Linguistically Motivated Evaluation of English-Latvian Statistical Machine Translation

Inguna SKADIŅA,
Kristīne LEVĀNE-PETROVA
Guna RĀBANTE-BUŠA
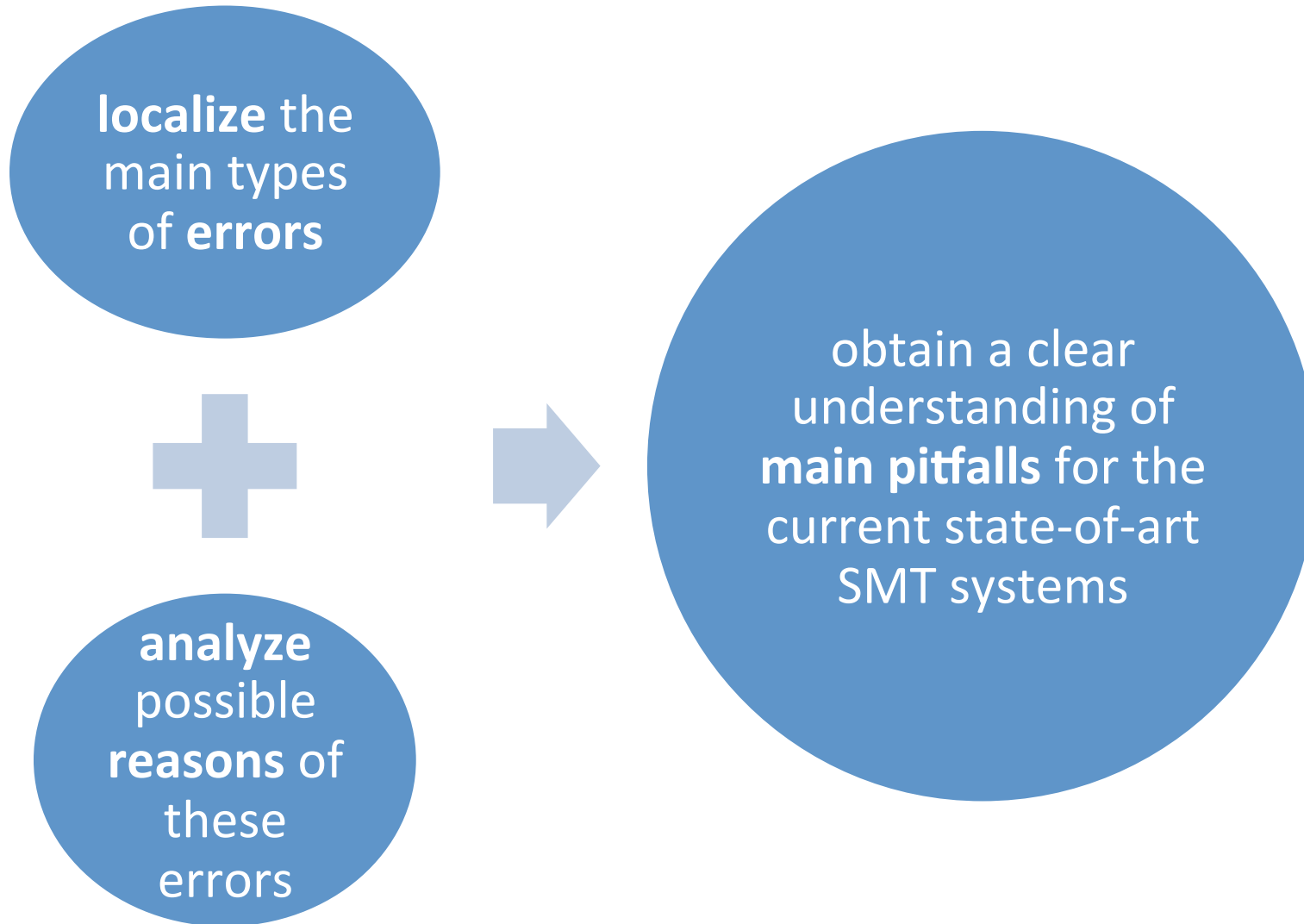
Institute of Mathematics and Computer Science,
University of Latvia

Outline

- Training resources and SMT System
- Human evaluation
  - Setup
  - Evaluation results
- Error analysis
  - Incorrect form
  - Errors in Word Order
  - Wrong Lexical Choice
  - Other Issues
- Additional results
  - Understandability
  - Comparison of Language models
- Conclusions

Aim of the study

localize the main types of **errors**

+

analyze possible **reasons** of these errors

→

obtain a clear understanding of **main pitfalls** for the current state-of-art SMT systems

TRAINING RESOURCES AND SMT SYSTEM

Training corpus

Only unique sentence (fragment) pairs included (ca. 3.3 million pairs):

- JRC Acquis parallel corpus version 2.2.
- DGT-TM corpus (released in 2007)
- EMEA corpus
- Some small additional corpora

Tools used

- *Moses toolkit* with standard training procedures
- *GIZA++ tool* for word alignments
- *SRI LM Toolkit* with recommended parameters (3-gram language model)
- Latvian tagger developed in project

English-Latvian factored SMT
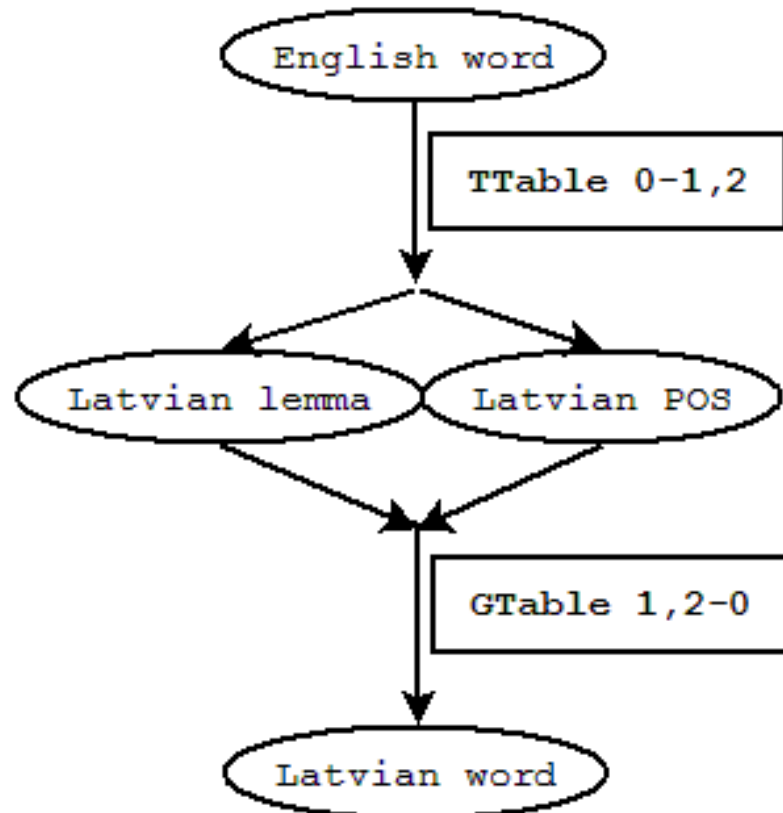
- Three factor model is chosen:
  - inflected form (0)
  - base form (or lemma) (1)
  - morphological tag (2)
- Two steps:
  - English sentence has been translated into sequence of Latvian factors 1 and 2, using translation table 0-1,2
  - Sequence of Latvian factors 1 and 2 were translated into factor 0, using generation table 1,2-0

Test corpus and automatic evaluation

- 1392 sentences were randomly selected from training data for automatic evaluation before training
- The system's performance in BLEU is **48.82**

HUMAN EVALUATION

Setup

- The evaluation was performed by three independent evaluators

- 200 sentences for evaluation

- For each sentence the annotators assigned a score:
  - **1** if they noticed this type of error occurring
  - a score of **0** if not

Evaluation framework

Vilar et al.	Our
Missing words: content words Missing words: filler words	Missing words
Error in local word order	Error in local word order
Error in local phrase order	Error in local phrase order
Error in global word order	Error in global word order
Error in global phrase order	Error in global phrase order
Wrong lexical choice	Wrong lexical choice
Incorrect form	Incorrect form
Wrong lexical choice Incorrect disambiguation	Wrong lexical choice
Extra words	Extra words
~~Style~~	
~~Idioms~~	
Unknown words	Unknown words
~~Punctuation~~	

Evaluation results

Error type	All evaluators	Two evaluators	Total (at least 2 evaluators)
Missing words	1	17	18 (7.62%)
Error in local word order	0	3	3 (1. 27%)
Error in local phrase order	8	17	25 (10.59%)
Error in global word order	4	11	15 (6.36%)
Error in global phrase order	0	3	3 (1.27%)
Wrong lexical choice	9	21	30 (12.71%)
Incorrect form	**68**	**46**	**114 (48.31%)**
Extra words	5	17	22 (9.32%)
Unknown words	5	1	6(2.54%)
Total	**99**	**137**	**236**

ERROR ANALYSIS

Incorrect form

- 114 sentences (57% from all translations)
- In 42 cases (36.84%) this was the only error in sentence.
- The rest of errors in incorrect word form occurs together with errors in
  - word order (39 cases)
  - wrong lexical choice (21 case)
  - extra words (14 cases)
  - missing words (14 cases)

EN: *.. financing schemes have to contribute to <u>high collection rates</u>..*

LV(MT): *.. finansēšanas shēmām ir jāsekmē <u>augsti savākšanas likmes</u>..*

Errors in word order: Translation of long genitive noun phrases

---

EN:
..council directive 91/492/eec of 15 july 1991 laying down the health conditions for <u>the production and placing on the market of live bivalve molluscs</u>..

LV (MT):
..padomes 1991. gada 15. jūlija direktīvu 91/492/ eek, ar ko nosaka veselības nosacījumus attiecībā uz <u>ražošanu un laišanu tirgū dzīvas gliemenes</u>..

LV (ref.)
..padomes 1991. gada 15. jūlija direktīva 91/492/ eek, ar ko nosaka veselības aizsardzības prasības attiecībā uz <u>dzīvu gliemeņu ražošanu un laišanu tirgū</u>..

Errors in word order

- The date — *16 may 2003* is incorrectly translated as *gada 16. maija 2003 (year 16 May 2003)*, instead of *2003. gada 16. maija*

- Wrong transformation of long genitive phrase consisting of two *of* phrases – the rightmost genitive phrase in English needs to be transformed into leftmost genitive phrase in Latvian

- Incorrect form — the word *lēmums (decision)* should be in accusative as it is determined by preposition *ar (with)*

EN:	*..annex xix to the agreement was amended <u>by decision</u> of the eea joint committee no <u>47/2003 of 16 may 2003</u>..*
LV (MT):	*..līguma xix pielikums ir grozīts ar eez apvienotās komitejas lēmums nr. 47/2003 gada 16. maija 2003..*
LV (ref.)	*..līguma xix pielikums grozīts <u>ar</u> eez apvienotās komitejas 2003. gada 16. maija lēmumu nr 47/2003..*

Wrong lexical choice: out of domain translation

EN: *medicinal product is used in the treatment of* <u>*conditions*</u> *which must be diagnosed in a hospital*

LV (MT): *zāles izmanto, ārstējot* <u>*nosacījumus*</u>*, kas ir jādiagnosticē slimnīcā*

LV (ref.) *zāles izmanto* <u>*slimību*</u> *ārstēšanai , kas jādiagnosticē slimnīcā*

Wrong lexical choice: influence of human translator

EN:	_conversion rates_ should be established for member states outside the euro zone
LV(MT):	_konversijas likmes_, kas jānosaka dalībvalstīm ārpus eiro zonas
LV(human):	_maiņas kurss_, kas jānosaka dalībvalstīm ārpus eiro zonas

Missing words

EN: *orders the applicants to bear , in addition to their own costs , the costs <u>incurred</u> by the council and the commission*

LV(MT): *pasūtījumu prasītāji atlīdzina , papildus savām izmaksām , [tiesāšanās] izmaksas , ko padome un komisija*

LV(ref.): *prasītājas sedz savus , kā arī atlīdzina padomes un komisijas <u>tiesāšanās</u> izdevumus*

SOME ADDITIONAL RESULTS

Understandability

Perfect	Understand-able	Hard to understand	Incomprehen sible
57 (28.5%)	86 (43%)	51 (25.5%)	6 (3%)

- 4 point scale:
  - 3 — perfect translation, no need for editing;
  - 2 —understandable translation, small form/grammar errors needs to be corrected;
  - 1 — hard to understand;
  - 0 — incomprehensible translation
- **71.5%** of translations can be classified as good

Comparison of Language Models: Automatic evaluation

From 200 sentences in our test corpus:

- 54 (27%) translations were identical for all three SMT systems
- 68 (34%)translations of 3-gram model and 5-gramm model were identical
- 121 translation (60.5%) was identical for 5-gram and 7-gram model

Language model	BLEU score
3-gram	48.82
5-gram	50.97
7-gram	51.67

Comparison of Language Models: Human evaluation

All Equal	3-gram	3-gram and 5-gram	3-gram and 7-gram	5-gram	5-gram and 7-gram	7-gram
77	8	3	1	11	33	13

146 sentences from the test corpus

Conclusion and future work

- English-Latvian SMT system can reach reasonably good quality (i.e. in 71.5% of cases translation is understandable and can be used with small modifications)
- Even using factored models with morphological information, still a lot of inflectional errors occur (57% of all translations)
- The most frequent error type is incorrect word form (48.31%), followed by different errors in word order (19.49%) and wrong lexical choice (12.71%)
- We see two ways how we can improve output of our SMT system:
  - by applying syntax-based techniques
  - by incorporating of additional linguistic knowledge through hybrid approach

Thank you!