

# Poznan Linguistic Meeting 2005 & Language and Technology Conference 2005

## Joint Panel on

### “Technology for Linguistics, Linguistics for Technology”

22 April 2005, 16:30 to 19:00/19:30

#### Aims

Computer-processed language and speech are becoming pervasive in everyday life – from GSM speech compression and VoIP (Voice over Internet Protocol) on the signal processing side to synthetic speech announcement in public places, reading software for the blind, dictation software to spell checkers, grammar correctors, word guessers, the text linguistic models which underlie word processor stylesheets and web documents and the powerful concordancing facilities of search engines like Google. If this is true of everyday life, then presumably the same is happening in linguistics. But what exactly is going on in this area? How are linguistic content and method being influenced by the Human Language Technologies? The idea of the panel on “Technology for Linguistics, Linguistics for Technology” is to summarise what technology and linguistics are currently contributing to each other, and what they may contribute to each another in the near future. The panelists have formulated statements on their views and their vision (see below), and participants at the two conferences are invited to exchange their views with the panelists.

#### Panelists

Nicoletta Calzolari (Istituto di Linguistica Computazionale del CNR, Pisa, Italy)  
Nick Campbell (ATR Network Informatics, Kyoto)  
Ron Cole (Center for Spoken Language Research, University of Colorado at Boulder)  
Grażyna Demenko (Adam Mickiewicz University, Poznań)  
Maria Gavrilidou (Institute for Language and Speech Processing, Athens)  
Dafydd Gibbon (Bielefeld), Moderator  
Zygmunt Vetulani (Adam Mickiewicz University, Poznań)

#### Schedule

16:30	Introduction (Moderator)
16:45	Statements by panelists (10 min + approximately 5 min discussion each)
18:30	General discussion – interventions by conference participants are very welcome!
19:00	Final comments by panelists

#### Statements

##### *Nicoletta Calzolari*

##### **Language Resources, Language Technology, Linguistics: isn't this too narrow?**

I'd like to touch a few issues related to some critical interaction between Language Resources (LR), Language Technology (LT)<sup>1</sup>, and Linguistics (L). But today this is not enough. For our field to have a real impact, we need to broaden our vision to other issues, such as interaction between different communities (also outside LT), and the importance of organisational aspects in addition to technical ones.

*Between LR and narrow LT (i.e. tools, systems, components, etc.):*

There is a loop between i) lack of suitable, large-size and knowledge intensive LR (lexicons and corpora, with rich syntactic and semantic annotation), and ii) systems' ability to use them effectively: are there systems (non ad-hoc toy systems) able to use real-size lexicons with very fine-grained semantic/conceptual information? The two paths should be pursued in parallel, closely interact with each other, and be gradually

<sup>1</sup> In a broad sense of LT (the one I prefer), LR are a component of LT. In a narrow sense, LT may be considered just as processing (tools, systems, etc.) vs LR as data.

integrated. This is not yet happening today, and requires more overall coordination.

#### *Between LR and L:*

A consequence of the corpus-based approach (e.g. to lexicon building) is that it compels to break hypotheses too easily taken for granted in mainstream linguistics. In actual usage – as revealed by corpus analysis –, one of the main characteristics of language is that of displaying many properties which behave as a continuum, not as “yes/no” properties. The same holds true for so-called “rules”, where we find more frequently “tendencies” towards a rule than precise rules, so that many of the theoretical rules appear to be simplifications or idealisations which are in fact dispelled by real usage. A number of dichotomies must then be reconciled, such as: rules vs. tendencies, absolute constraints vs. preferences, discreteness vs. continuum, theoretical vs. actual, theory-driven vs. data-driven, intuition/introspection vs. empirical evidence.

#### *LR and the future of LT:*

Broadening our perspective into the future, the need of ever growing LR for effective multilingual content processing requires a change in the paradigm, and the design of a new generation of LR, based on open content interoperability standards. The Semantic Web notion is going to crucially determine the shape of the LR of the future, consistent with the vision of an open space of sharable knowledge available on the Web for processing. The effort of making available millions of ‘words’ for dozens of languages is something that no single group is able to afford. This objective can only be achieved when working in the direction of an integrated Open and Distributed Linguistic Infrastructure, where not only the linguistic experts can participate, but which includes designers, developers and users of content encoding practices, and also many members of the society. We claim that the field of LR and LT is mature enough to broaden and open itself to the concept of cooperative effort of different set of communities (e.g. spoken and written, LT and Semantic Web, theoretical and application oriented).

#### *Importance of organizational aspects:*

The approach to realise such a linguistic infrastructure requires the coverage not only of a range of technical aspects (e.g. pertaining to linguistic modelling), but also – and maybe most critically – of a number of organisational aspects. In order to set up the required world-wide language infrastructure on the web, an essential aspect for ensuring an integrated basis is to enhance the interchange and cooperation among many communities that act now separately, such as LR and LT developers, Terminology, Semantic Web and Ontology experts, content providers, linguists and so on. This is one of the challenges for the next years, for a usable and useful “language” scenario in the global network. Moreover, such a language infrastructure may be inherently market driven, since the most widely used language portions may be the best developed and supported, and this has to be seriously considered.

### ***Nick Campbell***

#### **A Call for the Processing of Non-Verbal Speech Information**

The interface between people and computer-based information systems has become ubiquitous and is constantly evolving to include even more speech and language technology. Because it is widely assumed that we communicate primarily by means of language, most of the processing performed in current speech technology is still firmly based in linguistics. However, I shall argue that human-to-human communication is only partly linguistic, that the deliberate transfer of propositional content is in turn only a small part of speech communication, and that by far the greatest amount of information to be derived from human speech is paralinguistic or non-verbal in nature. Future speech technology will need to take this information into account if it is to meet the needs of an advanced media society.

Currently there is only very limited understanding about the transfer of paralinguistic information, since its study falls between the disciplines of psychology, sociology, acoustics, linguistics, and information theory, though it is neither central nor essential to any of them. Our recent collection and analysis of a very large corpus of naturally-situated conversational-speech has revealed that more than half of the utterances in a typical daily conversation are affect-bearing (A-type), as opposed to information-transmitting (I-type), and that they function to display the speaker-listener relationships, to control the flow of the discourse, and to reveal the speaker's intentions and affective states so that the text or linguistic content of a conversation can be properly interpreted.

The interpretation of paralinguistic (A-type) information requires knowledge of the prosody of a spoken utterance, but the processing of speech prosody has been confined to its linguistic (thematic, syntactic & semantic) function for speech synthesis, and largely ignored in most speech recognition. I shall argue in this talk that we need the help of linguists to formulate a new grammar of speech communication that is independent of text analysis, and that models the way that prosodic information (including voice-quality) is

used to determine the relationship between an utterance content (i.e., its linguistic information) and its spoken realisation (which is subject to paralinguistic modification).

I shall introduce a framework that can be used both in speech synthesis, for the signalling of non-verbal information, and in speech recognition, for its decoding and interpretation, and will draw a parallel with the way that an ADSL modem uses a simple twisted-pair telephone cable to simultaneously and transparently convey both speech and computer information, showing by implication that the speech signal can be decomposed into its linguistic component (which present speech technology can now process very well) and its paralinguistic or non-verbal component (which is currently as transparent to speech processing technology as would be the ADSL computer information to a traditional telephone engineer without a modem). In this panel discussion, I shall urge both Language Resource developers and Language Technology scientists to consider this important second channel of speech information so that future technology may become sensitive not just to what we say, but also to the way that we say it, by incorporating a new source of non-verbal speech information into the processing and interpretation of linguistic content. I shall aim to forge a union between speech technologists and linguists so that the different disciplines may both contribute to the development of this field in their future work.

*Ron Cole & Sarel van Vuure*

### **The Emerging Reality of Virtual Teachers and Virtual Therapists**

In March 2005 in Hanover Germany over 500,000 people attended Cebit, the world's largest technology trade show, to view the latest technology marvels. Germany's Siemens AG unveiled its new Animated Instant Voice Messages, in hopes of transforming text messages from boring print to a more interactive experience. "According to an Associated Press Report, "The program converts the text in a wireless message into speech that can be synchronized to play with moving animated lips superimposed on one of the user's own photographs. European users will get the first chance to see it, likely later this year." So I will soon be to call your cell phone, leave a message, and my pretty face will appear on the screen of your cell phone to say it.

We can expect more sophisticated systems to appear on cell phones and computers in the next few years. As speech recognition and semantic parsing technologies improve, the animated face will become more realistic and human-like, producing head movements and facial expressions that are consistent with the meaning and emotional content of typed or spoken messages.

At the Center for Spoken Language Research, my colleagues and I are developing computer programs that are being used to teach children to read and learn from text, and to help individuals with Parkinson disease or aphasia to improve their speech communication skills. While these programs are quite different in terms of the nature of the interaction between the virtual human and the student or patient, each requires understanding and modeling the behaviors of a human expert who is sensitive and effective in their task domain.

For both the virtual reading tutor and virtual speech therapist, I will present the theoretical and scientific rationale for the system, the design process, which aims to optimize the user experience and treatment outcomes, the challenges that had to be addressed to produce the current systems, and work now underway to create next generation experiences.

*Grażyna Demenko*

### **Linguistics in the advancement of speech technology**

#### **A) Speech Science in Speech Technology**

1. The synergy between SS and ST: for a successful ST system it is clearly and understandably difficult to build on anything other than on a strong and scientific foundation. And technology, in turn, provides infrastructure for the empirical and theoretical processes of research.
  1. Differing goals of SS and ST
  2. Methods of acquiring and representing knowledge
  3. Methods of processing and using knowledge
2. Integration of SS and ST
  1. How much linguistic knowledge is used in ST

**B) How much linguistic knowledge should be used in ST:** We are far from solving the ASR or TTS problem fully, and to the extent that human performance requires solving related AI problems (such as speech understanding), we might never fully achieve this goal.

1. State of the art
2. Directions of development of ST

ASR		TTS	
<b>Accept input:</b>	spontaneous	<b>Deliver output:</b>	intelligible
	emotional		understandable
	disfluent		personalized
	different styles/dialects		expressive
	different voices		communicative

### 3. Milestones in ST<sup>2</sup>

	1960- 70's	1980's	1990's
<b>Vocabulary</b>	Small, medium	Large	Very large
<b>Spoken corpus</b>	Isolated word, short phrases	Connected phrases Continuous speech	Speech understanding Spoken dialogue
<b>techniques</b>	Filter bank analysis DTW, LPC,	HMM, Stochastic language modelling	Machine learning Concatenative synthesis
<b>Corpora used</b>	Small	large	Very large

	2000	2010-2020
<b>Vocabulary</b>	Very large	unlimited
<b>Spoken corpus</b>	Dialog system, robust system	Multilingual, multimodal system
<b>Tasks</b>	Limited	unlimited
<b>Techniques</b>	Machine learning, concatenative synthesis with signal processing	?
<b>Corpora used</b>	Very large	?

### C) Limits of ST development

1. Creating a database
2. Database size - a dangerous tendency
3. Lack of perspectives for developing very large or unrestricted corpora for ST systems

### D) Prosody as a key to development of ST systems

1. Limited use of prosody in current ST systems
  1. Machine learning
  2. Prosody models for practical applications:
    1. In most prosodic models, too much emphasis was put on intonation and thus these models are not complete, since F0 cannot be varied in isolation without affecting other acoustic properties of the speech signal such as spectral tilt, voice quality and intensity<sup>3</sup>.
    2. Segmental variability is prosodically meaningful, but there is no algorithm for explaining relations
    3. Generalization of prosodical rules is difficult and could lead to faulty modelling
  3. Correlation between linguistic/phonetic information and prosody is not enough explained.
2. Potential for implementing prosody in ST
  1. The need for a quantitative working model of prosody: Because existing prosodical knowledge is not complete, it is best to use a combination of data-driven and knowledge-based approaches. Creative ideas about phonetic phenomena could come from well-controlled phonetic experiments. Subsequently, these concepts should be verified and quantified using very large speech corpora. In this way, prosodical knowledge can be acquired and more easily integrated into speech technology.
  2. Innovative approaches:
    1. Sophisticated databases: although a complete solution cannot be found in current phonetic/linguistic knowledge, this knowledge should certainly be taken into consideration while searching for new techniques for better systems.
    2. Paralinguistic annotation, access to paralinguistically transcribed text: this requires not only

<sup>2</sup> Partially based on : Challenges in Speech Recognition, Lawrence Rabiner. Gardner-Bonneau D. (red) (2003) Special Issue on Speech Synthesis, Int. Journal of Speech Technology, Kluwer Academic Publishers.

<sup>3</sup> Campbell, N., (2004): "Accounting for voice-quality variation", In SP-2004, Nara, 217-220.

phonetic knowledge, but also knowledge from many other disciplines. For instance, psycholinguistic models could be useful. In ASR the correct phone(me) and word sequences are not readily available. In order to integrate models from different disciplines, a lot of gaps still have to be bridged.

### ***Maria Gavrilidou***

#### **To predict what technology and linguistics may contribute to one another in the near future**

➤ I use the term linguistics in the broader sense of language study.

Historical evolution of the relationship of the two fields: close relationship

- initially: technology was practiced by and intended for the gurus, an elite of engineering → a cult evolved, there were devoted followers, "inside" people
- next step: technology led to applications useful for humanity in general
- the communication with applications was non-verbal, there was no use of language
- subsequently language was introduced at the interface level (made communication easier).

Language comes to be used as a means of communication with applications:

1. language makes man - machine communication more human
2. language brings technology closer to man (and especially) to the lay-man
3. in political terms, language brings an air of democracy to technology: language is a system everybody uses, so a much broader range of people can communicate with the technological applications
4. language is
  - (a) a communication system
  - (b) a system representing thought, i.e. one of the indications on how the human mind works (AI).

Technological advances provided to the study of language

- computing / crunching capacity (evidence, real data provides language study with objectivity, global view, measuring, quantitative aspects, comparing, archiving...) → verification or falsification of intuitions and working hypotheses
- language modeling possible
- learning mechanisms

Communication is not single-track, but a process which combines language with other modalities (gestures, facial expressions etc.). Recent advances in technology allow us to broaden the study of language and explore the study of communication in general – both between humans but also between humans and machines.

In this sense, what the two fields can provide each other is a transfusion of

- topics of study (from language (written, spoken) to communication (other modalities too) to the study of mind and AI
- methodologies and processes (e.g. methodologies used for one field where technology is applied can be used for language study)
- approaches.

In the near future we will encounter more of both language and technology; we will see them both separately (using each other as "infrastructural tool") but we will also see them joining hands to serve other fields (robotics, entertainment, education etc.).

### ***Zygmunt Vetulani***

#### **Human Language Technologies with respect to traditional disciplines**

What I remember from the class of physics is that electricity is generated on the borderline between two different environments. Also, that nothing new may emerge in the ideally homogeneous matter. From the history of thought we learn that new ideas often appear in the confrontation of different civilisations, cultures or traditions. Since relatively recently (some 20 years only) we have been present at the creation of the new emerging discipline: Human Language Technologies (HLTs). Let us try to localise its birth place.

In the spectrum of traditional fields I distinguish:

- Philosophy – which is about the main existential problems situating us with respect the Universe,
- Natural Sciences – about the human-perceptible physical world
- Humanities and Arts - about human activities depending on our intellectual faculties

- Social Sciences – about humans functioning within collectivities,
- Technical Sciences – about designing, producing and exploring artefacts, products or systems

It is hard to classify the HLTs with respect to this list of fields. HLTs is clearly a field emerging somewhere between them.

What I mean by HLTs are (mainly) technologies of interaction between humans and the technological environment, initially created by humans, which became a kind of extension of the natural, physical environment which its own autonomy. Elements of this environment, such as the internet, seems to have now their own identity, highly independent with respect to any individual human being or organisation (in most cases). What is essential for this environment is that it is based on information (information-rich).

This is a new situation as, until recently, the human technological environment was composed of artefacts which were information void (impossible to serve as an interactive information supplier for humans). In this new situation, humans may (and wish) to interact with this information-saturated environment, as they always used to interact with other humans – possible information suppliers. The HLTs are technologies enabling to use human language to implement this new kind of interaction (where some actors may not be humans).