# Leading by Example:

## Automatic Translation of Subtitles using Example-Based Machine Translation

Colm Caffrey, Marian Flanagan and Stephen Armstrong

Centre for Translation and Textual Studies &
National Centre for Language Technology
Dublin City University

# Outline

- Project background
- Example-based Machine Translation
- Corpus description
- Evaluation of output
- Demo: DVD clips with subtitles
- Future avenues

“Im Fernsehen reißt der Strom immer neuer Movies nicht ab, aber spielt nicht immer nur der eine ewig gleiche Film, zum Mitsingen, zum Mitgähnen, zum Totschlagen der Zeit?”

Von Festenberg, 2006

# Project Background

- MovieTrans
- Example-Based Machine Translation (EBMT) system
  - Interdisciplinary project
    - Centre for Translation and Textual Studies (CTTS)
    - National Centre for Language Technology (NCLT)
- Enterprise Ireland one year funded “Proof of Concept” phase
- Based on preliminary study (O’Hagan 2003)
- Studies using Rule-Based Machine Translation (RBMT)
- Develop a CAT tool to aid subtitler, NOT replace
- First time EBMT has been used with subtitles
- Currently no commercially available EBMT system
- Goal: increase throughput

“The price wars are fierce,
the time-to-market short,
the fears of piracy rampant.”
Carroll, 2004

“Given limited budgets and an ever-diminishing time-frame for the production of subtitles for films released in cinemas and on DVDs, there is a compelling case for a technology-based translation solution for subtitles.”
O’Hagan, 2003

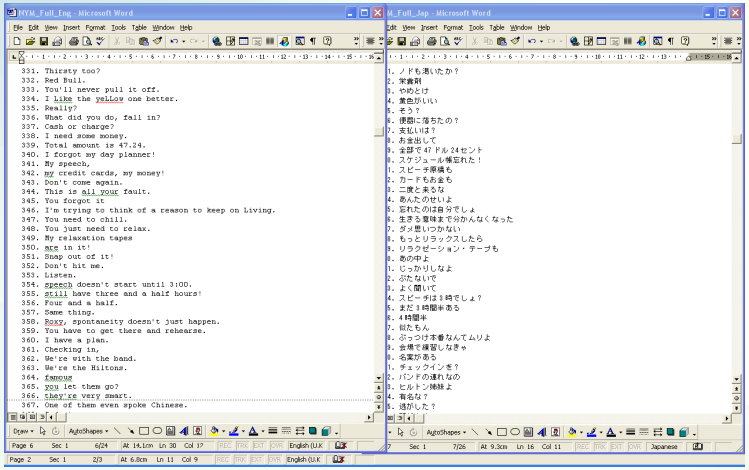
# Example-Based Machine Translation(2)

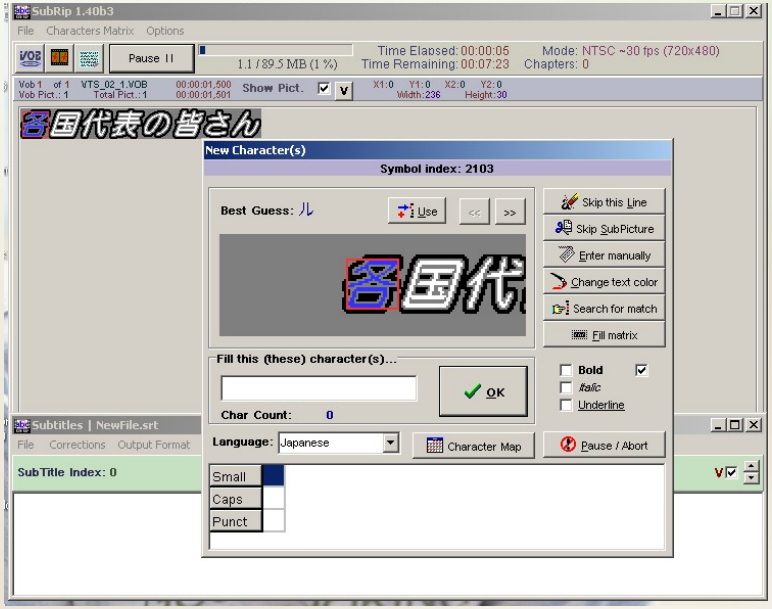
Example taken from (Somers, 2003)

| | | |
|---|---|---|
| The man ate a peach | ⇔ | 人は桃を食べた<br>hito ha momo o tabeta |
| The dog ate a peach | ⇔ | 犬は桃を食べた<br>inu ha momo o tabeta |
| The man ate the dog | ⇔ | 人は犬を食べた<br>hito ha inu o tabeta |

# Corpora

- Homogeneous and heterogeneous
- EBMT – Custom-made corpus
- Comparison between two sets
  - First with EBMT output
  - Two sides of argument
- Corpus creation issues
  - Format of subtitles on DVD
  - Japanese DVD availability
  - OCR on Japanese subtitles





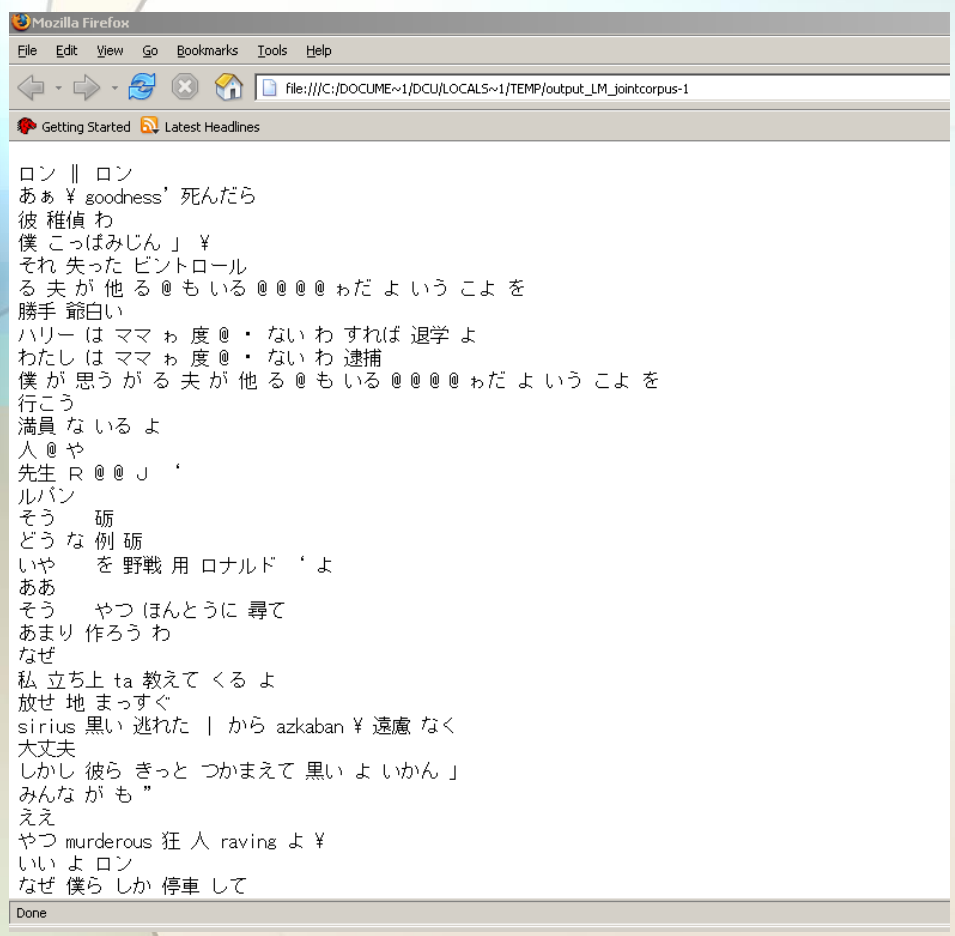
# German Corpora

- Homogeneous
  - DVD Subtitle Corpus
    - 40,000 sentence pairs
  - DVD Bonus Material Corpus
    - 10,000 sentence pairs
- Heterogeneous
  - Europarl Corpus (Koehn, 2005)
    - >1,000,000 sentence pairs

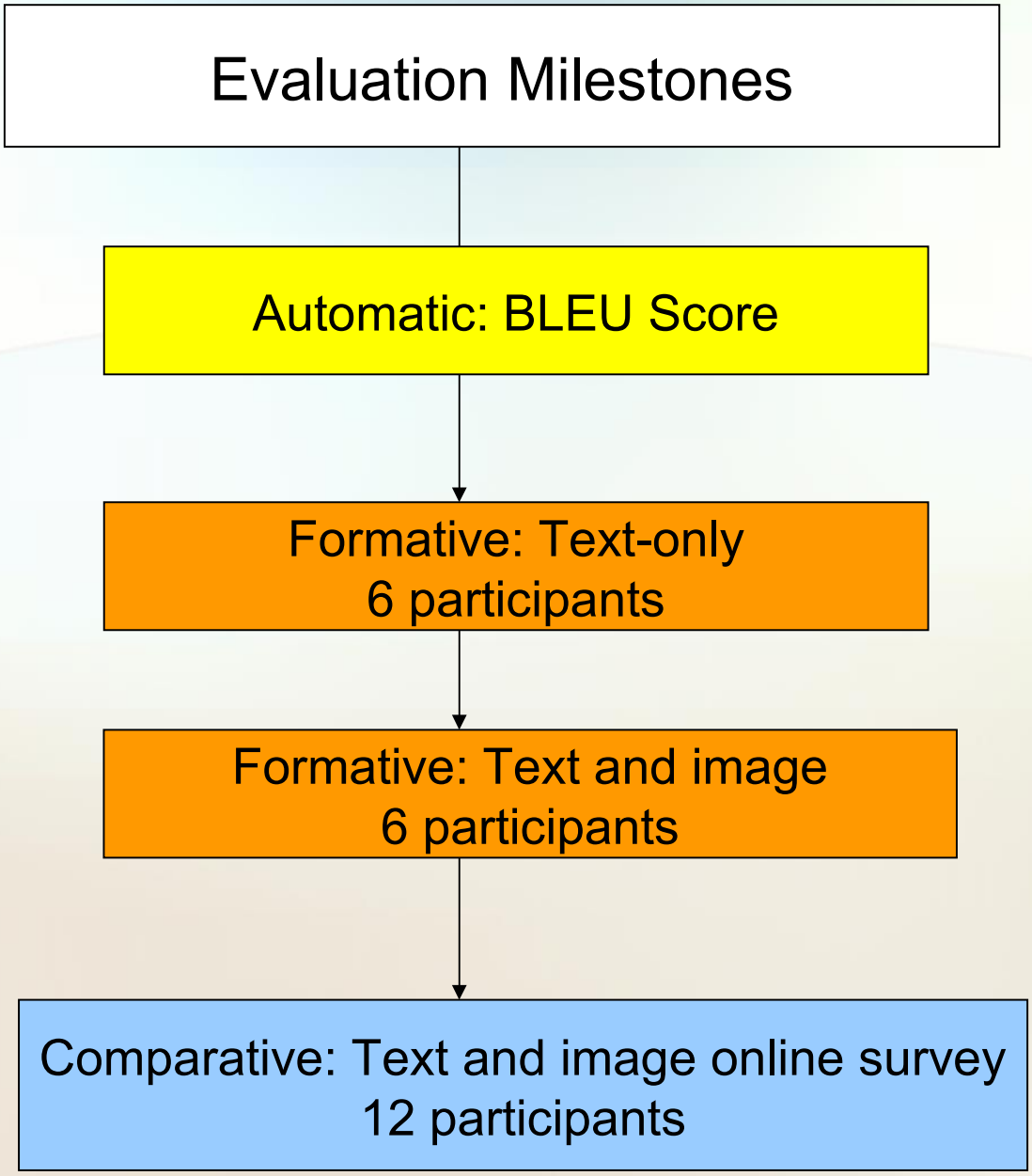
# Japanese Corpora

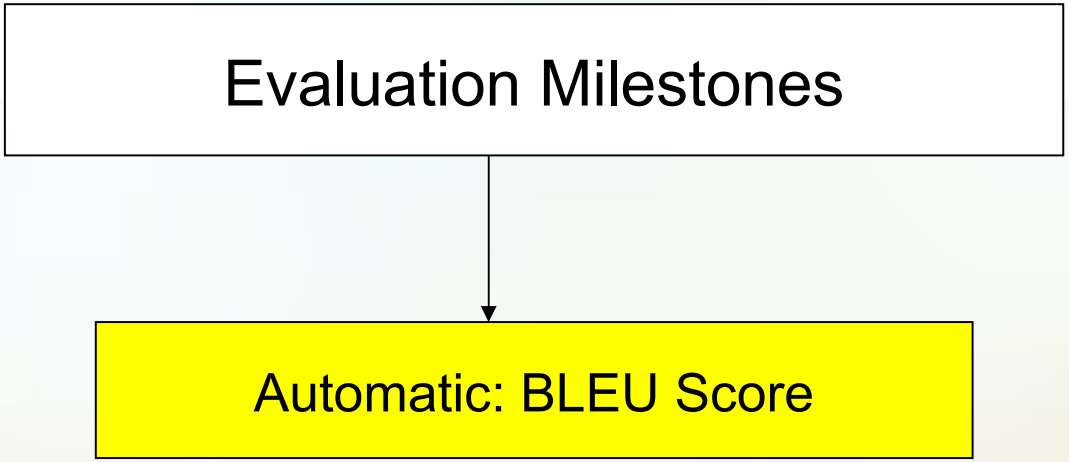
- Homogeneous
  - DVD Subtitle Corpus
    - 12,700 sentence pairs
    - 124,012 characters
- Heterogeneous
  - Book and Article Corpus (Utiyama, 2003)
    - 82,805 sentence pairs
    - 2,624,850 Japanese characters

# Evaluations




- Variety of methodologies
- German output
  - Automatic metrics
  - Human evaluation
- Japanese output
  - Very early output
  - Research design
    - German template
- Pros and cons of both
- Evaluation issues


Evaluation Milestones
Automatic: BLEU Score
Formative: Text-only
6 participants
Formative: Text and image
6 participants
Comparative: Text and image online survey
12 participants


Evaluation Milestones
Automatic: BLEU Score

# MT Automatic Evaluation Metrics

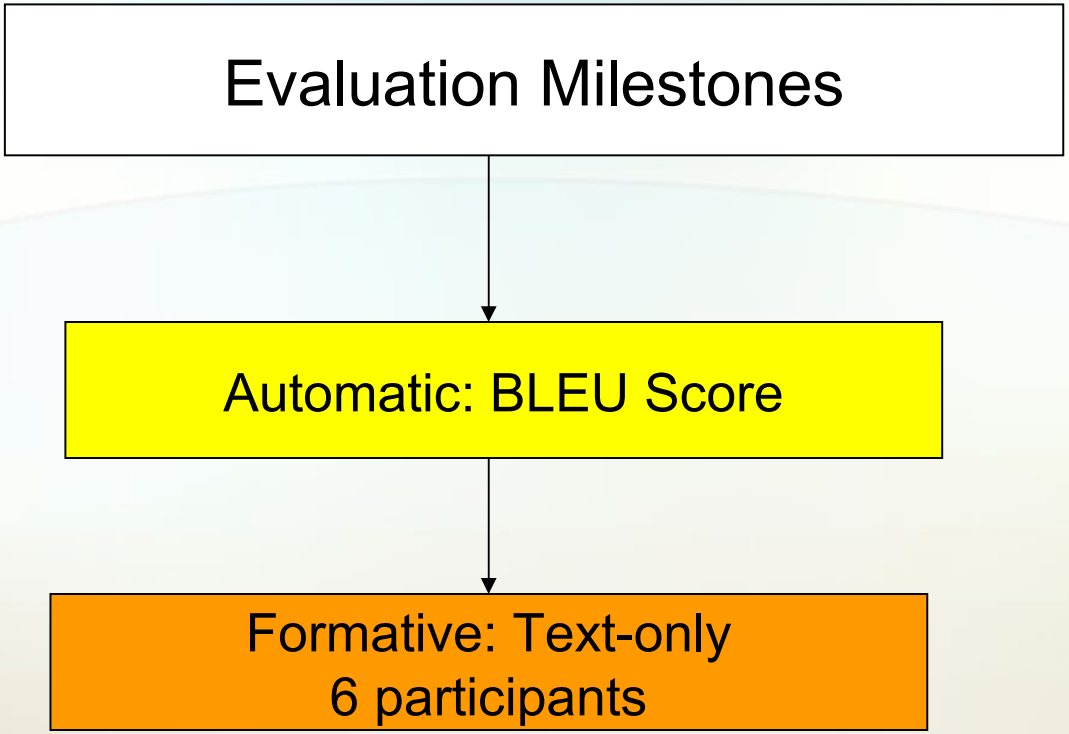
- MT community evaluation standard
- Measure large amounts of data
- Can be controversial, human evaluation addresses this
- BLEU automatic metric
  - Range 0 – 1
  - Human translation as ‘Gold’ standard
    - Length
    - Word strings

# BLEU Score Results

| NUMBER OF SENTENCES | HOMOGENEOUS DATA SCORE | HETEROGENEOUS DATA SCORE |
|---|---|---|
| **10,000** | **0.07** | **0.04** |
| **20,000** | **0.11** | **0.07** |
| **30,000** | **0.18** | **0.10** |
| **60,000** | **not enough data** | **0.13** |

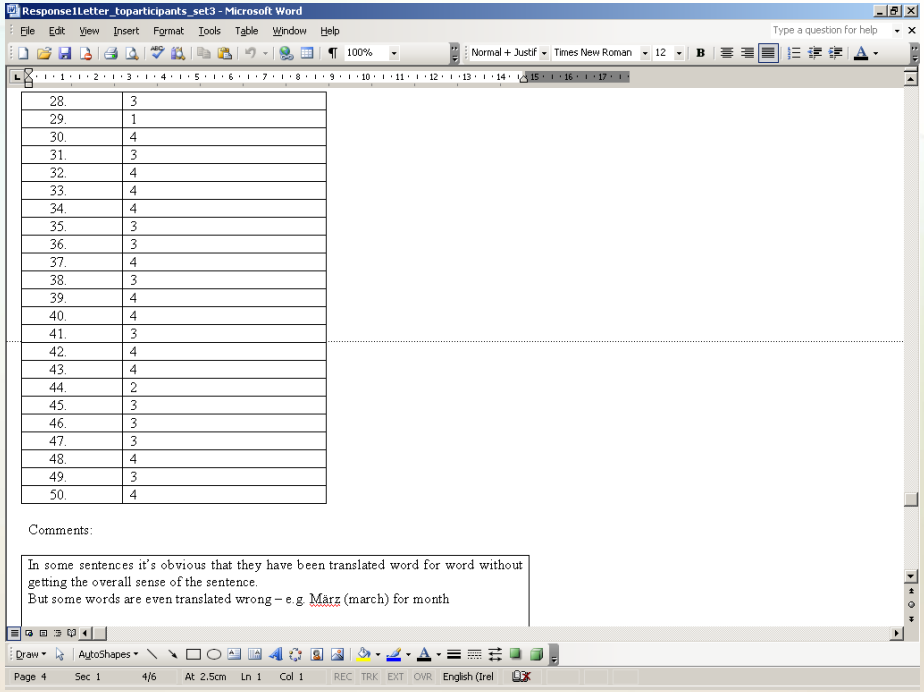
# Human Evaluation Methods

- Formative Evaluation
  - Text Only Evaluation
  - Audiovisual Evaluation (Face-to-Face)
- Comparative Evaluation
  - Audiovisual Evaluation (Online)


Evaluation Milestones
Automatic: BLEU Score
Formative: Text-only
6 participants

# Text-only Evaluation

- In-depth analysis
- 200 random sentences
- Harsh, lack context
- 6 participants
- Asked to score on
  - Intelligibility
  - Accuracy
- Comment on Output



# Intelligibility Scale (Wagner 1998)

1. gut verständlich
2. ziemlich verständlich
3. schwer verständlich
4. unverständlich

# Accuracy Scale (Wagner ibid.)

1. Satz übermittelt die Bedeutung des englischen Satzes

2. Satz übermittelt im großen und ganzen die Bedeutung des englischen Satzes

3. Satz übermittelt die Bedeutung des englischen Satzes nicht angemessen

4. Satz übermittelt die Bedeutung des englischen Satzes nicht

# Errors in Lexicon – Missing words

| | |
|---|---|
| English Original Subtitle | to dig it twice as wide. |
| German Original Subtitle | Doppelt so groβ graben. |
| German EBMT Subtitle | An dig sie zweimal ace wide. |

# Errors in Chunking

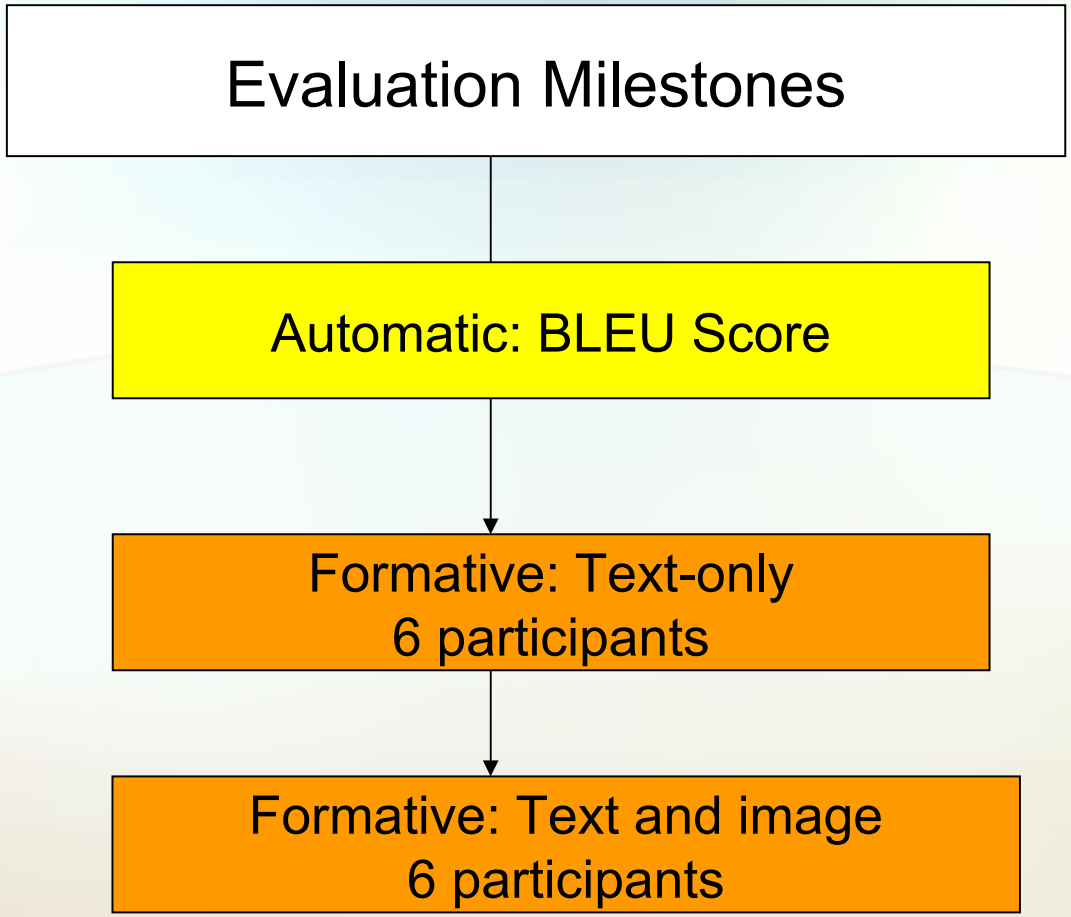
| | |
|---|---|
| English Original Subtitle | Now, Mr. Ewell, can you... |
| German Original Subtitle | Nun, Mr. Ewell, können sie... |
| German EBMT Subtitle | Nun, ich habe, können sie... |

# ‘Creativity’ of EBMT system

| | |
|---|---|
| English Original Subtitle | Shh, shh, shh!<br>Alright children, now quiet. |
| German Original Subtitle | Okay, Kinder, nun seid ruhig. |
| German EBMT Subtitle | Scht, scht, scht!<br>Gut Kinder, mehr Ruhe! |

# Results of Text-only Evaluation (1)

- Positive:
  - Small chunks and short sentences
  - ‘Creativity’
- Negative:
  - Vocabulary gaps
  - Word order


Evaluation Milestones
Automatic: BLEU Score
Formative: Text-only
6 participants
Formative: Text and image
6 participants

# Face-to-Face Evaluation

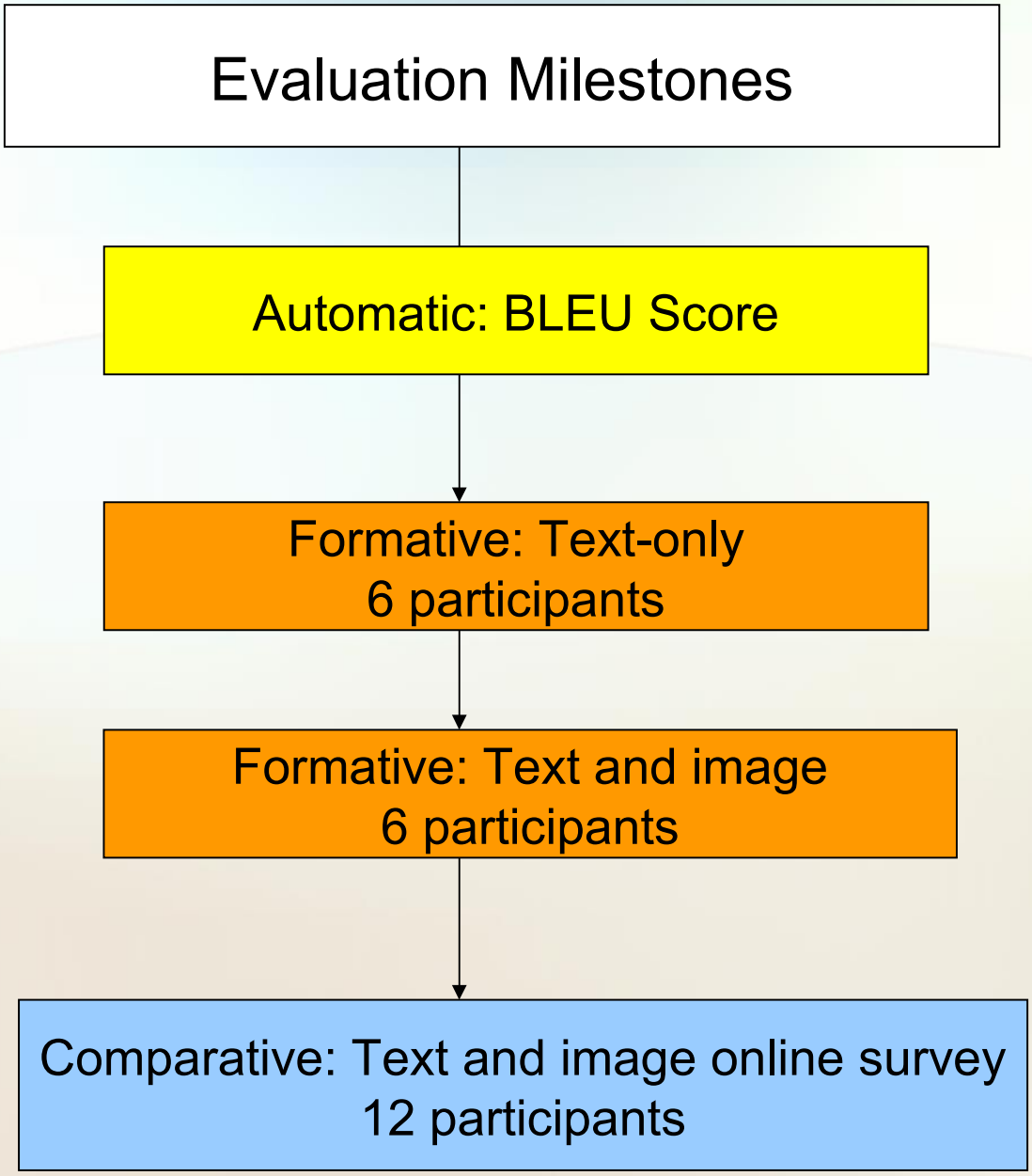
- 6 Participants (3 female 3 male)
- 45 Minute duration
- Holistic approach
  - Shown Text and Image
  - 6 Clips
  - 10 Categories: retrospective interview

# Retrospective Interview Topics

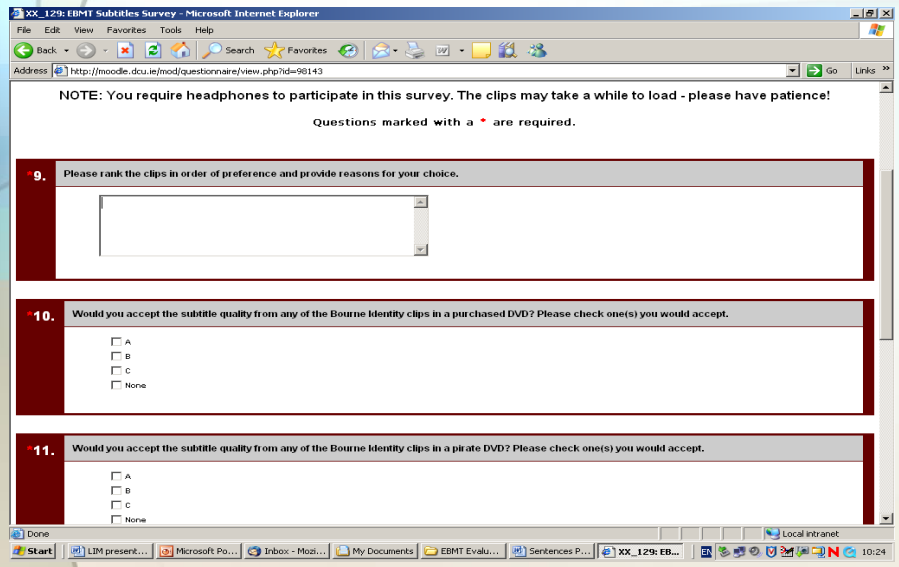
1. Background Information
2. Attitudes to Subtitles and Translation Technology
3. Influence of Source/Target Language Knowledge
4. Overall System Performance
5. Image Congruity
6. Subtitle Appearance
7. Subtitle Cues
8. Video Quality
9. Translation Technology
10. Commercial Avenues

# Results of Face-to-Face Evaluation (2)

- English words not translated
- No time to reread when errors occur
- Raw EBMT output – unacceptable, as expected
- Post-edited – possibly acceptable


Evaluation Milestones
Automatic: BLEU Score
Formative: Text-only
6 participants
Formative: Text and image
6 participants
Comparative: Text and image online survey
12 participants

# Online Evaluation




- Comparative Evaluation
- German-speaking participants: native & non-native
- Survey available on university virtual learning environment (Moodle)
- Pilot Study

# Sample Clips

- Three clips
  - Unedited EBMT output
  - Babelfish output
  - Post-edited EBMT output
- Post-edited by non-native speaker
  - 20 minute time-limit
- Harry Potter and the Prisoner of Azkaban

# Results of Evaluation (3)

I did not like it. I understand it, but it bothers me that there are no points after the sentences and that the nouns and the beginnings of the sentences are not capitalized. – Raw EBMT

- 12 Responses
  - German native speakers
  - English native speakers
- Best rated: EBMT post-edited

Acceptable and understandable, still evidently not done by native speaker. – Post-edited EBMT

- Positive results in relation to post-edited output – confirms research direction

Some slang translations were quite good. – Post-edited EBMT

- Babelfish scored better than raw EBMT, but Systran developed system

# Future development of system

- System tweaking
  - Punctuation and capitalisation
- Customisable MT
  - ‘Plug-in’ Personalised/Serialised Corpora
  - Genre-specific dictionaries
- Commercial Applications
  - In-flight film, Festival film
  - Server Architecture

# References

**Carroll, M.** (2004). Subtitling: Changing Standards for New Media [Online]. Available from: <http://www.translationdirectory.com/article422.htm> [Accessed January 2006].

**Koehn, P.** (2005). A Parallel Corpus for Statistical Machine Translation. In Machine Translation Summit X (pp. 79-86). Phuket, Thailand.

**O'Hagan, M.** (2003). Can language technology respond to the subtitler's dilemma? - A preliminary study. *IN*: Translating and the Computer 25. London: Aslib

**Utiyama, M.** and **Takahashi, M.** (2003) English-Japanese Translation Alignment Data. http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html

**Von Festenberg, N.** 2006. „Du willst es doch auch“. *Der Spiegel.* 29 May, pp140-141.

# Thank you!

Any Questions or Comments are Welcome.

Colm Caffrey: colm.caffrey@dcu.ie
Marian Flanagan: marian.flanagan23@mail.dcu.ie
Stephen Armstrong: sarmstrong@computing.dcu.ie