

Note from the editor: “Expositions of Romanian scientists on the design of text-to-speech synthesis and natural language understanding and generation systems”

Amy Neustein

Published online: 4 December 2009
© Springer Science+Business Media, LLC 2009

This special issue of the Journal is devoted to the assorted work of twelve Romanian speech scientists who have brought sharp analytical insights and new methodological approaches to the design of high performing text-to-speech (TTS) and natural language understanding and generation systems.

Volume 12, No. 2 consists of five papers. Each paper, albeit addressing a different aspect of speech synthesis and natural language processing, is like the others in the rigorous diligence and precision of the author’s scientific scrutiny. The contributors to this special issue have served their Romanian academic and research institutions with distinction, while adding to the vast body of research in speech technology through diverse publications in the professional literature, not to mention their presentations at various academic conferences and workshops.

The inspiration for this special issue comes from individuals such as Dan Tufis of the Romanian Academy, Institute for Artificial Intelligence. Back in 2004, I had the pleasure of hearing Dan give an enlightening presentation on multilingual word sense disambiguation (WSD) at the annual Natural Language Processing and Cognitive Science workshop, an event chaired by Bernadette Sharp of Staffordshire University. Several years later, when I became IJST’s Editor-in-Chief, I explored the Journal’s archives and found a compelling special issue dealing with speech technology research taking place in Romania (Volume 5, No. 3, September 2002). The guest editor of that issue, Dragos Burlieanu of Politehnica University of Bucharest, working under the auspices of the Journal’s then-Editor-in-Chief

Daryle Gardner-Bonneau, had assembled papers that “were intended to give the reader an accurate image of the state-of-the art [in] fundamental research and applied technology.”

This current special issue of the Journal, likewise, aims at informing the reader about new and exciting advances in Romanian speech technology. The contributing writers carefully explore, critique and compare new methodological approaches and practical applications, in a determined attempt to solve some of the most vexing problems in natural language processing and text-to-speech synthesis systems. For example, Dragos Burlieanu and his co-authors, Catalin Ungurean and Aurelian Dervis, confront head-on the problems arising from the fact that lexical stress (appearing at the level of word-syllables) is a fundamental component of the pronunciation of a word, making its correct placement crucial for prosody prediction and the resultant generation for high-quality TTS. The authors assert that spoken dialog systems must perform “an adequate modeling of prosody”; they point to the fact that “prosody is crucial for speech synthesis because it conveys aspects of meaning and structure that are not implicit in the segmental content of utterances.” In saying this, these authors don’t take lightly the conundrum facing designers of speech synthesis systems today—“Synthesized messages can become ambiguous or even unintelligible because the prosodic and rhetorical information is missing (or is wrongly positioned)”—but they stress the reality that “prosody is difficult to predict in TTS synthesis systems because the input text contains little or no explicit information about meaning and structure, and such information is very hard to deduce automatically.” To solve this problem, Burlieanu and his co-authors propound a statistical approach to lexical stress assignment for Romanian TTS synthesis, which is premised on “*n*-gram language models at character level and a modified Katz back-off smoothing

A. Neustein (✉)
Linguistic Technology Systems, 800 Palisade Avenue, Suite 1809,
Fort Lee, NJ 07024, USA
e-mail: amy.neustein@verizon.net

technique to solve the problem of data sparseness during training.”

Doina Jitca of the Romanian Academy, Institute of Computer Science, along with several colleagues, enhances this special issue with a presentation of an intonation description language derived from a robust analysis of a Romanian speech corpus. This intonation description language is based on the decomposition of complex intonational phrases (IP) into a tree of accentual units (AUs) and accentual unit groups (AUGs), which are then assigned functional labels at the communicative act level. The authors warrant their use of this intonation description language “to perform flexible micro-prosodic descriptions of the text in a Romanian Text-to-Speech (TTS) system in order to control the FO contour generation.” Jitca and her colleagues’ novel research on prosody is built on some of her earlier work that appeared in a prior issue of this Journal, dealing with speech technology research in Romania. That work described a method for increasing perceptual quality of synthetic speech by introducing a fuzzy paradigm (employing a rule-based formant synthesizer). Jitca’s approach aimed at solving the problem of “phonemes that are prone to multi-definitions in rule-based speech synthesis” in the implementation of a TTS system for Romanian language.

Dan Cristea of the Alexandru Ioan Cuza University of Iasi, Institute of Computer Science (Iasi branch of the Romanian Academy), offers a fascinating discussion of the cohesion component of global discourse processing. His article shows how Veins Theory “computes strings of discourse units, called veins, from which domains of accessibility can be determined for each discourse unit.” Cristea’s work is informed by a rich understanding of human cognition and the “innate cognitive mechanism that allows [humans] to obtain naturally the most plausible interpretation of a text . . . based on the maximal satisfaction of a constraints system.” In explicating how Veins Theory—validated by drawing on examples from corpora of English, French and Romanian—can account for the relationship between discourse structure and referentiality, Cristea points to some of the practical applications of this method:

- (1) to constrain a simultaneous parsing and anaphora resolution process so that discourse interpretations will require minimal inferential load in building the structure and in identifying the antecedents of referential expressions;
- (2) to correct the discourse structure when referential links are known; and
- (3) to guide a process aimed at producing focused summaries.

Vladimir Popescu of Politehnica University of Bucharest (and the University of Avignon, France) and his co-authors, Jean Caelen and Corneliu Burileanu, demonstrate through

case study the benefits of a constraint satisfaction approach to context-sensitive natural language generation in *multi-party dialog systems*. These authors argue the merits of providing a computational framework for handling important semantic and pragmatic features of discourse, such as the degree of illocutionary force, ellipsis and lexical selection of concessive connectors, for natural language generation. In so doing, the authors carefully examine the way in which “constraints influence the generation of context-sensitive . . . utterances in simultaneous interaction with several human speakers.” Popescu and his colleagues show how the utterance generation process is controlled in four stages. For each machine turn in multi-party dialogue, the following stages occur:

- (1) the discourse structure representing the dialogue is updated;
- (2) decisions are made regarding the adjustment of the degree of strength for the illocutionary/perlocutionary force of the utterance;
- (3) decisions are made with regard to the semantic ellipsis;
- (4) concessive connectors are selected between utterances.

Radu Ion and Dan Tufis of the Romanian Academy, Institute for Artificial Intelligence, take a close look at how linguistic ambiguity, expressed at many different levels—for example, phonological, morphological, lexical, syntactical and semantic—poses one of the most difficult challenges for Natural Language Processing. In tests that compared the performances of two different word sense disambiguation (WSD) systems—one that is composed of multilingual ontologies, applicable to parallel corpora and requiring aligned wordnets, and another that relies solely on monolingual data (text and wordnet)—the authors found that multilingual WSD is significantly more precise than monolingual WSD. To make certain that their testing methods were reliable and valid, the authors were at pains to control for the granularity of sense distinctions. Still they found that *supervised* WSD systems, which allow for a human translator to embed knowledge into parallel texts, provide a measurable advantage over unsupervised WSD systems.

While acknowledging that multilingual WSD can be compromised by “the existence of parallel data, and, at least, two aligned wordnets that might not be available in all cases,” Ion and Tufis recognize that “parallel resources are becoming increasingly available . . . and aligned wordnets are being produced for more and more languages.” The authors end on sanguine note, forecasting that “in the near future it should be possible to apply ours and similar methods to large amounts of parallel data and a wide spectrum of languages.”

I too will end on a sanguine note. Just look at the twentieth European Signal Processing Conference (EUSIPCO) scheduled to take place in Bucharest in 2012. Among the

main organizers of this event are none other than Dragos Burlieanu and his dedicated colleagues at Politehnica University of Bucharest. This is a remarkable accomplishment. When I asked Dan Tufis to explain why Romania has become a leader in speech technology, his answer was straightforward and revealing: “The scientific community grew out of enthusiasm and passion of research leaders and their students. The Romanian speech scientists, be they living in Ro-

mania or teaching abroad, are among the best scientists in their field because they love what they are doing and were raised in a respectful attitude toward languages. And this is why at many technological speech related competitions ... one can always see Romanians among the best competitors.” I agree. And this is why their work deserves a special issue of this Journal.