

Technologies which facilitate AVT

Andrew Lambourne

Overview

- Useful technologies
- Evolving approaches
- Current capabilities
- Remaining challenges
- Wider applications

Useful technologies for AVT

Transcription and alignment for timing

- language models and genres

Translation

- language pairs and genres

Editing

- challenging for informal dialogue

Workflow and productivity tools

- information management and flow

Parallel evolution

Translation

Rule-based systems

Statistical approaches

Neural net approaches

Knowledge-based approach?

New architectures?

Transcription

Formant identification

Statistical approaches

Neural net approaches

Knowledge-based approach?

New architectures?

“Big data” approach

- Translation systems are trained on parallel corpora (eg EU documents, UN documents, specific datasets)
- Transcription systems are based on language corpora (web crawling, news media, medical/legal documents)
- Aim is to build models/nets representing probabilities (to allow most likely hypothesis to be selected)
- Obviates the need for large complex sets of rules
- Makes it harder to find and correct error causes
- Relevant corpora / training time must be available

Domain specialisation

- Recognition results much better in restricted domains: vocabulary pre-preparation is key to quality
- Training on domain-specific texts is leading to claims: “Computers match humans in speech recognition”
- Adapting to a single speaker gives huge benefits: subtitle re-speakers can achieve >99% accuracy
- Restricted-domain parallel corpora can give rise to: “a leap forward; impressive fluency” in translation

Knowledge-based systems

- Semantics enable humans to interpret and disambiguate
- Knowledge of grammar underpins human information transfer
- Context is used to mediate communication

Models:

- *Adaptive discriminating sensors with extensive pre-processing*
- *Syntactic analyser built on wide-context real-world training*
- *Deep semantic extractor and discourse-thread follower*
- *Real-time disambiguator and sub-symbolic cue handler*
- *Extensive associative memory with hierarchical feedback*

New architectures?

- Studies of “how the brain works” lead to new insights: neocortical hierarchical systems, associative memory
- Having new architectures also implies new processing: except that we don’t really know how the brain works
- People are rooted and grounded in the physical world: capable of reasoning; huge breadth of common-sense; intuitions and instincts; can handle the unexpected
- Much of what is currently labelled “Artificial Intelligence” could in fact be regarded as “Simulated Intelligence”

Speaker independent transcription

- Microsoft research using Nat Inst Stds and Tech 2000 database of telephone conversations: some on assigned topic, some chat
- Most of the data used to train neural nets, the rest for transcripts
- Sample recordings sent to professional transcription service: 5.9% errors on defined topic; 11.3% errors on family conversation
- Convolutional multi-layer neural net transcription: 5.9% errors on defined topic; 11.1% errors on family conversation
- Errors between humans and computers were different: machine had more problem with back-channel like “uh-huh”

(ref <https://arxiv.org/abs/1610.05256> October 2016)

Machine Translation performance

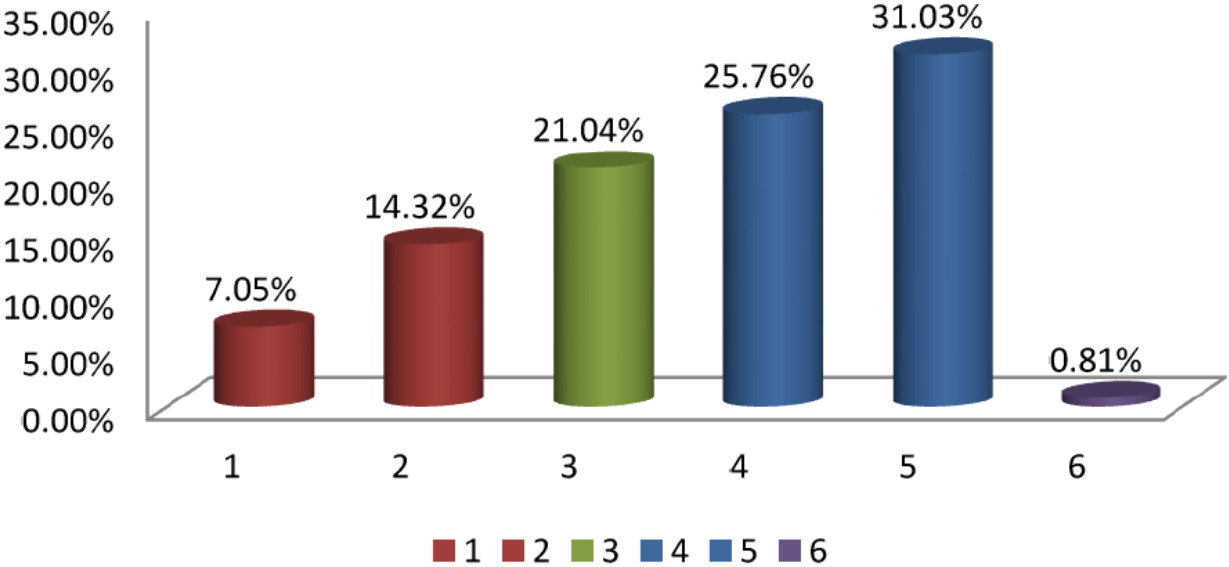


Figure 5 Rating distribution in evaluation round 1

1=unintelligible, 2=50-70% needs editing, 3=25-50% needs editing, 4=10-25% needs editing, 5=MT output perfectly usable

Image credit:
SUMAT project
final report Jun 2014

www.sumat-project.eu

Natural Language remains a challenge

- Computer vision and natural language are “strong AI” problems
- Computers can perform tasks that are complicated for people, but not tasks that are easy for people
- Difficult to translate into morphologically rich languages
- Certain professions (legal, patent, medical, technical writers) use natural language in a more structured and constrained way
- SLT works better in these constrained domains:
single speaker, medical/legal transcription, technical translation
- In the media area, there is an even more unhelpful trend:
pollution of perfectly clear dialogue with music and effects!

Understanding the limitations of NLP

- Best illuminated by considering some areas of ongoing research:
 - automatic summarisation
 - coreference (what refers to what) and agreements
 - discourse analysis
 - segmentation of words into morphological components
 - named entity recognition
 - transforming text to knowledge structures
 - sentiment analysis
 - topic identification
- The last part of the journey is likely to be the hardest!

Thank you

a.d.lambourne@leedsbeckett.ac.uk

