

# Using a Grammar Checker for Evaluation and Postprocessing of Statistical Machine Translation

Sara Stymne and Lars Ahrenberg  
Linköping University, Sweden

LREC  
May 20, 2010

GSLT



Linköping University

 **Santa Anna**  
IT Research Institute AB

# Outline

- 1 Introduction
  - Overview
  - SMT system
  - Grammar checker
- 2 Grammar checker for evaluation
- 3 Grammar checker for postprocessing
- 4 Conclusions

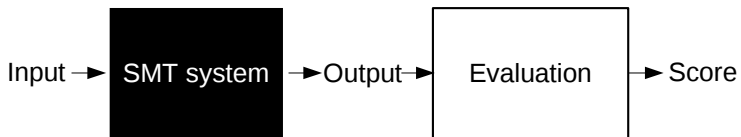
# Grammar checker for SMT

- Evaluation
  - Assess grammaticality of MT output
- Postprocessing
  - Improve the output of an SMT system by applying grammar checker suggestions

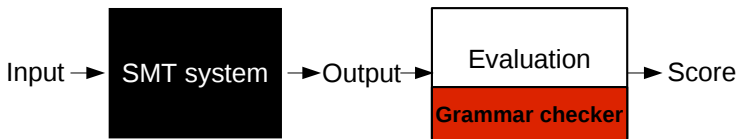
# Grammar checker for SMT

- Evaluation
  - Assess grammaticality of MT output
- Postprocessing
  - Improve the output of an SMT system by applying grammar checker suggestions
- *Preprocessing*
  - *Help a (rule-based) system by standardising its input*

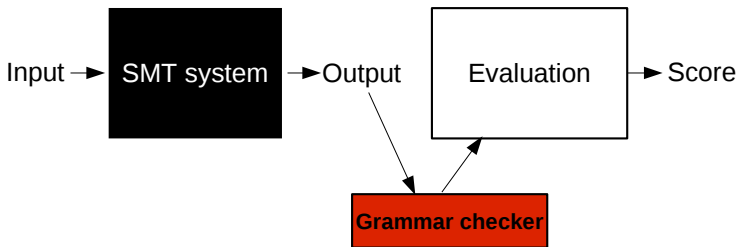
# Basic SMT system pipeline



# Pipeline with grammar checker for evaluation



# Pipeline with grammar checker for postprocessing



# SMT system

- Standard phrase-based statistical MT system:

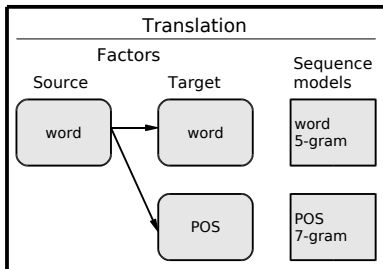
$$\hat{t} = \arg \max_t \left( \sum_{m=1}^M \lambda_m h_m(t, s) \right)$$

- Factored translation
- Tools
  - Moses
  - SRILM
  - Giza++
- English to Swedish
- Trained on Europarl data



# SMT with factors

- Standard SMT: words represented by surface form
- Factored SMT: words represented as vector of features



# SMT system variation

- 6 systems, varied on two dimensions:
  - Corpus size
    - Large (701157 sentences)
    - Small (100000 sentences)
  - Output factors
    - None (jag sover)
    - POS (jag|PN sover|VB)
    - Morph (jag|PN.utr.sin.def.sub sover|VB.prs.akt)

# Grammar checker

- Granska (Domeij et al., 1999)
- Swedish grammar checker
- Developed targeted at human texts
- Hybrid, mainly rule-based:
  - Probabilistic morphological tagger
  - Spell checker
  - Rule matcher (hand-written rules)
- 13 error categories

# Grammar checker tools

- Grammar checkers are normally authoring tools
- We use it as an automatic tool

# Grammar checker tools

- Grammar checkers are normally authoring tools
- We use it as an automatic tool
- Possible to use as an authoring tool for human MT postprocessing as well

# Grammar checker – sample output

Text: Averaging vore med tre timmar per dag , det är den mest omfattande mänskliga aktivitet efter sover och - för vuxna - arbete .

Rule: stav1@stavning Span: 1-1 Words: Averaging

Rule: kong10E@kong Span: 14-15 Words: mänskliga aktivitet  
mänskliga aktiviteten  
mänsklig aktivitet

# Granska: Error analysis on SMT output

Type	Error identification		Correction suggestions
	Correct	Wrong	
Agreement NP	64	10	
Agreement Pred.	21	1	
Split compounds	12	14	
Verb	31	18	
Word order	9	0	

161 spelling errors: foreign words (49.0%) and proper names (32.9%)

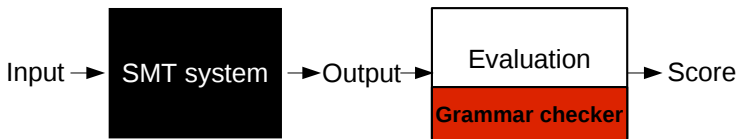
# Granska: Error analysis on SMT output

Type	Error identification		Correction suggestions			
	Correct	Wrong	Correct1	Correct2+	Wrong	None
Agreement NP	64	10	48	10	4+10	2+0
Agreement Pred.	21	1	20	–	1+1	–
Split compounds	12	14	8	–	3+13	1+1
Verb	31	18	11	2	–	18+18
Word order	9	0	8	–	1+0	–

161 spelling errors: foreign words (49.0%) and proper names (32.9%)



# Grammar checker for evaluation



# Grammar checker metrics

- Three new metrics based on Granska:
  - $GER_1$ : grammar errors/sentence (excl. bad categories)
  - $GER_2$ : grammar errors/sentence (all categories)
  - SGER: all errors/sentence

# Grammar checker metrics

- Three new metrics based on Granska:
  - $GER_1$ : grammar errors/sentence (excl. bad categories)
  - $GER_2$ : grammar errors/sentence (all categories)
  - SGER: all errors/sentence
- Only accounts for fluency, not accuracy

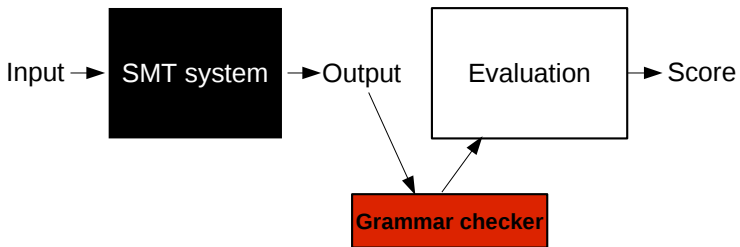
## Evaluation results

<b>Size</b>	<b>Factors</b>	<b>Bleu</b>	<b>GER<sub>1</sub></b>	<b>GER<sub>2</sub></b>	<b>SGER</b>
Large	none	22.18			
	POS	21.63			
	morph	22.04			
Small	none	21.16			
	POS	20.79			
	morph	19.45			

## Evaluation results

Size	Factors	Bleu	GER <sub>1</sub>	GER <sub>2</sub>	SGER
Large	none	<b>22.18</b>	0.196	0.293	0.496
	POS	21.63	0.228	0.304	0.559
	morph	22.04	<b>0.125</b>	<b>0.195</b>	<b>0.446</b>
Small	none	<b>21.16</b>	0.244	0.359	0.664
	POS	20.79	0.282	0.375	0.718
	morph	19.45	<b>0.121</b>	<b>0.245</b>	<b>0.600</b>

# Grammar checker for postprocessing



# Grammar checker for postprocessing

- Automatically apply first correction suggestion for the categories that had good suggestions on the error analysis:
  - Agreement errors (NP and pred)
  - Some verb errors
  - Word order errors
  - Capitalization of spelling errors

## Results and number of changes with Granska used for postprocessing

Size	Factors	Bleu	Improvement	No. of changes
Large	none	22.34	+0.16	382
	POS	21.81	+0.18	429
	morph	22.17	+0.13	259
Small	none	21.30	+0.14	456
	POS	20.95	+0.16	514
	morph	19.52	+0.07	249



## Results of postprocessing on affected subsets

<b>Size</b>	<b>Factors</b>	<b>Bleu</b>	<b>Improvement</b>	<b>No. of sentences</b>
Large	none	20.12	+0.68	335
	POS	19.61	+0.74	373
	morph	19.29	+0.82	238
Small	none	19.26	+0.54	395
	POS	18.27	+0.53	452
	morph	17.24	+0.45	241

# Analysis of the 100 first Granska-based changes for each system

<b>Size</b>	<b>Factors</b>	<b>Good</b>	<b>Neutral</b>	<b>Bad</b>
Large	none	73	19	8
	POS	77	17	6
	morph	68	19	13
Small	none	74	19	7
	POS	73	17	10
	morph	68	20	12

# Conclusions and future work

- Evaluation with grammar checker is complementary to metrics like Bleu
- Useful for postprocessing, but low coverage
- Future work:
  - Extend grammar checker coverage on SMT output
  - Create combination metric with GC features + adequacy
  - Integrate grammar checking techniques with SMT for postprocessing
  - Large scale investigation on common dataset
    - Looking for grammar checker for German, Spanish, or French!

Thank you for your attention!



Questions or comments?

# Example changes

First	Att ge nya befogenheter till en kommitté av ministrar främjas genom <i>en oansvarigt sekretariat</i> skulle inte <i>utgör</i> någon typ av framsteg ...
Changed	Att ge nya befogenheter till en kommitté av ministrar främjas genom <b>ett oansvarigt sekretariat</b> skulle inte <b>ha utgjort</b> någon typ av framsteg ...
First	Det är viktigt att fylla <i>den kulturella vakuum</i> mellan våra två regioner
Changed	Det är viktigt att fylla <b>ett kulturellt vakuum</b> mellan våra två regioner
First	Jag hör ibland sägas att rådet är så <i>engagerade</i> i Berlin ...
Changed	Jag hör ibland sägas att rådet är så <b>engagerat</b> i Berlin ...
First	Skulle det inte vara värt att ansvar på alla nivåer i <i>den beslutsfattande</i> processen tydligare, snarare än att försöka gå framåt ...
Changed	Skulle det inte vara värt att ansvar på alla nivåer i <b>det beslutsfattandet</b> processen tydligare, snarare än att försöka gå framåt ...
First	Dokumentet kommer att överlämnas till europeiska rådet i Biarritz i några <i>days'</i> tid .
Changed	Dokumentet kommer att överlämnas till europeiska rådet i Biarritz i några <b>daysar</b> tid .