The Past, the Present, the Future of Spoken Language Translation Evaluation

Marcello Federico
FBK, Trento, Italy

InterACT25, 14-15 July 2016, Baden Baden, Germany
What is IWSLT?

- International Workshop Spoken Language Translation
  - Annual event since 2004 (12 editions)
  - Evaluation Campaign + Technical Workshop
- Organisers
  - Chairs: KIT (Germany), FBK (Italy), CMU (USA)
  - International steering committee
- Important activities
  - WIT3.fbk.eu website with parallel data of TED talks
  - Evaluation methods
Core Tasks

- **Automatic Speech Recognition (ASR)**
  - English *speech* → English *transcript*

- **Machine Translation (MT)**
  - English text → French text

- **Spoken Language Translation (SLT)**
  - English *speech* → French text
  - English *transcript* → French text
Son of the STARs

from demonstrations to open evaluations

demonstration systems + internal evaluations

integration + open evaluations
Is this a C-STAR or a IWSLT dinner?
Son of the STARs

from demonstrations to open evaluations

demonstration systems + internal evaluations

integration + open evaluations
TC-STAR – (Seville, 2003)

We should evaluate all our systems.

...and make them accessible via UIMA.

Couldn’t we just simply evaluate?
IWSLT participants from 70 different teams
TED Talks

TED LLC is non-profit

- Two annual events
- Short talks
- Variety of topics
- Website with:
  - Videos
  - Transcripts
  - Translations
  - CC License
Human task: subtitling and translating

- segment audio
- transcribe and annotate
- split into captions
- translate captions

e come sarebbero potuti essere automatizzati il più possibile.

et comment ils pourraient être aussi automatisés que possible.

and how they could be automated as much as possible.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Talks (EN)</td>
<td>800</td>
<td>1,080</td>
<td>1,395</td>
<td>1,650</td>
<td>1,875</td>
<td>2,095</td>
</tr>
<tr>
<td>Languages</td>
<td>80</td>
<td>83</td>
<td>93</td>
<td>103</td>
<td>105</td>
<td>109</td>
</tr>
<tr>
<td>Translators</td>
<td>4,000</td>
<td>6,823</td>
<td>8,382</td>
<td>11,010</td>
<td>18,699</td>
<td>15,487</td>
</tr>
<tr>
<td>Translations</td>
<td>12,500</td>
<td>24,287</td>
<td>32,707</td>
<td>49,607</td>
<td>65,290</td>
<td>83,265</td>
</tr>
<tr>
<td></td>
<td>+94%</td>
<td>+34%</td>
<td>+52%</td>
<td>+32%</td>
<td>+28%</td>
<td></td>
</tr>
</tbody>
</table>
Challenges

- **Language modelling**
  - Limited in-domain training data
  - Variability of topics and styles

- **Acoustic modelling**
  - Speaker: accent, fluency, speaking rate, style, ...
  - Noise: some mumble, applauses, laughs, music, ...

- **Translation modelling**
  - Distant and under-resourced languages
  - Morphologically rich languages

- **Speech Translation**
  - Detection and removal of non-speech events
  - Subtitling and translating
Best Measured MT Performance
Progress in ASR En (best systems WER%)
Best MT Performance over Time

![Graph showing the best MT performance over time with lines for EN-DE, DE-EN, RU-EN, and EN-ZH, indicating a steady increase in DELTA% from 2012 to 2015.]
Best MT Performance over Time

![Line Graph showing Best MT Performance over Time with data points for EN-FR, AR-EN, and ZH-EN]
Human Evaluation

- **Pairwise ranking** [2010-2012]
  - *native volunteers judge quality of MT outputs*
  - pros: relatively cheap
  - cons: no re-use of data

- **Post-editing effort** [2013-2015]
  - *professional translators post-edit MT outputs*
  - cons: more expensive (~7 cents/word)
  - pros: multiple references as by product
  - pros: permits profiling of MT systems
Skype Translator Task
Challenges

- **Language modelling**
  - Limited in-domain training data
  - Variability of topics and styles

- **Acoustic modelling**
  - Speaker: accent, fluency, speaking rate, style, ...
  - Noise: spontaneous speech, disfluencies

- **Translation modelling**
  - Morphologically rich languages

- **Speech Translation**
  - From spontaneous speech to polished text
  - *Simultaneous translation*
IWSLT 2016
Seattle, US
December 08-09, 2016
www.iwslt.org

International Workshop on Spoken Language Translation
Credits

- **Language resources**
  - TED LLC, USA (Talk data)
  - Workshop Machine Translation (Giga and news data)
  - DFKI, Germany (United Nations data)
  - Microsoft Research (Skype Translator Task)
  - Qatar Computing Research Institute (QED Lectures)

- **Funding**
  - H2020 CSA CRACKER
  - H2020 QT21

**Questions?**