“The More, the Better” in Human and Machine Translation
what do we have in common?

Ekaterina Lapshinova-Koltunski

Germersheim
January 29th, 2015
Overview

1. Aims and Motivation
2. Related Work
3. Methods and Data
   - Data
   - Feature Selection
4. Analysis and Interpretations

January 29th, 2015
Aims and Motivation
Aims and Motivation

Translation Phenomenon

January 29th, 2015 “The More, the Better”
Aims and Motivation

Translation Phenomenon
Aims and Motivation

Translation Phenomenon: Dimensions

January 29th, 2015

“The More, the Better”
Aims and Motivation

Translation Phenomenon: Dimensions

- “translation varieties” distinguished not only according to
  1. languages (both source and target)
  2. registers (text types they belong to)

- but also:
  1. translation methods (human vs. machine) used in translation process,
  2. experience (knowledge and data) involved

Assumption: experience relate HT and MT and provide us with new perspectives on their research.
Aims and Motivation

Assumption: Similarities

<table>
<thead>
<tr>
<th>HUMAN TRANSLATION</th>
<th>MACHINE TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>amount of human translator experience</td>
<td>amount of data used in SMT</td>
</tr>
<tr>
<td>similar outcomes in both translation varieties</td>
<td></td>
</tr>
<tr>
<td>professional vs. novice translators</td>
<td>systems based on small vs. large training data</td>
</tr>
<tr>
<td>learn in a similar way, collecting experience, data or knowledge</td>
<td></td>
</tr>
<tr>
<td><strong>However:</strong> they both differ in the type of the experience acquired, which is reflected in their linguistic features</td>
<td></td>
</tr>
</tbody>
</table>
Assumption: Similarities
Related Work
studies addressing both human and machine translations: [White, 1994], [Papineni et al., 2002], [Babych et al., 2004], [Popović and Burchardt, 2011], [Popovic and Ney, 2011]

all focus solely on translation error analysis, using human translation as a reference

studies operating with linguistically-motivated categories: [Popović and Burchardt, 2011], [Popovic and Ney, 2011] or [Fishel et al., 2012]

However: none of them provides a comprehensive analysis of specific linguistically motivated features of different translation methods
Related Work

Human and Machine Translation

- works on differentiation between human and machine translation:
  - (1) [Volansky et al., 2011] and (2) [El-Haj et al., 2014]:
    - (1) analysis of human and machine translations, and comparable non-translated texts
    - a range of features based on the theory of translationese, see [Gellerstam, 1986]
    - claim that the features specific for human translations can be used to identify MT
    - coinciding and diversifying features
    - (2) compare translation style and consistency in human and machine translations of Camus’ novel “The Stranger” (French-English and French-Arabic)
    - measure: readability as a proxy for style
    - evaluative and not descriptive character
Experience: SMT

- the relation between large vs. small data plays an important role
- Gains from using large corpora for translation/language models:
  - [Estrella et al., 2007]
  - [Bertoldi and Federico, 2009]
  - [Koehn and Haddow, 2012]
Experience: HT

- relation between experienced vs. inexperienced translators:
  [Jääskeläinen, 1997], [Jakobsen, 2002],
  [Martinez Melis and Hurtado Albir, 2001],
  [Englund-Dimitrova, 2005], [Muñoz Martín and Conde, 2007],
  [Carl and Buch-Kromann, 2010]
  - [Jääskeläinen, 1997]: translational behaviour of professionals and non-professionals who perform translation from English into Finnish
  - [Carl and Buch-Kromann, 2010]: translation processes for student and professional translators, relating properties of the translators’ eye movements and keystrokes to the quality of the produced translations
    ⇒ process rather than the product
    (no description of linguistic features involved)
  - [Englund-Dimitrova, 2005]: a combined process and product analysis
    ⇒ only cohesive explicitness
Experience Combined

- some parallelism in the observations for both translation varieties
  - [Göpferich & Jääskeläinen, 2009]:
    - with increasing translation competence, translators focus on larger translation units
  - [Koehn, 2010]:
    - in (phrase-based) MT, large training data sets make it possible to learn longer phrases

- experienced translators take into account more aspects, e.g. text function, context, register

- systems trained with larger data set can also capture more contextual information
Methods and Data
Questions and Methodology

- compare relations between novice translations and statistical machine translations with small data
- between professional translations and statistical machine translations produced with much data
- trace the patterns specific for both human and machine translation:
  - define similarities and differences between the varieties
## Corpus-Based Approach

<table>
<thead>
<tr>
<th>REQUIREMENT</th>
<th>CHOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexico-grammatical features</td>
<td>corpus-based approach</td>
</tr>
<tr>
<td>corpus</td>
<td>contain human and machine translation</td>
</tr>
<tr>
<td>setting 1</td>
<td>two variants of HT</td>
</tr>
<tr>
<td>setting 2</td>
<td>two SMT systems</td>
</tr>
<tr>
<td>features</td>
<td>frequencies of lexico-grammatical patterns</td>
</tr>
<tr>
<td>feature distributions</td>
<td>across translation varieties</td>
</tr>
</tbody>
</table>
Corpus

VARTRA cf. Lapshinova (2013)

contains:

- variants of translation from English into German
  = translation varieties produced by:
  1. human professional translators (PT1)
  2. human inexperienced translators (PT2)
  3. a rule-based MT system (RBMT)
  4. 2 statistical MT systems (SMT1 and SMT2)

TOTAL number of tokens in translations ca. 600,000
Corpus

- PT1 – CroCo, [Hansen-Schirra et al., 2012]
- PT2 – trained translators (over BA) with no/little experience
- RBMT – SYSTRAN
- SMT1 – Google Translate (big undefined data)
- SMT2 – Moses system (small known data)

Each translation covers 7 registers:

- political essays – ESSAY
- fictional texts – FICTION
- instruction manuals – INSTR
- popular-scientific articles – POPSCI
- letters of share-holders – SHARE
- prepared political speeches – SPEECH
- touristic leaflets – TOU
Features: Patterns
Feature Requirements

- reflect linguistic characteristics of all texts under analysis
Feature Requirements

- reflect **linguistic characteristics** of all texts under analysis

- **content-independent** (do not contain terminology or keywords)
Feature Requirements

- reflect **linguistic characteristics** of all texts under analysis

- content-independent *(do not contain terminology or keywords)*

- easy to interpret
Methods and Data

Feature Selection

Nature of Features

Data-driven
Nature of Features

Data-driven

Theory-driven

January 29th, 2015
Our Features

1. **data-driven:**
   - joint work with Marcos Zampieri (DFKI, Saarbrücken)

2. **theory-driven:**
   - in [Lapshinova, forthcoming],
   - in joint work with Mihaela Vela (UdS)
## Theory-Driven Features

<table>
<thead>
<tr>
<th>patterns</th>
<th>register</th>
<th>translationese</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 content vs. grammatical words</td>
<td>mode</td>
<td>simplification</td>
</tr>
<tr>
<td>2 nominal vs. verbal word classes and phrases</td>
<td>field</td>
<td>normalisation / shining through</td>
</tr>
<tr>
<td>3 <em>ung</em>-nominalisation</td>
<td>field</td>
<td>normalisation / shining through</td>
</tr>
<tr>
<td>4 nominal vs. pronominal and demonstrative vs. personal</td>
<td>mode</td>
<td>explicitation, normalisation / shining through</td>
</tr>
<tr>
<td>5 abstract or general nouns vs. all other nouns</td>
<td>fields</td>
<td>explicitation</td>
</tr>
<tr>
<td>6 logico-semantic relations: additive, adversative, causal, temporal, modal</td>
<td>mode</td>
<td>explicitation</td>
</tr>
<tr>
<td>7 modal meanings: obligation, permission, volition</td>
<td>tenor</td>
<td>normalisation / shining through</td>
</tr>
<tr>
<td>8 evaluative patterns</td>
<td>tenor</td>
<td>normalisation / shining through</td>
</tr>
</tbody>
</table>
Hierarchical Cluster Analysis (HCA)

- discover differences and similarities between subcorpora
- discover ‘interesting structures’
- group according to different dimensions, i.e. variety and register

cf. Baayen (2008); Everitt et al. (2001)

- set of dissimilarities for the $n$ objects is used
- each subcorpus is assigned to its own cluster
- iteratively, two most similar clusters
- statistics behind hierarchical cluster analysis quantifies how far apart (or similar) two (sets of) subcorpora are
- on the top node, all clusters are joined together

⇒ subcorpora similar to each other have a common node on the tree
Hierarchical Cluster Analysis (HCA)
Analysis and Interpretations

Statistical test: Correspondance Analysis

- **Input:** frequencies of **features**
- **Output:** a two dimensional graph with:
  - **arrows** for the observed feature frequencies
  - **points** for translation varieties
- **Interpretation:**
  - the length of the **arrows** indicates how pronounced a particular feature is
  - the position of the **points** in relation to the **arrows** indicates the relative importance of a feature for a register.
  - the **arrows** pointing in the direction of an axis indicate a high contribution to the respective dimension

cf. [Baayen, 2008], [Glynn, 2012]
Correspondence Analysis

January 29th, 2015

“The More, the Better”
Summary and Discussion

- parallelism in HT and MT in terms of experience
- influence on quality?
- realise that both differ in the type of experience (or data) acquired
  ⇒ need to be modelled differently
- more data
- more features
- other methods?
Thank you!

Questions? Comments? Suggestions?

e.lapshinova@mx.uni-saarland.de
Modelling legitimate translation variation for automatic evaluation of mt quality.

Domain adaptation for statistical machine translation with monolingual resources.

Correlating translation product and translation process data of professional and student translators.
In Annual Conference of the European Association for Machine Translation, Saint-Raphaël, France.

Language independent evaluation of translation style and consistency: Comparing human and machine translations of camus’ novel “the stranger”.

Expertise and Explicitation in the Translation Process.
John Benjamins, Amsterdam/Philadelphia.

How Much Data is Needed for Reliable MT Evaluation? Using Bootstrapping to Study Human and Automatic Metrics.
ID: unige:3461.

Terrorcat: a translation error categorization-based mt quality metric.
In 7th Workshop on Statistical Machine Translation.

Translationese in Swedish novels translated from English.


From human to automatic error classification for machine translation output.
In 15th International Conference of the European Association for Machine Translation (EAMT-2011), Leuven, Belgium. European Association for Machine Translation.

Towards automatic error analysis of machine translation output.

More human or more translated? original texts vs. human and machine translations.

The ARPA MT evaluation methodologies: Evolution, lessons, and further approaches.