

A Syntactic Skeleton for Statistical Machine Translation

Bart Mellebeek Karolina Owczarzak Declan Groves
Josef van Genabith Andy Way

National Centre for Language Technology
School of Computing
Dublin City University

European Association for Machine Translation, 2006



Outline

- 1 Motivation
- 2 Implementation
 - Architecture
 - Worked example
- 3 Evaluation
 - Experimental Setup
 - Results

Outline

- 1 Motivation
- 2 Implementation
 - Architecture
 - Worked example
- 3 Evaluation
 - Experimental Setup
 - Results

Motivation

Difficulties for SMT

- Structural aspects of language hard to model.
- At best, only local context taken into account.

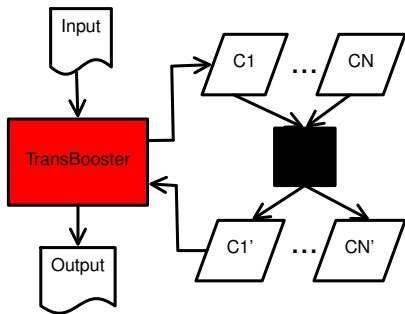
Solution = Enhance SMT with Syntax

- Incorporate syntax into SMT models (previous research).
- Use syntax as wrapper around SMT (our approach).

Outline

- 1 Motivation
- 2 Implementation**
 - Architecture
 - Worked example
- 3 Evaluation
 - Experimental Setup
 - Results

TransBooster = Wrapper Technology for MT



- Boost the output quality of MT systems by automatic sentence decomposition.
- Baseline MT system performs translation.
- Baseline MT system treated as black box.
- Maximise MT system's own possibilities.

1. Decomposition of Input

$[SAT_1] \dots [SAT_l]$ pivot $[SAT_{l+1}] \dots [SAT_{l+r}]$

- Input = syntactic parse of string.
- Decompose into pivot and satellites.

Example

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
 [fast and confidential deals]_{ARG₂}.

2. Skeletons and Substitution Variables

$$[V_{SAT_1}] \dots [V_{SAT_i}] \text{ pivot } [V_{SAT_{i+1}}] \dots [V_{SAT_{i+r}}]$$

- V_{SAT_i} = a simpler string substituting SAT_i
- Reduce complexity original constituents \rightarrow better translation pivot.
- Track location satellites in target.

Example

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
 [fast and confidential deals]_{ARG₂}.

2. Skeletons and Substitution Variables

$$[V_{SAT_1}] \dots [V_{SAT_i}] \text{ pivot } [V_{SAT_{i+1}}] \dots [V_{SAT_{i+r}}]$$

- V_{SAT_i} = a simpler string substituting SAT_i
- Reduce complexity original constituents \rightarrow better translation pivot.
- Track location satellites in target.

Example

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
 [fast and confidential deals]_{ARG₂}.

2. Skeletons and Substitution Variables

$$[V_{SAT_1}] \dots [V_{SAT_i}] \text{ pivot } [V_{SAT_{i+1}}] \dots [V_{SAT_{i+r}}]$$

- V_{SAT_i} = a simpler string substituting SAT_i
- Reduce complexity original constituents \rightarrow better translation pivot.
- Track location satellites in target.

Example

[The chairman] _{V_{ARG_1}} [likes] _{$pivot$} [deals] _{V_{ARG_2}} .

2. Skeletons and Substitution Variables

$$[V_{SAT_1}] \dots [V_{SAT_i}] \text{ pivot } [V_{SAT_{i+1}}] \dots [V_{SAT_{i+r}}]$$

- V_{SAT_i} = a simpler string substituting SAT_i
- Reduce complexity original constituents \rightarrow better translation pivot.
- Track location satellites in target.

Example

[The boy] _{V_{ARG_1}} [likes]_{*pivot*} [books] _{V_{ARG_2}} .

3. Translating Satellites

- If complexity SAT_i not OK \rightarrow recursion.
- If complexity SAT_i OK \rightarrow translate.
- Embed SAT_i in context for optimal translation.

Example

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

3. Translating Satellites

- If complexity SAT_i not OK \rightarrow recursion.
- If complexity SAT_i OK \rightarrow translate.
- Embed SAT_i in context for optimal translation.

Example

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
 [fast and confidential deals]_{ARG₂}.

3. Translating Satellites

- If complexity SAT_i not OK \rightarrow recursion.
- If complexity SAT_i OK \rightarrow translate.
- Embed SAT_i in context for optimal translation.

Example

[The chairman likes]_{Context} [fast and confidential deals]_{ARG₂}.

3. Translating Satellites

- If complexity SAT_i not OK \rightarrow recursion.
- If complexity SAT_i OK \rightarrow translate.
- Embed SAT_i in context for optimal translation.

Example

[The boy sees]_{Context} [**fast and confidential deals**]_{ARG₂}.

4. Composing the Translation

- Recursive procedure.
- Only syntactically simple chunks sent to MT engine.
- Substitution variables track location of translation chunks in target.
- Therefore composing translation is straightforward.

Worked example

Source

The chairman, a long-time rival of Bill Gates, likes fast and confidential deals.

Target

El presidente, rival de largo plazo de Bill Gates, gustos ayuna y los repartos confidenciales.

Worked example

Source

The chairman, a long-time rival of Bill Gates, **likes** fast and confidential deals.

Target

El presidente, rival de largo plazo de Bill Gates, **gustos** ayuna y los repartos confidenciales.

Worked example

Source

The chairman, a long-time rival of Bill Gates, likes **fast** and confidential deals.

Target

El presidente, rival de largo plazo de Bill Gates, gustos **ayuna** y los repartos confidenciales.

Worked example

Decomposition

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

Substitution Variables

[The chairman]_{V_{ARG₁}} [likes]_{pivot} [deals]_{V_{ARG₂}}.

Source

The chairman likes deals.

Target

El presidente tiene gusto de repartos.

Worked example

Decomposition

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

Substitution Variables

[The chairman]_{V_{ARG₁}} [likes]_{pivot} [deals]_{V_{ARG₂}}.

Source

The chairman likes deals.

Target

El presidente tiene gusto de repartos.

Worked example

Decomposition

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

Substitution Variables

[The chairman]_{V_{ARG₁}} [likes]_{pivot} [deals]_{V_{ARG₂}}.

Source

The chairman likes deals.

Target

El presidente tiene gusto de repartos.

Worked example

Decomposition

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

Substitution Variables

[The chairman]_{V_{ARG₁}} [likes]_{pivot} [deals]_{V_{ARG₂}}.

Source

The chairman **likes** deals.

Target

El presidente **tiene gusto de**
repartos.

Worked example

Decomposition

[The chairman, a long-time rival of Bill Gates,]_{ARG₁} [likes]_{pivot}
[fast and confidential deals]_{ARG₂}.

Substitution Variables

[The chairman]_{V_{ARG₁}} [likes]_{pivot} [deals]_{V_{ARG₂}}.

Source

The chairman likes deals.

Target

El presidente tiene gusto de repartos.

Worked example

ARG₁ is simple enough

→ translate 'fast and confidential deals' in context.

Source

The chairman likes fast and confidential deals.

Target

El presidente tiene gusto de repartos rápidos y confidenciales.

ARG₂ is too complex

→ recurse into 'The chairman, a long-time rival of Bill Gates'.

Worked example

ARG₁ is simple enough

→ translate 'fast and confidential deals' in context.

Source

The chairman likes fast and confidential deals.

Target

El presidente tiene gusto de repartos rápidos y confidenciales.

ARG₂ is too complex

→ recurse into 'The chairman, a long-time rival of Bill Gates'.

Worked example

ARG₁ is simple enough

→ translate 'fast and confidential deals' in context.

Source

The chairman likes **fast and confidential deals**.

Target

El presidente tiene gusto de **repartos rápidos y confidenciales**.

ARG₂ is too complex

→ recurse into 'The chairman, a long-time rival of Bill Gates'.

Worked example

ARG₁ is simple enough

→ translate ‘fast and confidential deals’ in context.

Source

The chairman likes fast and confidential deals.

Target

El presidente tiene gusto de repartos rápidos y confidenciales.

ARG₂ is too complex

→ recurse into ‘The chairman, a long-time rival of Bill Gates’.

Worked example

Source

The chairman, a long-time rival of Bill Gates, likes fast and confidential deals.

Baseline MT

El presidente, rival de largo plazo de Bill Gates, gustos ayuna y los repartos confidenciales.

TransBooster

El presidente, un rival de largo plazo de Bill Gates, tiene gusto de repartos rápidos y confidenciales.

Worked example

Source

The chairman, a long-time rival of Bill Gates, **likes** fast and confidential deals.

Baseline MT

El presidente, rival de largo plazo de Bill Gates, **gustos** ayuna y los repartos confidenciales.

TransBooster

El presidente, un rival de largo plazo de Bill Gates, **tiene gusto de** repartos rápidos y confidenciales.

Worked example

Source

The chairman, a long-time rival of Bill Gates, likes **fast** and confidential deals.

Baseline MT

El presidente, rival de largo plazo de Bill Gates, gustos **ayuna** y los repartos confidenciales.

TransBooster

El presidente, un rival de largo plazo de Bill Gates, tiene gusto de repartos **rápidos** y confidenciales.

Outline

- 1 Motivation
- 2 Implementation
 - Architecture
 - Worked example
- 3 Evaluation
 - Experimental Setup
 - Results

Baseline MT = phrase-based SMT

- English → Spanish.
- Training data: 501K sentence pairs Europarl (Max. 40 words).
- Translation Model: IBM4 (Giza++).
- Language Model: trigram LM, modified Kneser-Ney smoothing (SRI toolkit).
- Decoder: Pharaoh → phrase tables from Giza++ word alignments.

Two Test Sets: Europarl and Wall Street Journal

- Test set 1: 800 sentences WSJ
- Test set 2: 800 sentences Europarl
- Reason: two 'out-of-domain' phenomena.
- TransBooster performs better on test set 1.
- Pharaoh performs better on test set 2.

Results on Europarl

English→Spanish	BLEU	NIST
Pharaoh	0.1986	5.8393
TransBooster	0.2052	5.8766
Percent. of Baseline	103.3%	100.6%

Results on Wall Street Journal

English→Spanish	BLEU	NIST
Pharaoh	0.1343	5.1432
TransBooster	0.1379	5.1259
Percent. of Baseline	102.7%	99.7%

Manual Evaluation

200 Sentences Europarl	
Better	20%
Similar	73%
Worse	7%

Analysis improvements

- 70% word order.
- 30% lexical selection.

Conclusions

- Phrase-based SMT improves when adding layer of syntactic guidance.
- Improvement of 3.3% BLEU score relative.
- Approach shown to improve rule-based systems.
- Approach viable for various MT architectures.

Questions

Thanks for listening.

Example 1

Original	Despite an impressive number of international studies , there is still no clear evidence of any direct link between violence and media consumption.
Pharaoh	a pesar de los estudios internacionales , todavía no existe ninguna relación directa entre la violencia y media un número impresionante pruebas claras de consumo
TB	pese a un número impresionante de estudios internacionales , todavía no hay pruebas claras de ninguna relación directa entre la violencia y los medios consumo

Example 1

- Original Despite **an impressive number** of international studies , there is still no clear evidence of any direct link between violence and media consumption.
- Pharaoh a pesar de los estudios internacionales , todavía no existe ninguna relación directa entre la violencia y media **un número impresionante** pruebas claras de consumo
- TB pese a **un número impresionante** de estudios internacionales , todavía no hay pruebas claras de ninguna relación directa entre la violencia y los medios consumo

Example 1

Original	Despite an impressive number of international studies , there is still no clear evidence of any direct link between violence and media consumption.
Pharaoh	a pesar de los estudios internacionales , todavía no existe ninguna relación directa entre la violencia y media un número impresionante pruebas claras de consumo
TB	pese a un número impresionante de estudios internacionales , todavía no hay pruebas claras de ninguna relación directa entre la violencia y los medios consumo

Example 2

Original	Presidency communication on the situation in the Middle East
Pharaoh	presidencia comunicación sobre la situación en el mediterráneo
TB	presidencia comunicación sobre la situación en el cercano oriente

Example 2

Original	Presidency communication on the situation in the Middle East
Pharaoh	presidencia comunicación sobre la situación en el mediterráneo
TB	presidencia comunicación sobre la situación en el cercano oriente