Document Level
Statistical Machine Translation

Eva Martínez García

Research Plan

Advisors:
Lluís Màrquez Villodre
Cristina Espeña Bonet

Barcelona; February 12, 2014
Outline

1. Introduction
   - MT Nowadays
   - Statistical Machine Translation
   - Automatic Evaluation

2. Motivation and Research Goals

3. Related Work

4. Current Work
   - Analysis of the Phenomena
   - Experiments: Postprocess
   - Experiments: Decoding

5. Temporal Planning
Machine Translation is the use of a computer to translate a message from one natural language to another.

- **Frequent uses**
  Translation web services: *Google Translate, Bing, Reverso, etc.*

- **Important research conferences**
  ACL, MT Summit, DiscoMT, etc.

- **Research projects**
  USA (GALE, BOLT) Europe (FP7 projects)
  GPLN recent projects: Spanish (OpenMT, OpenMT2, TACARDI), and European (FAUST, MOLTO)
MT History

1940: MT appears

1950: 1st systems
Promising results

1960: Lack of productivity
1966
ALPAC report

1970: Powerful computers
MT revives

1980: SMT
Statistical approaches

1990: NOW

2000: Nowadays
Online interfaces
Open domain
Many language pairs

Current Work
Analysis of the Phenomena
Experiments:
Postprocess
Experiments:
Decoding

Temporal Planning

Related Work
MT at Document Level
Automatic Evaluation

Motivation and Research Goals
State-of-the-art
Limits
Research Goals
Summary
Framework

Introduction
MT Nowadays
SMT
Automatic Evaluation

Document-Level SMT
E. Martínez
MT Systems Classification

Machine Translation systems

- Human Translation with Machine Support
- Machine Translation with Human Support
- Fully Automated Translation

Machine Translation systems

- Human Translation with Machine Support
- Machine Translation with Human Support
- Fully Automated Translation
MT Systems Classification

- Machine Translation systems
  - Human Translation with Machine Support
  - Machine Translation with Human Support
    - Fully Automated Translation
      - Empirical systems
      - Rule-based systems
MT Systems Classification

- Machine Translation systems
  - Human Translation with Machine Support
  - Machine Translation with Human Support
  - Fully Automated Translation
    - Empirical systems
    - Rule-based systems
      - Statistical Machine Translation
      - Example-based Translation

MT Systems Classification

- Machine Translation systems
  - Human Translation with Machine Support
  - Machine Translation with Human Support
  - Fully Automated Translation
    - Empirical systems
    - Rule-based systems
      - Statistical Machine Translation
      - Example-based Translation
Mathematically

\[
P(e|f) = \frac{P(e)P(f|e)}{P(f)}
\]

\[
T(f) = \arg\max_e P(e|f) = \arg\max_e P(e)P(f|e)
\]

decoder \( \arg\max_e \) 

language model \( P(e) \) 

translation model \( P(f|e) \)
SMT Systems Structure

Phrase Extraction → Preprocessing → Parallel corpus

Word/Phrase Alignment

Translation Model

Language Model

Preprocessing

Monolingual corpus
SMT Systems Structure

- **Source Language Sentence**
- **Preprocessing**
- **Decoding**
- **Postprocessing**
- **Target Language Sentence**
- **Phrase Extraction**
- **Preprocessing**
- **Parallel corpus**
- **Word/Phrase Alignment**
- **Translation Model**
- **Language Model**
- **Preprocessing**
- **Monolingual corpus**

Example of SMT system

*Moses (Koehn, P., et al. 2007)*
Automatic Evaluation

Compare automatic translation with reference manual texts to get the semantic similarity.

- Necessity of evaluating MT systems and having a guidance for development and tuning
- Usual measures are based on lexical information: BLEU, NIST, ROUGE (n-gram matching), TER, WER, (error rate), etc.
- Measures that include more linguistic information: lexical, syntactic, semantics, discourse.
Outline

1. Introduction

2. Motivation and Research Goals
   - State-of-the-art Limits
   - Research Goals Summary
   - Framework

3. Related Work

4. Current Work

5. Temporal Planning
MT Limitations

- **Limitation:**
  Translation is produced **sentence by sentence** (no discourse-level context is taken into account)

- **Consequence:**
  **Lack of cohesion and coherence** when translating long texts
MT Limitations

**Goal:** Improving SMT at document level

- **Limitation:** Translation is produced *sentence by sentence* (no discourse-level context is taken into account)

- **Consequence:** Lack of cohesion and coherence when translating long texts
Contextual coherence

The window is open. It is blue. La fenêtre est ouverte. **Elle est bleue.**  CORRECT
The window is open. It is black. La fenêtre est ouverte. **Il est noir.**  WRONG
The oven is open. It is new. Le four est ouverte. **Elle est neuve.**  WRONG
The door is open. It is new. La porte est ouverte. **Elle est neuve.**  CORRECT

Translation errors due to lack of co-reference resolution (created with Google Translate). (Nagard-Koehn, 2010)

Consistency translating ambiguous words

- house → **cámara, casa**
- floor → **piso, planta, suelo**
- desk → **ventanilla, despacho, mostrador**
MT Limitations
Examples

**Subtasks:** Treatment of different linguistic phenomena that confer consistency to the translation.

- Contextual coherence

The window is open. *It is blue.* La fenêtre est ouverte. **Elle est bleue.** CORRECT

The window is open. *It is black.* La fenêtre est ouverte. **Il est noir.** WRONG

The oven is open. *It is new.* Le four est ouverte. **Elle est neuve.** WRONG

The door is open. *It is new.* La porte est ouverte. **Elle est neuve.** CORRECT

Translation errors due to lack of co-reference resolution (created with Google Translate). (Nagard-Koehn, 2010)

- Consistency translating ambiguous words

  house → cámara, casa
  floor → piso, planta, suelo
  desk → ventanilla, despacho, mostrador
Limitations of Automatic MT Evaluation Measures

- Translation quality evaluation at **sentence level**
  Measures do not take into account information at document level

- Improvements of MT systems in lexical **cohesion and coherence or discourse structure** are not captured
Limitations of Automatic MT Evaluation Measures

**Goal:** Research in Automatic Evaluation at document level
- Translation quality evaluation at **sentence level** Measures do not take into account information at document level
- Improvements of MT systems in lexical **cohesion and coherence or discourse structure** are not captured
Research Goals
Summary

- **Improving translation quality**
  Use document level information to improve translation coherence and cohesion

- **Evaluation at document level**
  *Enhance* current automatic evaluation metrics and *design* new ones to measure quality at document level
**TACARDI**  
Context-aware Machine Translation Augmented using Dynamic Resources from the Internet

*(TIN2012-38523-C02-00)*

Improve current SMT system attending to:

- *the lack of linguistic resources*
  gathering resources from the Internet
- *contextual incoherences*
  Document Level MT strategies and automatic evaluation metrics
Outline

1. Introduction
2. Motivation and Research Goals
3. Related Work
   - MT at Document Level
   - Automatic Evaluation
4. Current Work
5. Temporal Planning
Main Ideas

- Sentences are not independent: use **discourse information**
- One sense per discourse: **topic** helps understanding some words/phrases/sentences
MT at Document Level (II)

### Problems

- Parallel corpora without annotation of **document boundaries**
- Difficult to capture document-level information via features
- **Few** evaluation measures exist at document level
- No corpora with **human assessments** exist at document level to train or compute correlations
Related Work (I)

- **Topic-based** language models (Tam et al., 2007)
- “One translation per discourse” hypothesis (Carpuat, 2009)
- **Cache-based** approach for domain adaptation (Tiedemann, 2010)
- Using **Coreference resolution** to help pronoun translation (Nagard-Koehn, 2010)
Related Work (II)

- Post-editing process treating **ambiguous words** (Xiao et al., 2011)
- Modelling **pronominal anaphora** (Hardmeier et al., 2010)
- **Stochastic local search** decoding method (Hardmeier et al., 2012)
- **Docent decoder**: a document level decoder (Hardmeier et al., 2013)
Related Work (III) - Document level measures

- Precision / accuracy of discourse level phenomena.
- Document level automatic MT evaluation based on discourse representations (Giménez et al., 2008)
- Extension of standard machine translation evaluation metrics with lexical cohesion (Wong et al., 2012)
Outline

1. Introduction
2. Motivation and Research Goals
3. Related Work
4. Current Work
   - Analysis of the Phenomena
   - Experiments: Postprocess
   - Experiments: Decoding
5. Temporal Planning
Analysis of the Phenomena (I)

- **Ambiguous words**
  Same word with different meanings translated in several incompatible forms through a document

Batlló *House*, located in the Eixample district on a busy shopping street Passeig de Gracia No. 43, was rebuilt from the original *house* in 1877 by the architect Emilio Sal Cortes.

→ La *cámara* Batlló, situada en el distrito Eixample en la calle Passeig de Gracia No. 43, fue reconstruida desde la *casa* original en 1877 por el arquitecto Emili Sal Cortes.
Analysis of the Phenomena (II)

**Coreference chains**

*The window is open. It is blue.*  *La fenetre est ouverte. Elle est bleue.*  CORRECT

*The window is open. It is black.*  *La fenetre est ouverte. Il est noir.*  WRONG

*The oven is open. It is new.*  *Le four est ouverte. Elle est neuve.*  WRONG

*The door is open. It is new.*  *La porte est ouverte. Elle est neuve.*  CORRECT

Translation errors due to lack of co-reference resolution (created with Google Translate). (Nagard-Koehn,2010)
Analysis of the Phenomena (III)

- **Gender/number agreement**
  Disagreements in gender or number among nouns/pronouns and their determiners and adjectives
  
  the surveillance of the estate of the gasduct is provided by an undersea robot
  
  → la vigilancia del estado de la gasoducto es **proporcionados** por un robot en el mar.
Analysis of the Phenomena (III)

- **Gender/number agreement**
  Disagreements in gender or number among nouns/pronouns and their determiners and adjectives

  the surveillance of the estate of the gasduct is provided by an undersea robot

  → la vigilancia del estado de *la gasoducto* es proporcionados por un robot en el mar.
Analysis of the Phenomena (IV)

- **Discursive structure**
  Source document and translation must follow the same structure: connectors, if/then clauses, etc.

  It’s a good idea. *However*, I don’t think we have the money to fund it.
  → Es una buena idea. *Sin embargo*, no tenemos dinero para financiarla
  → Es una buena idea. *Además*, no tenemos dinero para financiarla
Experiments: A proposal

1. Linguistic annotation of the input
2. Transfer annotations into the output via the SMT alignments
3. Identify mismatches between input and output at document level
4. Treat the mismatches
   - postprocess
   - generating information for a document based SMT decoder
Experiments: Resources

- **Europarl**
  Version 7: Spanish-English, 1,965,734 sentences, 51,575,748 Spanish words, 49,093,806 English words

- **Newscommentaries 2011**
  110 news documents with 680 words in average

- **Moses baseline**
  Trained with the Europarl v7 and tested over Newscommentaries2011

- **Automatic analysis tools**: Asiya toolkit, Freeling, RelaxCor
Experiments: Postprocess (I)
Ambiguous Words
Experiments: Postprocess (I)

Ambiguous Words
Experiments: Postprocess (I)
Ambiguous Words - Results

<table>
<thead>
<tr>
<th>System</th>
<th>TER</th>
<th>BLEU</th>
<th>NIST</th>
<th>METEOR-ex</th>
<th>ROUGE-L</th>
<th>ULC</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>0.5545</td>
<td>0.2673</td>
<td>7.3360</td>
<td>0.2733</td>
<td>0.5151</td>
<td>0.7638</td>
</tr>
<tr>
<td>restrictive</td>
<td>0.5539</td>
<td>0.2676</td>
<td>7.3435</td>
<td>0.2735</td>
<td>0.5157</td>
<td>0.7650</td>
</tr>
<tr>
<td>probabilistic</td>
<td>0.5541</td>
<td>0.2673</td>
<td>7.3395</td>
<td>0.2732</td>
<td>0.5155</td>
<td>0.7644</td>
</tr>
</tbody>
</table>

Slight improvements:

- Few number of changes (500/1000 aprox. changes in 75000 words)
- Measures do not capture cohesion improvements
Experiments: Postprocess (II)
Agreement

![Diagram showing the process of postprocess agreement]

Source → SMT → Target

Det Noun  Noun Adj
Det Noun Adj

Freeling → POS Parse Trees → Post-Process

Fixed Translation
Experiments: Postprocess (II)
Agreement - Results

<table>
<thead>
<tr>
<th>System</th>
<th>TER</th>
<th>BLEU</th>
<th>NIST</th>
<th>METEOR-ex</th>
<th>ROUGE-L</th>
<th>ULC</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>0.5545</td>
<td>0.2673</td>
<td>7.3360</td>
<td>0.2733</td>
<td>0.5151</td>
<td>0.7633</td>
</tr>
<tr>
<td>agreement_nn+dets</td>
<td>0.5540</td>
<td>0.2676</td>
<td>7.3404</td>
<td>0.2737</td>
<td>0.5161</td>
<td>0.7649</td>
</tr>
<tr>
<td>agreement_nn+dets+adj</td>
<td>0.5538</td>
<td>0.2669</td>
<td>7.3437</td>
<td>0.2734</td>
<td>0.5156</td>
<td>0.7642</td>
</tr>
</tbody>
</table>

Slight improvements:
- Few number of changes
- Measures do not capture cohesion improvements
Experiments: Decoding Docent (I)
Experiments: Decoding Docent (II)

Design of

- **Feature Function**
  attending to the studied phenomena

  \[
  f_{amb}(x) = \frac{\#\text{ambiguos words}}{\#\text{words}}
  \]

- **Change operation**
  future work: detect inconsistencies at translation time
1. Introduction

2. Motivation and Research Goals

3. Related Work

4. Current Work

5. Temporal Planning
Introduction
MT Nowadays
SMT
Automatic Evaluation

Motivation and Research Goals
State-of-the-art
Limits
Research Goals
Summary
Framework

Related Work
MT at Document Level
Automatic Evaluation

Current Work
Analysis of the Phenomena
Experiments: Postprocess
Experiments: Decoding

Temporal Planning

First Year
(Feb. 2013 - Sept. 2013)

- **Background**
  Study bibliography about MT, SMT and MT at document level

- **Study of the data**
  Automatic and manual study of the data
  Building a baseline SMT system

- **Study and treatment of the discourse phenomena**
  Build a prototype able to run some preprocess, fix ambiguous words translation and disagreements with a postprocess
Second Year
(Sept. 2013 - Sept. 2014)

- **Improve the first models**
  Heuristics to detect and suggest new possible translations
  Fix/prevent gender/number disagreements also in verbs, pronouns using coreferences, etc
  Study discourse structure
Second Year (II)  
(Sept. 2013 - Sept. 2014)

- **Document level decoder**
  Integrate our models inside Docent decoder

- **Evaluation Metrics**
  Enhance current metrics with discourse level information or use new metrics that are developed currently by other researchers
Second Year (II)  
(Sept. 2013 - Sept. 2014)

- **Document level decoder**
  Integrate our models inside Docent decoder

- **Evaluation Metrics**
  Enhance current metrics with discourse level information or use new metrics that are developed currently by other researchers

### Short stay abroad

3 months at the Department of Linguistics and Philology of Uppsala University (Sweden)  
*Founded by Spanish Ministry of Economics and Competitivity grant (EEBB-I-14-08614)*
Third Year
(Sept. 2014 - Sept. 2015)

- Improve our models
  Domain adaptation by topic

- Participation in international evaluation campaigns
  Validate our system(s) in international frameworks

- Documents
  Thesis summary for external reviewers
  Final PhD thesis manuscript
Thank you for your attention!
Support Slides
Gong, Z., Zhang, M., Zhou, G.

Hardmeier, C., Federico, M.
Modelling pronominal anaphora in statistical machine translation.
Hardmeier, C., Nivre, J., Tiedemann, J.  
Document-Wide Decoding for Phrase-Based Statistical Machine Translation.  

Hardmeier, C. and Stymne, S. and Tiedemann, J. and Nivre, J.  
Docent: A Document-Level Decoder for Phrase-Based Statistical Machine Translation.  
Papineni, K. and Roukos, S. and Ward, T. and Zhu, W. J.
Bleu: a method for automatic evaluation of machine translation

Doddington, G.
Automatic evaluation of machine translation quality using n-gram co-occurrence statistics.
References

Snover, M. and Dorr, B.J. and Schwartz, R. and Micciulla, L. and Makhoul, J.
A Study of Translation Edit Rate with Targeted Human Annotation.

Lin, C.-Y and Och, F.J.
References

Denkowski, M. and Lavie, A.
METEOR-NEXT and the METEOR Paraphrase Tables: Improved Evaluation Support for Five Target Languages.

Giménez, J. and Màrquez, L. and Comelles, E. and Castellón, I. and Arranz, V.
Document-level automatic MT evaluation based on discourse representations

Giménez, J. and Màrquez, L.
Linguistic Measures for Automatic Machine Translation Evaluation
Machine Translation, 24 (3-4) (2010)
Le Nagard, R., Koehn, P.
Aiding pronouns translation with co-reference resolution.

Xiao, T., Zhu, J., Yao, S., Zhang, H.
Document-level consistency verification in machine translation.
Support Slides
Automatic Evaluation (I)

Lexical metrics

- Edit distances
  PER, TER, WER

- n-gram matching
  BLEU, NIST, ROUGE

- F-measure
  GTM, METEOR, OI

- Uniform Linear Combination (ULC)
Automatic Evaluation (II)

Current metrics designed at sentence level
- BLEU (Papineni et al., 2001, 2002),
- NIST (Doddington, 2002),
- TER (Snover et al., 2006, 2009),
- ROUGE (Lin and Och, 2004),
- METEOR (Denkowski and Lavie, 2012a)
Experiments: Postprocessing
Ambiguous Words - Results per new

<table>
<thead>
<tr>
<th>System</th>
<th>TER</th>
<th>BLEU</th>
<th>NIST</th>
<th>METEOR-ex</th>
<th>ROUGE-L</th>
<th>ULC</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline_new1</td>
<td>0.6887</td>
<td>0.1079</td>
<td>3.8257</td>
<td>0.1966</td>
<td>0.4275</td>
<td>0.7647</td>
</tr>
<tr>
<td>restrictive_new1</td>
<td>0.6887</td>
<td>0.1080</td>
<td>3.8318</td>
<td>0.1964</td>
<td>0.4252</td>
<td>0.7647</td>
</tr>
<tr>
<td>probabilistic_new1</td>
<td>0.6909</td>
<td>0.1098</td>
<td>3.8173</td>
<td>0.1953</td>
<td>0.4244</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new2</td>
<td>0.6369</td>
<td>0.2073</td>
<td>4.2684</td>
<td>0.2457</td>
<td>0.4721</td>
<td>0.7647</td>
</tr>
<tr>
<td>restrictive_new2</td>
<td>0.6369</td>
<td>0.2073</td>
<td>4.2684</td>
<td>0.2457</td>
<td>0.4721</td>
<td>0.6471</td>
</tr>
<tr>
<td>probabilistic_new2</td>
<td>0.6369</td>
<td>0.2073</td>
<td>4.2684</td>
<td>0.2457</td>
<td>0.4721</td>
<td>0.7059</td>
</tr>
<tr>
<td>baseline_new3</td>
<td>0.6679</td>
<td>0.1415</td>
<td>4.1891</td>
<td>0.2153</td>
<td>0.4234</td>
<td>0.7647</td>
</tr>
<tr>
<td>restrictive_new3</td>
<td>0.6672</td>
<td>0.1416</td>
<td>4.1929</td>
<td>0.2154</td>
<td>0.4239</td>
<td>0.7647</td>
</tr>
<tr>
<td>probabilistic_new3</td>
<td>0.6597</td>
<td>0.1423</td>
<td>4.2232</td>
<td>0.2169</td>
<td>0.4258</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new4</td>
<td>0.6755</td>
<td>0.1869</td>
<td>4.0303</td>
<td>0.2181</td>
<td>0.4365</td>
<td>0.7647</td>
</tr>
<tr>
<td>restrictive_new4</td>
<td>0.6740</td>
<td>0.1893</td>
<td>4.0638</td>
<td>0.2207</td>
<td>0.4411</td>
<td>0.6471</td>
</tr>
<tr>
<td>probabilistic_new4</td>
<td>0.6711</td>
<td>0.1894</td>
<td>4.0731</td>
<td>0.2214</td>
<td>0.4431</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new5</td>
<td>0.6915</td>
<td>0.1374</td>
<td>3.7287</td>
<td>0.2006</td>
<td>0.3999</td>
<td>0.7647</td>
</tr>
<tr>
<td>restrictive_new5</td>
<td>0.6915</td>
<td>0.1374</td>
<td>3.7287</td>
<td>0.2006</td>
<td>0.3999</td>
<td>0.7647</td>
</tr>
<tr>
<td>probabilistic_new5</td>
<td>0.6906</td>
<td>0.1375</td>
<td>3.7319</td>
<td>0.2004</td>
<td>0.4009</td>
<td>0.7059</td>
</tr>
</tbody>
</table>
### Experiments: Postprocessing Agreement - Results per new

<table>
<thead>
<tr>
<th>Model</th>
<th>Agreement</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Measure</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline_new1</td>
<td>0.6887</td>
<td>0.1079</td>
<td>3.8257</td>
<td>0.1966</td>
<td>0.4275</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets_new1</td>
<td>0.6821</td>
<td>0.1187</td>
<td>3.8778</td>
<td>0.1997</td>
<td>0.4302</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets+adj_new1</td>
<td>0.6799</td>
<td>0.1188</td>
<td>3.8900</td>
<td>0.2002</td>
<td>0.4310</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new2</td>
<td>0.6369</td>
<td>0.2073</td>
<td>4.2684</td>
<td>0.2457</td>
<td>0.4721</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets_new2</td>
<td>0.6350</td>
<td>0.2079</td>
<td>4.2859</td>
<td>0.2469</td>
<td>0.4759</td>
<td>0.6471</td>
</tr>
<tr>
<td>agreement_nn+dets+adj_new2</td>
<td>0.6350</td>
<td>0.2072</td>
<td>4.2597</td>
<td>0.2449</td>
<td>0.4752</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new3</td>
<td>0.6679</td>
<td>0.1415</td>
<td>4.1891</td>
<td>0.2153</td>
<td>0.4234</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets_new3</td>
<td>0.6687</td>
<td>0.1416</td>
<td>4.1833</td>
<td>0.2150</td>
<td>0.4225</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets+adj_new3</td>
<td>0.6679</td>
<td>0.1416</td>
<td>4.1791</td>
<td>0.2148</td>
<td>0.4236</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new4</td>
<td>0.6755</td>
<td>0.1869</td>
<td>4.0303</td>
<td>0.2181</td>
<td>0.4365</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets_new4</td>
<td>0.6740</td>
<td>0.1873</td>
<td>4.0474</td>
<td>0.2189</td>
<td>0.4377</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets+adj_new4</td>
<td>0.6740</td>
<td>0.1883</td>
<td>4.0510</td>
<td>0.2191</td>
<td>0.4402</td>
<td>0.7647</td>
</tr>
<tr>
<td>baseline_new5</td>
<td>0.6915</td>
<td>0.1374</td>
<td>3.7287</td>
<td>0.2006</td>
<td>0.3999</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets_new5</td>
<td>0.6879</td>
<td>0.1385</td>
<td>3.7687</td>
<td>0.2032</td>
<td>0.4029</td>
<td>0.7647</td>
</tr>
<tr>
<td>agreement_nn+dets+adj_new5</td>
<td>0.6861</td>
<td>0.1392</td>
<td>3.7956</td>
<td>0.2043</td>
<td>0.4014</td>
<td>0.7647</td>
</tr>
</tbody>
</table>
Docent details

- Built on top of the Moses
- Two different parts 
  *decoder* and *scorer*
- Change operations
  *change-phrase translation, resegment, swap-phrases*
- Included Feature Functions
  compatible with Moses features: *phrase table, n-gram language models (KenLM toolkit), unlexicalised distortion cost model, word penalty cost*
  document-level model: *length parity model, semantic language model, readability models*