ASR-MT Interface: Problems and Future Directions

Evgeny Matusov

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University
D-52056 Aachen

Multiple ASR hypotheses for MT

- motivation: take advantage of another knowledge source
  - goal: find best target language translation,
    the exact transcription in the source language is not that important
- strong theoretical base (Bayes decision rule)
- in practice, acoustic/source LM scores or the ASR posterior probability
  are combined with the translation models using log-linear model combination
- representations:
  - N-best lists (simple)
  - word lattice (fits the theory best)
  - confusion network (fits the MT search best)
- multiple research groups working on the problem
  - ATR, CMU, FBK(ITC-irst), IBM, JHU/Cambridge, U.Maryland, RWTH, ...
  - languages: Arabic, Chinese, Italian, Japanese, ...
- so far: small or moderate improvement only

Perspectives of coupling speech recognition and MT

Progress so far (RWTH, in %):

	TC-STAR Es→En	IWSLT Ce→En	IWSLT It→En	LC-STAR Es→Ca
ASR WER, %	12.0	42.0	21.0	32.0
BLEU for single best ASR, %	36.1	33.1	55.1	47.6
BLEU for lattice/CN input, %	36.4	35.1	57.7	53.7

What are the chances for future improvements?

	TER (on correct transcriptions)		
ASR WER	low (0-20%)	middle (20-40%)	high (40-60%)
low (0-15%)	???	+ TC-STAR ES↔EN	− GALE CH-EN
middle (15-25%)	+	+ GALE AR-EN, IWSLT IT-EN	−
high (25-50%)	+ LC-STAR ES-CA	+ IWSLT CH-EN	+ LC-STAR ES-EN

Future Challenges

- sophisticated re-ordering strategies, syntax-based MT
  - is this going to help for speech at all?
  - applicable to confusion networks with low densities
  - open problem for word lattices
- rule-based preprocessing, morphosyntactic analysis/transformations, sentence segmentation, punctuation, etc.
  - modification of existing algorithms is required to handle multiple ASR hypotheses
- scaling to larger tasks
- real-time systems