Experimental research on Machine Translation in Translation Studies curriculum

Module “Principles and Applications of Machine Translation”

Bogdan Babych
b.babych@leeds.ac.uk
Overview

- Collaborative translation workflow and modern Machine Translation (MT)
- Relevance on teaching MT in translation studies curriculum
- Module structure: professional skills and ‘intellectual adventure’
- Research & impact in students’ projects
Problem: Finding a proper place of MT in translation workflow

- From early days to the ‘big data’ era: industry & teaching
- Answer for evolving MT technology & industrial demand
  - MT useful in a range of scenarios (controlled language, closely-related languages)
  - Human translation outperforms MT, MT always needs to be checked
  - MT usefulness increases with time: cannot afford to stay skeptical
- Two converging industrial and technological trends
  - Demand for high-volume collaborative translation & strict deadlines
  - Increasing MT output quality with new tech developments + datasets
- New generation: respond to technological MT paradigm shift
• Large-volume multilingual projects
  • Localization, product documentation & global release deadlines, style guidelines, terminology consistency within & across projects
  • Require ‘weapon-grade’ technology support: content, TMs + (increasingly) computational linguistic technologies

• Steady evolution of MT (what skeptics might have missed)
  • Improvements in usability, coverage, domains & translation directions
  • New ways of integrating MT into professional translation workflow: MT API in TMs + customization and terminology support
  • Collaborative translation platforms with interactive & off-line MT
  • Monolingual and parallel corpus support, post-editing feedback loop, project-specific engine tuning
Technological developments in MT: translation quality improvements

• For translators and students: awareness of tools & resources for developing modern high-quality wide-coverage MT
  • Architectures, new ‘training’ and runtime decoding methods, models
    • word-based → phrase-based → hierarchical, syntax-based, neural MT, hybrid MT (RBMT → SMT, e.g., Systran; SMT → RBMT)
  • Systematic linguistic improvements with other NLP technologies: Named Entity recognition; MWEs, WordNet, WSD, LVCs, agreement…
  • System customization and tuning on task-specific MT evaluation scores, terminology extraction & management, domain adaptation
  • Leveraging shared and open-source TMs: domain-specific training
  • Fast MT development for under-resourced languages
• Result: measurable improvements: time, cost, consistency and fitness for purpose
Module “Principles & Application of MT”
for MA TS students

• 1sem (spring), 15-credit, elective

• “Research-lead; informing future research; impact-driven”

• Students Should be able to:
  • Understand capabilities & limitations of MT systems for different translation scenarios; be able to use the potential of MT technology for optimizing the workflow in their translation projects
  • Work in the rapidly changing technological MT landscape in future,
    • follow new developments and literature MT
    • assess their impact on translation workflow
  • understand and exploit synergies between MT and other CAT tools, e.g., take advantage of MT customization and domain adaptation capacities for own projects and tasks

Students’ research component in MT module

- Rapidly-changing industrial landscape with no ready answers to future project needs in professional context
  - E.g., system evaluation & comparison, MT integration into the workflow
  - In future will need to set up experiments to systematically find answers relevant for a specific purpose, translation project, or organization

- Research component in taught classes and assessment
  - Teaching methods, techniques and tools for experimental research
  - Assessment: an experimental case study on MT evaluation, improvement, workflow design etc. within a theoretical framework
  - Session on principles of experiment design
  - One-to-one tutorial on students’ individual projects
MT module structure

- Theoretical linguistic and computational issues in MT
  - MT architectures, system combination, productivity of MT in workflow
  - Unresolved issues and future developments
- Usage scenarios for imperfect MT output, post-editing, controlled language, sublanguage
- Customizing Rule-based, Statistical and hybrid MT
  - Dictionary update, terminology extraction and management: Systran
  - Domain adaptation & translation corpus training
- Human and automated methods of evaluating MT quality
  - Error analysis schemes, system improvement, concordance evaluation
  - Calibration of automated scores, automated error analysis
MT module: professional relevance and intellectual adventure

- Research skills for modern professional context …
- … + Academically rewarding intellectual adventure:
  - MT development vs. evaluation: ‘there will be no good engineering where there is no good science’ (M.Kay)
    - Evaluation as a way to understand what we do not know about translation, feeding back into MT development cycle.

- Perspectives for students’ research path
  - Final-year dissertation
  - MA(Res) and PhD projects

- Impact component: actual or potential value for NGOs, volunteer translation groups, translation companies.
Students’ projects

- Experimental study + 2000 words report
  - For ‘distinction’ – innovative, make a potential contribution to state-of-the-art MT evaluation research (methodology, approach, but not necessarily the material or results)
    - Describe at a level suitable for publication / presentation
    - Follow the structure of research papers in the area of MT
    - Include references to recent publications in journals and conferences proceedings on the relevant topics
  - Systematic analysis of the material, e.g., corpus-based + linguistic interpretation or insights & discussion with linguistic examples
  - Structured feedback (external examiner liked it!)
Module feedback

Feedback for
MODL5003, Principles and Applications of Machine Translation

Name: 200543679
Marked by: BB
Grade score (Dictionary * 20% + Case Study * 80% = Total) 72 + 74 = 74

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>y good</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>completeness of the task</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coverage of error types</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comments</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Study: Content</th>
<th>y good</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>fit with essay topic</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quality of research</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coverage of essential points</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>substantiating evidence</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pitch at MA level</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Study: Argumentation</th>
<th>y good</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>clarity of introduction</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>links between ideas</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>justification of conclusion</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Study: Examples</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>pertinence</td>
<td>y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clarity of gloss in English</td>
<td>y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Study: Citations</th>
<th>y good</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>appropriate references</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>presentation of bibliography</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Study: Presentation</th>
<th>y good</th>
<th>good</th>
<th>weak</th>
<th>y weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>sections and paragraphs</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spelling</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grammar and fluency</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Further comments:

This is an excellent case study with a clear identification of an important and innovative research problem in the area of MT evaluation, and well-designed methodology to address it, a very good execution of the experiment and a clearly described interpretation of the results. The chosen problem is also important for MT evaluation in several respects, as was clearly motivated in the introduction, with appropriate reference to literature sources. The case study shows your critical ability and makes an interesting contribution to knowledge: currently the results are on the small scale pilot experiment, but the methodology, properly refined, can be the basis of a serious published study of intelligibility of MT in future.

A very strong methodological point is that you started with the two hypotheses based on reasonable psychological assumptions, and identified the way to test them with evaluation data (that clearly did not support your initial hypotheses). You also clearly indicated another finding (done alongside the main investigation) that there is a dependency between error types and intelligibility – i.e., idioms are more important than e.g., word order for understanding texts. This is an important positive result, and can become a matter of an independent investigation in future.

The point that can be improved:
- In future you may wish to look into more advanced mathematical techniques to characterise the relations between responses in gap filling exercise for individual participants and their intelligibility scores: current visualisation gives a clear qualitative understanding of the relation, but for a more systematic way to address this problem you may look into finding quantitative measures of agreement / disagreement.

The level of investigation and understanding of the problem is very high.
How technical?

- Students with humanities background
- Module introduces experiment design and data analysis techniques, but no technical pre-requisites
- Exploits synergies of computational linguistic methods with knowledge of the professional translators’ workflow.
- New generation of technologically aware translators?
  - Translators who collaborate with CAT engineering teams
  - Who define and create the next generations of MT and CAT tools, which would be more capable and useful for professional translators
Students projects and results

- Impact of segmentation methods for Chinese text on accuracy of automated BLEU evaluation
  - Res: character-based segmentation better correlates with human scores
- Designing a controlled language for drafting Spanish language loan overviews at Kiva for automated translation into English
  - Outcome: Rules for non-specialists and effect on MT quality (adequacy)
  - Tested with non-for-profit organization
- Comparative error analysis on the output of rule-based and statistical machine translation systems before and after the implication of a Spanish Control Language
- An investigation into the potential for human evaluation of Machine Translation to be objective
Students projects and results

- An exploration of the predictability of human evaluation scores of machine translation system output from automated evaluation scores across varied text types
- Evaluating Error Analysis Results From MT Systems across Genres
- Relation between pre-editing and quality of translation
- Declarative evaluation of MT in legal domain
- Analysis of the feasibility of using Systran 7 and Google Translate for translation of a literary text, with consequent evaluation of application of MT in this field
- Creation of new grammatical controlled language rules for French through error analysis and self-evaluation of machine translation quality
- The Use of Controlled Language for Improving the Translation of Scientific Texts by Systran
Students projects and results

• Human Evaluation of Machine Translated Patient Information Leaflets (PILs)
• Using the BLEU evaluation metric to determine the suitability of rule-based compared with statistical MT systems in the translation of medical texts
• Comparative analysis of task-based vs. Proximity-based evaluation methodologies
• Setting up an MT evaluation framework for a translation company, an MT development team, etc.
• Automated MT evaluation methods: current limitations and future prospects.
• Designing human evaluation experiment for FEMTI framework
• Use of error analysis and test suites for general-purpose MT systems