

Capturing Natural Interactions

Nick Campbell, Trinity College, Dublin

CLARIN w/s @ KTH, 25-26 November 2009

Speech recognition and synthesis technologies can now be considered mature, but their simple incorporation into speech-based human-computer interfaces reveals shortcomings in their capabilities. Perhaps the biggest reason for this is that each technology was designed explicitly to convert between text and spoken modalities, without taking into consideration the complexities of human spoken interaction as the joint creation of mutually understood meaning.

If the goal of this meeting is to improve these interfaces and make them more intelligent and human-like, through a better understanding of human interaction and communication, then we might start by designing improved techniques for efficiently capturing, storing, annotating, and distributing large corpora of natural spoken interactions

Speech & Multimodal Databases

*One impulse from a vernal wood / May teach you more of man, / Of moral evil and of good,
Than all the sages can. / Sweet is the lore which Nature brings; / Our meddling intellect
Mis-shapes the beauteous forms of things:- / We murder to dissect.* (William Wordsworth, 1798)

We might discuss at length here about how to collect much more representative natural data.

Annotating, Viewing and Distributing New Data

There are presently several tools for manual annotation of data that each store the results in a prescribed format, easy for dissemination, but my experience of working with these and of talking with people who use them regularly is that the task is tedious, and the framework often restrictive. Rather than prescribe a standard at this time, we might benefit more from creating a support group whereby people who annotate data regularly can communicate and share samples, tools, and formats for rapid assisted evolution. My LREC 2010 paper (submitted) may be relevant here.

Two-Way Dissemination

We should also examine crowd-sourcing in this respect. As with our own FreeTalk corpus (www.speech-data.jp), by making the initial data public and co-operating worldwide with interested partners, the annotations can be grown as researchers with different interests contribute their own layers of knowledge. Since the world of multimodal corpora is still young, perhaps the most we might expect from this November meeting is the opening up of channels whereby the exchange of sources and resources might take place.

*Enough of Science and of Art; / Close up those barren leaves;
Come forth, and bring with you a heart / That watches and receives.* (from *The Tables Turned*)