



The oldest known dictionaries came long before Noah Webster. They were Sumerian-to-Akkadian bilingual dictionaries written on clay tablets in the 2300s BCE.

INIT

## New Directions in Language Processing

**A**lan Turing, one of artificial intelligence's most visionary scientists, ended his famous 1950 paper titled "Computing Machinery and Intelligence" with the following warning note of realism: "We can only see a short distance ahead, but we can see plenty there that needs to be done."

The idea to use computers to simulate the human ability to communicate in natural language is as old as the field of computer science itself. When work in what would later become the field of natural language processing (NLP) began in the 1950s, the goals and predictions for the field were truly ambitious: to create a true-to-life translation system between English and Russian by the turn of the decade. Yet, nearly 65 years later, we are still struggling to get software to perfectly distinguish between common and proper nouns in a sentence. Instead of attaining the language-capable systems we set out to achieve, the field has taught us a lesson in humility on the complexity of human language. When results fell short of their early ambitions, disillusioned researchers and funding agencies concluded

language technology was not yet ready for real-world applications.

Despite these early setbacks, the field has grown, matured, and branched out over the last two decades. Its many areas have seen steady



Daniel Bauer

**What makes us human is inextricably linked to our ability to use language, we believe the future belongs to intelligent and language-able computers.**

and remarkable progress, coupled with many shifts of paradigm. What started out as an approach with a heavy bearing on linguistics and hand-crafted rules eventually shifted toward statistics-driven machine learning, and is still evolving. For instance, the established Bayesian algorithms of machine learning are starting to lose ground to a recent re-emergence of neural networks. There is also an increasing perception that NLP needs to focus not only on the form, but on the meaning of natural language, and on the context in which language is used.

As a result of these transitions and tedious leaps forward, we can now boast credible, if imperfect, machine translation systems, or being able to give voice commands to our smartphones. As our lives move more and more online, language technology has become increasingly important in helping us communicate and find information more efficiently.

### INSIDE THE ISSUE

Today, the question is no longer whether NLP can provide solutions for real-world problems, but rather which steps to take next in order to bring us closer to our goal of truly intelligent language

software. The current issue of *XRDS* was conceived to give you a taste of the breadth and scope of current work in language processing and its applications.

Percy Liang's article provides an introduction to past, current, and future research on the problem of making computers understand natural language.

Traditionally, much of the work in NLP has focused on English and closely related languages. Today the field is becoming more and more multilingual. Two of the articles in this issue address how software can help students acquire a new foreign language. Mariano Felice and Zheng Yuan discuss the problem of automatically correcting grammar errors made by language learners; and Daniel Bauer and Billy Rathje interview Miriam Plieninger about mobile apps for language learning. While Heba Elfardy's article describes work on automatically detecting when Arabic speakers switch from one dialect into another within the same conversation.

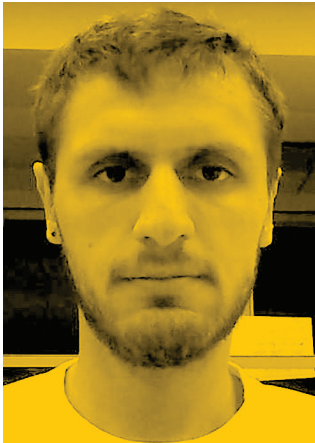
Language processing research spans different modalities. Per Ola Kristenson looks at new intelligent methods to make text entry easier and faster

# 1888

The international phonetic alphabet, which contains 74 consonants and 28 vowels, was established.

on touch interfaces. Pierre Lison and Raveesh Meena report on speech input and output for computational agents that are truly conversational.

The articles in this issue also highlight the explora-



Adrian Scoică

**We can now boast credible, if imperfect, machine translation systems, or being able to give voice commands to our smartphones.**

tion of different types of text genres from different time periods, ranging from Ancient Sumerian accounting records to short social media messages. They also demonstrate how NLP has started to reach out to other disciplines, in particular the digital humanities. Terhi Nurmikko-Fuller's article describes work on digitizing Ancient Sumerian texts to make them available to historians and archeologists, while Sara Rosenthal's article describes methods to automatically detect influencers on social media based on concepts taken from social science.

## FINAL WORD

As so much of what makes us human is inextricably linked to our ability to use language, we believe the future belongs to intelligent and language-able computers. With a bit of luck, a dash of inspiration, and plenty of hard work and perseverance, the days in which you will be able to sit down and discuss the contents of this magazine with your laptop or smartphone are not beyond reach. We hope you will enjoy exploring this issue of *XRDS*.

—Adrian Scoică and  
Daniel Bauer, Issue Editors

## MILESTONES

# Technology for Talking

The field of natural language processing (NLP) is broad and spans many subdisciplines including understanding language, generating language, extracting information and sentiment, and more. Here, we examine just a few of the many major events in the history of NLP:

**1950** Allen Turing publishes his famous “Turing test,” which defines a standard for a machine to be considered intelligent: A human interrogator should not be able to tell the difference between a computer and human in a conversation.

**1968** During the Cold War SYSTRAN is founded by Dr. Peter Toma. The company translates Russian texts to English for the United States Air Force.

**1971** Les Earnest and Ralph Gorin creates the first spell checker program, which displays corrected alternatives of misspelled words that differed by one letter and small transpositions.

**1997** The search engine, AltaVista, releases BabelFish, which is based on SYSTRAN. The service provides free machine translation in 10 major European languages to the public.

**2011** IBM's Watson appears on the television game show “Jeopardy!” and utilizes a combination of NLP, information retrieval, and representation algorithms to win the first place prize of \$1 million.

—Jay Patel