EU-BRIDGE
Bridges Across the Language Divide

Sebastian Stüker, KIT
Not this project, + different architecture

![BABEL FISH from "The hitch-hiker's Guide to the Galaxy" (TV series) 1981 © BBC orig. animation artwork by Rod Lord]
It’s Happening Right Now

Speech and language technologies (posing hard problems)

A strong need, in particular in Europe

Availability of resources

Growth of computing power/memory

Scientific progress

A strong need, in particular in Europe

Availability of resources

Growth of computing power/memory

Scientific progress

Speech and language technologies (posing hard problems)
The Vision

• Bridges Between Languages in Europe:
  • Technology to bridge the European language
  • Building/maintaining a sustainable European language infrastructure
  • Language divide is expressed in language and speech
  • Easily accessible services

• Bridges Between Science and Society
  • Language technology requires continued scientific attention
  • Exploitation and insertion requires suitable adaptation
  • Metrics:
    • WER, BLEU, TER, F-Scores
    • User Friendliness, Productivity, Sales, Distribution Channels, Customer Support
  • Identify use cases and applications
  • Effective transition and insertion
EU-BRIDGE in a Nutshell

• Here: Not science fiction, but results coming soon (2014/2015)
• FP7 IP EU-BRIDGE: Bridges Across the Language Divide
• Development of a speech translation service infrastructure
• Targeted to make progress in particular regarding market insertion

• Project footprint:
  • Feb 2012 - Jan 2015
  • Budget € 10.5m, EC funding € 7.8m
  • 10 partners
  • 1 service infrastructure, 4 use cases
Goals and Project Plan

Four use cases
1. Captioning and translation of subtitles for TV programs
2. Simultaneous translation of academic lectures
3. Speech translation services for the European Parliament
4. Translation of webinars

Four major objectives
1. Performance
2. Language portability
3. Reduction of dependency on data
4. Rapid technology transition and market insertion
EU-BRIDGE Partners

Karlsruher Institut für Technologie

FONDAZIONE BRUNO KESSLER

RWTH AACHEN UNIVERSITY

RED BEE

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

PerVoice

Andrexen

End-to-End Unified Communications

ACCIPIO PROJECTS
Language Service

Engines

Services

Use Cases

Use Case 1

Use Case 2

Use Case 3

Language Services for User and Developer Communities

Develop and Insert Improved Technology

Customization, Adaptation

ASR

MT

Sebastian Stüker: EU-BRIDGE — Advanced text and speech translation services for educational applications

www.eu-bridge.eu
Language Service

Engines | Services | Use Cases
---|---|---
ASR | | Cross-Lingual Captioning
| MT | | Use Case 1

Use Case 1
Lecture Translation
Use Case 2
Webinar Translation
Use Case 3

24 May 2012
Sebastian Stüker: EU-BRIDGE *** Advanced text and speech translation services for educational applications
www.eu-bridge.eu
Technical Challenges

Languages vs. Domain

Robustness (Speaking Style Noise)

- EuroMatrixPlus
- C-STAR
- TC-STAR
- Babel
- GAIA
- Verbmobil
- Nespolo
- Verbmobil
- Clean, read
- Planned
- Spontaneous
- Informal Comm.
- Far distance
- Close talking
- Domain
- Domain unlimited
- Domain adaptive
- Small domain
## Project Organization

<table>
<thead>
<tr>
<th>Laboratory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
</tr>
<tr>
<td>WP 1</td>
</tr>
<tr>
<td>WP 2</td>
</tr>
<tr>
<td>WP 3</td>
</tr>
<tr>
<td>WP 4</td>
</tr>
<tr>
<td>WP 5</td>
</tr>
<tr>
<td>WP 6: Evaluation</td>
</tr>
</tbody>
</table>

### WP 1: Transcription and Translation
- **Research:** Transcription and Translation
- **Engines:** Deployable Engines
- **Services:** Transcription and Translation Services
- **Products:**
  - Caption Transl.
  - Lecture Transl.
  - Webinar

### WP 2: Data Dependency
- **Research:** Data Dependency
- **Engines:**
- **Services:**
- **Products:**

### WP 3: Standarized, Inhouse Tests
- **Research:**
- **Engines:**
- **Services:**
- **Products:**

### WP 4: External Campaigns
- **Research:**
- **Engines:**
- **Services:**
- **Products:**

### WP 5: Market Insertion
- **Research:**
- **Engines:**
- **Services:**
- **Products:**

### WP 6: Evaluation
- **Research:**
- **Engines:**
- **Services:**
- **Products:**
Task 3.1 ASR and MT
Task 3.2 Multithreaded+Stream Decoding
Task 3.3 Language Dependent Linguistic Processing
Task 3.4 LID
Task 3.5 Text Processing

Task 4.1 Core Services: ASR, MT, Text Processing
Design and implementation of core service APIs and integration of core engines
Task 4.1 Core Services: ASR, MT, Text Processing
Design and implementation of core service APIs and integration of core engines

Task 4.2 Running Services in the cloud
Design and implementation of client-server architecture + APIs

WP 4
Services

WP 5
Products

Task 5.1
Caption Translation

Task 5.2
Lecture Translation

Task 5.3
Multilingual Tele- and Video Conferencing
Evaluations

“You Improve what you Evaluate”

- Evaluations organized around Use Cases
- Align Metrics with Product/Service Goals

Co-opetition

- Partners engage in friendly competition
- Goal is progress/complementarity, not site-to-site “horse-race”
- Partners participate with standard or optimized/tuned systems
Evaluations

• Standard open benchmarks
  • Calibrate technology internationally in open competition
  • IWSL, WMT: cover lecture task and multilingual MT
  • EU-Bridge partners are (co-) organizers of these evals and can thus influence the process to suit EU-BRIDGE’s needs
  • Don’t need to create/market a new campaign

• Internal EU-Bridge Tasks
  • Need to evaluate & optimize technology for EU-BRIDGE use cases
  • Different goals/sub-goals pursued (e.g. captioning, NE, shortening, ..)
  • Evaluals/Assessment need to change frequently during the project
  • Comparisons with outside teams not needed and not helpful

• Field Tests and Measures:
  • Evaluate on Living online ‘Organism’
  • Optimize for extrinsic measures, not intrinsic ones
IWSLT 2014

- Evaluation of Speech Translation Technology on Talks and Lectures
- Working on TED data, because of efficiency of creating training and evaluation data
Use Cases

Captioning and Translation of Multimedia Content
  • BBC Weather Data
  • Euronews
  • Skynews

Multilingual Lectures, Meetings
  • TED Lectures
  • University Lectures
  • European Parliament Voting Sessions

European Parliament Interpreter Support
  • Terminology Support: Tool Field Tests
  • Named Entities: Eval, Tool Integration & Tests

Webinar Translation
  • Integration into the Andre xen platform
Use Cases

Captioning and Translation of Multimedia Content
- BBC Weather Data
- Euronews
- Skynews

Multilingual Lectures, Meetings
- TED Lectures
- University Lectures
- European Parliament Voting Sessions

European Parliament Interpreter Support
- Terminology Support: Tool Field Tests
- Named Entities: Eval, Tool Integration & Tests

Webinar Translation
- Integration into the Andrexen platform
Lecture Translation

• Continuous Monologue

• Speaking-Style
  • Planned, not read, not memorized
  • Fast, spontaneous, fragmentary, and no punctuation!
  • Noises, coughing, non-verbal events (e.g., singing)

• Vocabulary
  • Very large
  • Specialized terms
  • Foreign Words

• Speed, Real-time

• Service-Infrastructure
  • Many parallel lectures;
  • Automatic, robust assignment of compute power
Speech-Translation Systems

- Human Interpreters are too expensive
- Automatic Speech Translation an affordable solution:
  - Still lots of errors, room for improvement
  - But, better than nothing
Service Infrastructure

Client -> Mediator/Load balancer -> Client

Client -> Mediator/Load balancer -> Client

Client -> Mediator/Load balancer -> Client

Worker

Worker

Worker

Service request

Distribution

Process request

Receive results

Combination/Re-distribution

Return results

Mobile Devices

Web Browsers

Loudspeakers

ASR

MT

TTS

Sebastian Stüker: EU-BRIDGE *** Advanced text and speech translation services for educational applications

24 May 2012

www.eu-bridge.eu
Mehr als nur Bahnhof verstehen - Weltweit erster Vorlesungübersetzer

Deployment
Data Resources

KIT lecture corpus:

• Collected to fit the needs for training the systems:
  • ASR AM: Large amounts of in-domain audio data with careful transcriptions
  • MT Translation Model: Parallel sentences of in-domain data for the translation
  • ASR+MT LM: Large amounts of monolingual sentences in the required domain
  • Any kind of meta data that might help (e.g., lecturers’ slides)

• Data collection took place at KIT’s lecture halls:
  • Started with computer science lectures
  • Gradually spread to lectures from all departments at the university
Data Resources

- Presenter Notebook
- Slide grabbing with Timestamps
- Videograbber
- Beamer
- Microphone 1 (M1) + Emitter
- Microphone 2 (M2)
- Microphone 3 (M3)
- Emitter for M2, M3 (x2)
- Receiver (x3)
- Videocamera
- Soundcard (x2)
- Data Repository
- Assistant (collecting Data)

Video
Audio
Slides and Timestamps
Automatic Speech Recognition

- Janus based ASR system: HMM/GMM based quinphone system with 4,000 models, MVDR front-end, 4gram LM
- Acoustic Model:
  - Trained on all lecture data in order to get a speaker independent model
  - Created 5 speaker adapted AMs: speaker independent model + Viterbi training and bMMIE training on the speaker dependent data
- Language Model:
  - Interpolation from 28 text corpora
  - Interpol. weights tuned on random selection from AM training data
  - 300k vocab selected by ML count estimation method
- German has a lot of compounds:
  - Sub-word vocabulary for compounds
- Vocabulary adaptation by deriving queries from slides (OOV 2.25% → 0.75% at 300k vocab size)
ASR Performance

- Tested the ASR system on the dev set of six lecturers

<table>
<thead>
<tr>
<th>Lecturer</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker Independent AM</td>
<td>34.8%</td>
<td>21.1%</td>
<td>28.4%</td>
<td>22.9%</td>
<td>22.7%</td>
<td>19.0%</td>
</tr>
<tr>
<td>Speaker Dependent AM</td>
<td>–</td>
<td>18.9%</td>
<td>27.6%</td>
<td>22.6%</td>
<td>21.5%</td>
<td>17.8%</td>
</tr>
<tr>
<td>Adapted LM+Vocabulary</td>
<td>23.9%</td>
<td>17.3%</td>
<td>–</td>
<td>18.1%</td>
<td>–</td>
<td>15.4%</td>
</tr>
</tbody>
</table>
OOV Reduction

Vocabulary size vs. out of vocabulary rate for different vocabulary models:
- Baseline
- Query Vocab
- Subword Vocab
- Subword + Query Vocab

24 May 2012
Sebastian Stüker: EU-BRIDGE: Advanced text and speech translation services for educational applications
www.eu-bridge.eu
ASR Post-Processing

- Numbers are converted to digit sequences
- Common symbols are substituted, e.g., “Prozent” → ’%’
- Punctuation, i.e. ‘.’ and ‘,’ is inserted:
  - Prediction via a 4gram model
  - Pause information used to adjust the priors
- Simple, rule based conversion of equations:
  - “F of x” → ‘f(x)’
- Punctuation is used to structure text into sentences and chunks for translation
Machine Translation

Statistical Phrase Based MT System

- Trained on different out-of-domain data and the lecture corpus
- Applied the same compound word splitting as for the ASR system for consistency; names etc. were excluded from splitting by applying a named entity tagger
- Used a discriminatively trained word alignment approach
- Specific models for short and long range reorderings (also on the training data)
- Online system:
  - Phrase table filtered with ASR vocabulary
  - Simplified POS tagging
Machine Translation

Adaptation to the lecture domain

- Domain independent translation model trained on all data
- In-Domain TM only on the lecture data (re-use alignments)
- Combined via log-linear combination
- For LM log-linear combination of large LM, in-domain LM and TED LM
- Translations or special terms learned from Wikipedia and Wiktionary
### MT Results

<table>
<thead>
<tr>
<th>Lecturer</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecturer 1</td>
<td>13.80</td>
</tr>
<tr>
<td>Lecturer 2</td>
<td>22.58</td>
</tr>
<tr>
<td>Lecturer 3</td>
<td>14.24</td>
</tr>
<tr>
<td>Lecturer 4</td>
<td>20.83</td>
</tr>
<tr>
<td>Lecturer 5</td>
<td>24.50</td>
</tr>
<tr>
<td>Lecturer 6</td>
<td>24.13</td>
</tr>
</tbody>
</table>
Thoughts on the Output Modality

Text instead of synthesized speech

• Text can be easily distributed over the WWW
  • Laptops, smartphones, tablets
  • Nowadays ubiquitous
  • No proprietary software, just a browser

• Listening to synthesized speech can be very tiresome
  • Artificial voices not perfect
  • Original speech present in addition

• Translation system commits errors
  • Translated text contains errors; synthesis quality suffers from that

• Temporal Navigation
  • Once translation has been heard, it is gone
  • Text enables users to move back and forth in the translation; supports understanding the translation
Interface

Client for the lecturer:

• Must be as simple as possible!
• Client needs to know:
  • Who is speaking?
  • What lecture is it?
• Lecturer needs to know:
  • Is it turned on?
  • Is it doing something?
Interface

Displaying the results:

- Translation result + ASR result (?!)
- Scrolling back and forth in time
- Make text readable !!!

Speaker 1
Vorlesung 5: Kognitive Systeme


That means, you also need to be able to interpret this statement, and then a. Formulate question. We also have briefly mentioned. And that means, in the end it is the content of this statement to understand.
Where do we go?

• Interfaces need to get simpler!

• Systems need to become an omnivore:
  • Get all meta information as early as possible (before the lecture!)
  • Slides, papers, web sites, text books, etc.
  • Find a way for obtaining comparable corpora

• Make system self maintaining and autonomous
  • Automatic unsupervised acoustic model adaptation after every lecture
  • Automatically detect speaker, lecture and language: Access the university information system

• Offer additional services
  • Archive of the lectures for the students (for search)
  • Translation of the slides
  • Summary of the lecture

• Get the students in the loop
  • Automatic corrections by the students: during the lecture and afterwards
  • Make it a game
Andreessen - Webinar

• Role definition (Speaker / Listener)
  • Functionality was defined and implemented, and relevant MCloud services integrated

• Data flow (Architecture / Slideshow / Voice / Text)
  • Features implemented to support the parallel transmission of slides, speaker audio, translation text and intelligent user location to provide integrated UC experience
Andrexen - Webinar

Adapting to industry requirements:

- Real time streaming

- Support legacy systems:
  - Training narrow bandwidth (8KHz) acoustic model

- Specific vocabulary:
  - Vocabulary and language model adaptation
Andrexen - Webinar

- Translator integrated as virtual participant
- Is “chatting” the translation results