

# Natural Language Processing

## Scenario

Researchers at Langdrum University are exploring ways to apply tools and techniques of natural language processing (NLP) to automate the scoring of student writing. Such work explores the integration of text-mining, machine-learning, and NLP-specific algorithms to evaluate aspects of writing for educational purposes. This approach has particular value for assessing standardized tests. Grading essays by hand is time-consuming and subjective; automated grading shows promise for speeding assessments and getting feedback to students faster and with greater consistency and efficiency than a human scorer might provide. (As one example, the prominent testing company [ETS](#) has published considerable research about its work with automated text scoring.) Langdrum researchers have consulted with peers at the [University of British Columbia](#) who are developing technology to automatically extract and analyze different features of argumentation from written text/essays and with researchers at the [University of Technology Sydney](#) who are working on technology to assess reflective writing.

Automated testing is just one of many research threads in the Center for Natural Language Processing at Langdrum. One research group is exploring how NLP-informed voice-recognition software can be used in classrooms. One model has students asking voice-recognition devices questions that they might otherwise research in a library. Another application focuses on ways the technology can help learners with disabilities. The same team is also developing NLP-related tools to help students conduct research in different languages. A team of graduate students has started to map how voice recognition may one day be integral to pedagogy and integrated with university curriculum.

Another research group is engaged in innovating ways NLP can help businesses. This group published an algorithm that quickly converts findings from “sentiment analyses” of customer comments on a business’s website into actionable practices that can increase sales. Another effort explored ways NLP could identify and analyze patterns in customer inquiries about products.

Still another group is exploring some of the ethical questions that NLP raises. Protecting privacy is a major concern, but other issues range from whether NLP will unfairly pigeonhole people based on its analyses (such as when it is used to screen applicants for jobs), whether data sets constructed using NLP contain inadvertent biases, and what kinds of guidelines might be needed to infer findings from user data.

## 1 What is it?

With roots that go back to translation projects at MIT in the 1950s that were funded by the Department of Defense, the field of **NLP marries the power of artificial intelligence with linguistics to process and analyze language-based data.** By one [definition](#), NLP “helps computers understand, interpret, and manipulate human language.” NLP enables researchers to go deep into text and speech to mine language for correlations, patterns, and other findings—teasing out meaning that might be detectable only through powerful computation as well as expediting analyses that would take humans relatively significant effort to complete on their own.

## 2 How does it work?

**NLP draws on a variety of techniques for interpreting human language**, including statistical and machine learning methods, as well as rules-based and algorithmic approaches. NLP considers the building blocks of language—phonemes, syllables, morphemes, words, phrases, and sentences—as data. NLP tools analyze these data segments, looking for latent structures and patterns in language. In this way, NLP seeks to understand relationships between these pieces of data in ways that create new and deeper understanding. Commonly used for text mining, translation, and generation of automated speech recognition dialogue systems, NLP is useful for quickly summarizing text-based content and finding and interpreting useful connections in text. For example, a researcher might write a system that analyzes patterns to derive deeper meaning from large bodies of text found in the real world, such as a batch of newspaper articles or a grouping of medical records. The NLP analysis of language as data can be conducted at the relatively narrow level of syllables and morphemes—if, for example, the goal is to detect the mother tongue or origins of a speaker—but NLP can also be invaluable in helping to solve big, complex problems, such as voice-controlled applications in self-driving cars or voice-interactive appliances.

## 3 Who’s doing it?

Providing intellectual underpinnings for such well-publicized examples of NLP in action such as Siri, Alexa, and Watson, **research in colleges and universities on NLP is expanding.** Reflecting its inherent multidisciplinary nature, for example, NLP at [Columbia University](#) is conducted in the computer science department, the Center for Computational Learning

# Natural Language Processing

Systems, and the biomedical informatics department. A recent [project](#) there explored new ways to analyze writer sentiments in complex postings written in response to Arabic newspaper articles. Research at the [University of Delaware](#) includes work to apply NLP techniques to assistive technologies for people with disabilities, particularly those with severe speech or sight impairments. Recent [NLP research and applications efforts](#) at the University of British Columbia's computer science department used text-mining and summarization techniques to synthesize and summarize contents of lengthy online conversations and blogs. UBC's [Science Centre for Learning and Teaching](#) and [Centre for Teaching, Learning, and Technology](#) used latent semantic analysis methods and rule-based grammars to develop learning technologies applications to [assess](#) the writing and argumentation patterns in student essays.

## 4 Why is it significant?

NLP helps linguists understand aspects of a language and conduct studies comparing one language to another. It **informs the technology that drives online translation software and speech-recognition applications**. NLP is especially relevant today as a foundational pillar in the voice command technology behind Apple's Siri, Amazon's Alexa, and other digital assistants. Researchers are developing added value for NLP as they use it to take a deeper dive into text than has previously been possible, to mine information in text for connections that help lead to new knowledge. That value has been shown, for example, in the insights that IBM's Watson is able to derive by scanning huge text-based databases, with applications in such fields as health care, education, and energy.

## 5 What are the downsides?

NLP is a fundamentally difficult challenge, in part because it **seeks to analyze the complex structures and meanings of natural language using tools and methods from mathematics, statistics, and machine learning**. A related issue is that while extensive pattern dictionaries, semantic taggers, and other tools that can help NLP researchers parse meaning and structure have been developed

for some of the world's more popular languages (such as English), such resources are not readily available for other languages. Deeper concerns include such issues as whether reliance on NLP will diminish learners' abilities to analyze texts on their own and whether NLP might be another technology that erodes human contact.

## 6 Where is it going?

With the emergence and increasing importance of big data in many businesses and industries, **NLP is primed to become an even more important tool for analytics and knowledge generation**. In the field of medicine and health care alone, for example, the value of a tool like Watson is only beginning to be realized. Many other fields—including but not limited to biology, genetics, and science education—will deepen their use of NLP for fundamental research. Further, applications based on NLP, such as digital assistants, will become more ubiquitous. Concomitantly, researchers will continue to hone NLP techniques and develop more sophisticated NLP methodologies.

## 7 What are the implications for teaching and learning?

While NLP has gained a foothold in education through applications like Turnitin plagiarism-detection software and Carnegie Mellon University's [cognitive tutors](#), other applications will evolve in the years ahead. Researchers at the University of Arizona, for example, have begun to explore how digital assistants like Alexa might be [used in the classroom](#) as part of automated systems for teaching and learning. NLP will likely be essential in advancing efforts to scale automated techniques for assessing learners' work, including assessments of student writing, such as the work mentioned above at UBC. Another potentially significant application is likely to be the use of NLP in **processing narrative input that is part of learning analytics and efforts to support student success**. Broadly speaking, NLP will continue to provide powerful channels for researchers to bridge the gap between humans and computers as they derive new and deeper meaning about language, communication, and the human experience.