Empirical Machine Translation and its Evaluation
EAMT Best Thesis Award 2008

Jesús Giménez

(Advisor, Lluís Màrquez)

Universitat Politècnica de Catalunya

May 28, 2010
Empirical Machine Translation

MT System Developer

Error Detection

Error Analysis

Refinement

Implementation

Test

OK? NO

Keep Discard

Unfruitful Results
Empirical Machine Translation

1. Error Detection
2. Error Analysis
3. Refinement
4. Implementation
5. Test
   - OK?
     - YES: Keep
     - NO: Discard

Unfruitful Results
Empirical Machine Translation

Error Detection

Error Analysis

Refinement

Implementation

Test

OK?

Yes

Keep

No

Discard

Unfruitful Results

No explicit use of linguistic information + Source context is not fully exploited
Empirical Machine Translation

EAMT Best Thesis Award 2008

Empirical Machine Translation and its Evaluation

Jesús Giménez

Error Detection → Error Analysis → Refinement → Implementation → Test

Incorporation of linguistic knowledge + Discriminative Learning

OK? → Keep

NO → Discard

Unfruitful Results
Discriminative Phrase Translation

A brilliant play written by William Locke

Una obra brillante escrita por William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”

→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Discriminative Phrase Translation

A brilliant play written by William Locke

Una obra brillante escrita por William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”

→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Discriminative Phrase Translation

A brilliant play written by William Locke

Una obra brillante escrita por William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”

→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Discriminative Phrase Translation

A brilliant play written by William Locke

Una obra brillante escrita por William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”

→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Discriminative Phrase Translation

A brilliant play written by William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”
→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Discriminative Phrase Translation

A brilliant play written by William Locke

Una obra brillante escrita por William Locke

vs. “A brilliant play by Lionel Messi that produced a wonderful goal”

→ “Una brillante jugada de Lionel Messi que resultó en un bello gol”
Idea

- Use discriminative Machine Learning techniques in SMT to estimate $P(e_i|f_j)$, actually, $P(e_i|f_j, context_{f_j})$
- Translation modeling is addressed as a classification problem
Idea

- Use discriminative Machine Learning techniques in SMT to estimate $P(e_i|f_j)$, actually, $P(e_i|f_j, \text{context}_{f_j})$
- Translation modeling is addressed as a classification problem
Contributions

1. **Discriminative Phrase Selection for SMT**
   - Shallow-syntactic features
     - word, lemma, parts of speech, chunking
     - local / global context
   → Improved translation quality

2. **Domain Dependence in SMT**
   - Parliament proceedings → Dictionary definitions
   - Adaptation based on:
     - EuroWordNet
     - out-of-domain data
     - a small amount of in-domain data
   → Improved translation quality
Contributions

1. **Discriminative Phrase Selection for SMT**
   - Shallow-syntactic features
     - word, lemma, parts of speech, chunking
     - local / global context
   → Improved translation quality

2. **Domain Dependence in SMT**
   - Parliament proceedings → Dictionary definitions
   - Adaptation based on:
     - EuroWordNet
     - out-of-domain data
     - a small amount of in-domain data
   → Improved translation quality
Contributions

1. Discriminative Phrase Selection for SMT
   - Shallow-syntactic features
     - word, lemma, parts of speech, chunking
     - local / global context
   → Improved translation quality

2. Domain Dependence in SMT
   - Parliament proceedings → Dictionary definitions
   - Adaptation based on:
     - EuroWordNet
     - out-of-domain data
     - a small amount of in-domain data
   → Improved translation quality
Contributions

1. Discriminative Phrase Selection for SMT
   - Shallow-syntactic features
     - word, lemma, parts of speech, chunking
     - local / global context
   \[\rightarrow\] Improved translation quality

2. Domain Dependence in SMT
   - Parliament proceedings \[\rightarrow\] Dictionary definitions
   - Adaptation based on:
     - EuroWordNet
     - out-of-domain data
     - a small amount of in-domain data
   \[\rightarrow\] Improved translation quality
... and its Evaluation

- Error Detection
- Error Analysis
- Refinement
- Implementation
- Test
- OK?
  - YES: Keep
  - NO: Discard

Unfruitful Results

Jesus Giménez
Empirical Machine Translation and its Evaluation

Jesús Giménez

![Diagram]

- Error Detection
- Error Analysis
- Refinement
- Implementation
- Test
- OK?
  - YES: Keep
  - NO: Discard

Evaluation Methods

Unfruitful Results
... and its Evaluation

Error Detection
Error Analysis
Refinement
Implementation
Test

Evaluation Methods

Keep
Unfruitful Results

Discard
YES
NO

Jesus Gimenez

EAMT Best Thesis Award 2008
Empirical Machine Translation and its Evaluation
Jesús Giménez
## Limits of Lexical Similarity

<table>
<thead>
<tr>
<th>Candidate Translation</th>
<th>On Tuesday several missiles and mortar shells fell in southern Israel, but there were no casualties.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Translation</td>
<td>Several Qassam rockets and mortar shells fell today, Tuesday, in southern Israel without causing any casualties.</td>
</tr>
</tbody>
</table>

Only one 4-gram in common!
## Limits of Lexical Similarity

<table>
<thead>
<tr>
<th>Candidate Translation</th>
<th>On Tuesday several missiles and mortar shells fell in southern Israel, but there were no casualties.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Translation</td>
<td>Several Qassam rockets and mortar shells fell today, Tuesday, in southern Israel without causing any casualties.</td>
</tr>
</tbody>
</table>

Only one 4-gram in common!
Limits of Lexical Similarity

<table>
<thead>
<tr>
<th>Candidate</th>
<th>On Tuesday several missiles and mortar shells fell in southern Israel, but there were no casualties.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>Several Qassam rockets and mortar shells fell today, Tuesday, in southern Israel without causing any casualties.</td>
</tr>
</tbody>
</table>

Only one 4-gram in common!
Idea

- Define similarity measures based on deeper linguistic information
  - Compare linguistic structures and their lexical realizations

Linguistic levels

- Syntax
  - Parts-of-speech
  - Base phrase chunks
  - Phrase constituents
  - Dependency relationships

- Semantics
  - Named entities
  - Semantic roles
  - Discourse representations
On Tuesday several missiles and mortar shells fell in southern Israel no casualties.

, but

EAMT Best Thesis Award 2008

Empirical Machine Translation and its Evaluation

Jesús Giménez
On Tuesday several missiles and mortar shells fell in southern Israel. There were no casualties, but in southern Israel.
Several Qassam rockets and mortar shells fell, in southern Israel without any casualties, causing casualties today Tuesday in VP PP LOC PP ADV S NP TMP, NP, without S VP causing any casualties EAMT Best Thesis Award 2008 Empirical Machine Translation and its Evaluation Jesús Giménez
Several Qassam rockets and mortar shells fell without causing any casualties in southern Israel, in without today Tuesday TMP, in without southern Israel VP causing any casualties NP.

NP A1 A0 NP and NP fell NP PP LOC VP PP ADV.

Linguistic measures provide more reliable rankings when the systems under evaluation are based on different paradigms (statistical vs. rule-based, fully-automatic vs. human-aided).

Linguistic measures have proven effective in shared tasks (WMT 2007-2009) both at the system and sentence levels.

Some linguistic measures suffer a substantial quality decrease at the sentence level (due to parsing errors!).

Lexical and Linguistic measures are complementary → suitable for being combined!
Linguistic measures provide more reliable rankings when the systems under evaluation are based on different paradigms (statistical vs. rule-based, fully-automatic vs. human-aided).

Linguistic measures have proven effective in shared tasks (WMT 2007-2009) both at the system and sentence levels.

Some linguistic measures suffer a substantial quality decrease at the sentence level (due to parsing errors!).

Lexical and Linguistic measures are complementary → suitable for being combined!
Contributions

1. Linguistic measures provide more reliable rankings when the systems under evaluation are based on different paradigms (statistical vs. rule-based, fully-automatic vs. human-aided)

2. Linguistic measures have proven effective in shared tasks (WMT 2007-2009) both at the system and sentence levels

3. Some linguistic measures suffer a substantial quality decrease at the sentence level (due to parsing errors!)

4. Lexical and Linguistic measures are complementary → suitable for being combined!
Linguistic measures provide more reliable rankings when the systems under evaluation are based on different paradigms (statistical vs. rule-based, fully-automatic vs. human-aided).

Linguistic measures have proven effective in shared tasks (WMT 2007-2009) both at the system and sentence levels.

Some linguistic measures suffer a substantial quality decrease at the sentence level (due to parsing errors!)

Lexical and Linguistic measures are complementary → suitable for being combined!
Contributions

1. Linguistic measures provide more reliable rankings when the systems under evaluation are based on different paradigms (statistical vs. rule-based, fully-automatic vs. human-aided)

2. Linguistic measures have proven effective in shared tasks (WMT 2007-2009) both at the system and sentence levels

3. Some linguistic measures suffer a substantial quality decrease at the sentence level (due to parsing errors!)

4. Lexical and Linguistic measures are complementary
   → suitable for being combined!
Acknowledgements

1. **Spanish Government**
   - Ministry of Science and Technology
   - Ministry of Education

2. Organizers and participants of the NIST, WMT and IWSLT Evaluation Campaigns

3. A number of NLP researchers worldwide sharing their software

4. Enrique Amigó and German Rigau
Acknowledgements

- Spanish Government
  - Ministry of Science and Technology
  - Ministry of Education

- Organizers and participants of the NIST, WMT and IWSLT Evaluation Campaigns

- A number of NLP researchers worldwide sharing their software

- Enrique Amigó and German Rigau
Acknowledgements

1. Spanish Government
   - Ministry of Science and Technology
   - Ministry of Education

2. Organizers and participants of the NIST, WMT and IWSLT Evaluation Campaigns

3. A number of NLP researchers worldwide sharing their software

4. Enrique Amigó and German Rigau
Acknowledgements

1. Spanish Government
   - Ministry of Science and Technology
   - Ministry of Education

2. Organizers and participants of the NIST, WMT and IWSLT Evaluation Campaigns

3. A number of NLP researchers worldwide sharing their software

4. Enrique Amigó and German Rigau
Empirical Machine Translation and its Evaluation
EAMT Best Thesis Award 2008

Thanks!