software program that references a dictionary of grammatical, psychological, and content word categories. They go on to show that LIWC has been used to efficiently classify texts along psychological dimensions and to predict behavioral outcomes in a wide variety of studies in social sciences.

In Chapter 13, Cyrus Shaoul and Chris Westbury present the *Hi*gh *Di*mensional *Ex*plorer (HiDEx). HiDEx is a tool for exploring a class of models of lexical semantics derived from the Hyperspace Analog to Language (HAL). The authors describe HAL as a high-dimensional model of semantic space that uses the global co-occurrence frequency of words in a large corpus of text as the basis for a representation of semantic memory.

In Chapter 14, Mihai Lintean, Vasile Rus, Zhiqiang Cai, Amy Witherspoon, Arthur C. Graesser, and Roger Azevedo present the architecture of the intelligent tutoring system MetaTutor. The system trains students to use metacognitive strategies while learning about complex science topics. The authors particularly focus on MetaTutor's natural language components.

In Chapter 15, G. Tanner Jackson and Danielle S. McNamara discuss the intelligent tutoring system Interactive Strategy Training for Active Reading and Thinking (iSTART). iSTART utilizes a complex set of algorithms to evaluate student input and subsequently select real-time appropriate responses.

In Chapter 16, Suguru Ishizaki and David Kaufer present a corpus-based text analysis tool along with a research approach to conducting a rhetorical analysis of individual text and text collections. The tool, DocuScope, supports both quantitative and quantitatively-informed qualitative analyses of rhetorical strategies found in a broad range of textual artifacts.

Section 3

In Chapter 17, Danielle S. McNamara, Roxanne Raine, Rod Roscoe, Scott Crossley, G. Tanner Jackson, Jianmin Dai, Zhiqiang Cai, Adam Renner, Russell Brandon, Jennifer L. Weston, Kyle Dempsey, Diana Lam, Susan Sullivan, Loel Kim, Vasile Rus, Randy Floyd, Philip M. McCarthy, and Arthur C. Graesser present Writing-Pal (W-Pal), an intelligent tutoring system that provides writing strategy instruction to high school students and students entering college. The chapter describes the W-Pal system itself, as well as various NLP projects geared toward providing automated feedback to students using the system.

In Chapter 18, Philip McCarthy, Shinobu Watanabe, and Travis Lamkin present the Gramulator, a freely available tool for qualitative and quantitative computational textual analysis. The Gramulator is designed to allow researchers and materials designers to identify indicative lexical features of texts and text types. It also offers a wide range of text assessment metrics, and useful analysis tools such a concordancer, a lemmatizer, and a parser

In Chapter 19, Bryan Rink, Cosmin Adrian Bejan, and Sanda Harabagiu present a novel method for discovering causal relations between events encoded in text. In order to determine if two events from the same sentence are in a causal relation or not, they first build a graph representation of the sentence that encodes lexical, syntactic, and semantic information. From such graph representations, the authors automatically extract multiple graph patterns (or *subgraphs*). The authors sort the patterns according to their contribution to the expression of intra-sentential causality between events.

In Chapter 20, Nate Blaylock, William de Beaumont, Lucian Galescu, and Hyuckchul Jung describe a system for task learning and its application to textual user interfaces. The system, PLOW, uses observation of user demonstration, together with the user's play-by-play description of that demonstration,

to learn complex tasks. The authors suggest that PLOW may make it possible for users without any programming experience to create tasks via natural language.

In Chapter 21, Jennifer L. Weston, Scott A. Crossley, and Danielle S. McNamara examine the relation between the linguistic features of freewrites and human assessments of freewrite quality. This classical example of ANLP shows how one system (Coh-Metrix) can be used to address issues in development with another system (W-Pal).

In Chapter 22, Khaled Shalaan, Marwa Magdy, and Aly Fahmy address issues related to the morphological analysis of ill-formed Arabic verbs. Edit distance and constraint relaxation techniques are used to demonstrate the capability of the proposed system in generating all possible analyses of erroneous Arabic verbs written by language learners.

In Chapter 23, Kyokoa Baba and Ryo Nitta investigate the longitudinal effects of repeating a timed writing activity on English language learners. The authors analyze the texts using a variety of *Coh-Metrix* indices.

In Chapter 24, Wei Xiong, Min Song, and Lori Watrous-deVersterre evaluate SENSATIONAL, a novel unsupervised word sense disambiguation technique. The authors define word sense disambiguation as the problem of selecting a sense for a word from a set of predefined possibilities.

In Chapter 25, Scott A. Crossley and Danielle S. McNamara investigate the production *of* and exposure *to* lexical features when non-native speakers (NNS) converse with each other. The authors focus on lexical features that are associated with *breadth* of lexical knowledge including lexical diversity and lexical frequency.

In Chapter 26, Adam M. Renner, Philip M. McCarthy, Chutima Boonthum-Denecke, and Danielle S. McNamara describe the Harmonizer, a system that addresses the problem of user input irregularities (e.g., typos). The Harmonizer is specifically designed for intelligent tutoring systems (ITSs) that use NLP to provide assessment and feedback based on the typed input of the user. The performance of the tool is evaluated using various computational approaches on *un*edited input from high school students in the context of an ITS (i.e., iSTART).

In Chapter 27, Philip M. McCarthy, David Dufty, Christian Hempelman, Zhiqiang Cai, Danielle S. McNamara, and Arthur C. Graesser address the problem of identifying *new* versus *given* information within a text. The authors discuss a variety of computational new/given systems and analyze four typical expository and narrative texts.

In Chapter 28, Andrew J. Neel and Max H. Garzon take a new approach to the problem of recognizing textual entailment (RTE). They show that semantic graphs can provide a very competitive performance. The semantic graphs are made of synonym sets (*synsets*) and selected relationships between those synsets.

In Chapter 29, Aqil Azmi and Suha Al-Thanyyan present *Ikhtasir*, an automatic extractive Arabic text summarization system. The system integrates a Rhetorical Structure Theory (RST) based system with a sentence scoring system, where individual sentences are scored.

In Chapter 30, Kirk Roberts, Cosmin Adrian Bejan, and Sanda Harabagiu discuss an ontology-based method for improving the disambiguation of ambiguous location names (or toponyms) using limited event semantics. Location names are often ambiguous, as the same name may refer to locations in different states, countries, or continents.

In Chapter 31, René Venegas approaches three automatic methods for the evaluation of summaries from narrative and expository texts in Spanish. This task consists of correlating the evaluation made by human raters with results provided by latent semantic analysis.

In Chapter 32, Courtney M. Bell, Philip M. McCarthy, and Danielle S. McNamara use Coh-Metrix and LIWC to investigate gender differences in language use within the context of marital conflict.

THE FUTURE OF ANLP

The future of ANLP is bright. It is inconceivable that the coming years will see anything less than a continuing rise in the number, availability, and scope of computational systems that address real world issues through the medium of language. These systems will develop the ever growing need of users to request and retrieve information quickly, easily, and accurately. Each avenue of daily life will increase its dependency on language related applications: governmental, commercial, educational, recreational; system designers will seek out new approaches, methods, and techniques that address issues such as speech recognition, question answering, information extraction, and all such computationally linguistic tasks that are discussed in this book. Soon enough, other researchers will collect the algorithms that make these approaches, methods, and techniques possible, and with them, they will create newer, faster, and more accessible analysis systems, which, in turn, will find yet newer researchers who use these algorithms in novel applications. In short, the *identification* of computationally solvable language issues will be addressed by a broad *investigation* of developing textual analysis systems, which will lead to a *resolution* through *applied natural language processing*.

REFERENCES

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *136*, 180–193. doi:10.3758/BF03195563

Hearst, M. (2005). *Teaching applied natural language processing: Triumphs and tribulations*. The Second ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing And Computational Linguistics, Ann Arbor, MI, June 2005.

Kintsch, W., & Kintsch, E. (in press). LSA in the classroom. In McCarthy, P. M., & Boonthum-Denecke, C. (Eds.), *Applied Natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

McCarthy, P. M., Guess, R., & McNamara, D. S. (2009). The components of paraphrase evaluations. *Behavior Research Methods*, *41*, 682–690. doi:10.3758/BRM.41.3.682

McCarthy, P. M., Rus, V., Crossley, S. A., Bigham, S. C., Graesser, A. C., & McNamara, D. S. (2007). Assessing entailer with a corpus of natural language from an intelligent tutoring system. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference* (pp. 247-252). Menlo Park, CA: The AAAI Press.

McCarthy, P. M., Rus, V., Crossley, S. A., Graesser, A. C., & McNamara, D. S. (2008). Assessing forward-, reverse-, and average-entailment indices on natural language input from the intelligent tutoring system, iSTART. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference* (pp. 165-170). Menlo Park, CA: The AAAI Press.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA systems. In Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.), *Handbook of latent semantic analysis* (pp. 227–241). Mahwah, NJ: Erlbaum.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*, 222–233. doi:10.3758/BF03195567

Rus, V., Feng, S., Brandon, R., Crossley, S. A., & McNamara, D. S. (in press). A linguistic analysis of student-generated paraphrases. In C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference* (pp. 293-298). Menlo Park, CA: The AAAI Press.

Rus, V., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2008). A study of textual entailment. *International Journal of Artificial Intelligence Tools*, *17*, 659–685. doi:10.1142/S0218213008004096