Better Punctuation Prediction with Hierarchical Phrase-Based Translation

Stephan Peitz, Markus Freitag and Hermann Ney

peitz@cs.rwth-aachen.de

IWSLT 2014, Lake Tahoe, CA
December 4th, 2014

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Outline

- Introduction
- Modeling Punctuation Prediction as Machine Translation
- Hierarchical Phrase-based Translation
- Experimental Evaluation
- Conclusion

Introduction

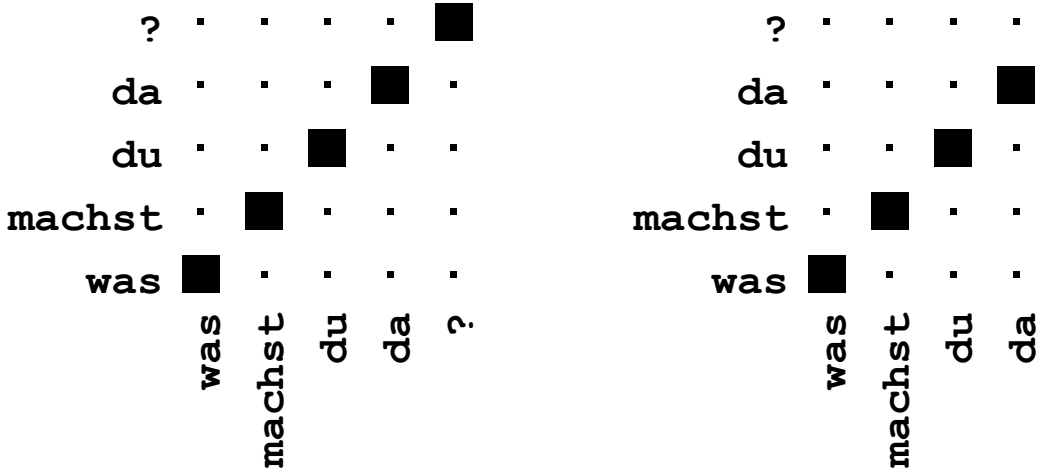
- Spoken language translation (SLT)
  - Automatic speech recognition (ASR)
  - Machine translation (MT)
- In speech punctuation is not made explicitly
  - ASR systems provide an output without punctuation marks
  - MT systems are trained on data with proper punctuation
- Reintroduce punctuation marks with monolingual translation
  - Translate from unpunctuated text to text with punctuation
  - Based on phrase-based translation
- In this work
  - Use hierarchical instead of phrase-based translation
  - Investigation of the optimization criterion

Introduction

- **Monolingual translation system**
  - **More features besides the language model**
  - **Scaling factors can be tuned**
- **Phrase-based translation**
  - **Sequence of words are translated at once**
  - **Local contextual information is preserved**
  - **Useful to predict punctuation depending of its surrounding words (e.g. commas)**
  - **Limitation: dependencies beyond the local context**
- **Hierarchical phrase-based translation**
  - **Discontinuous phrases with “gaps”**
  - **Capture long-range dependencies between words and punctuation marks**

Modeling Punctuation Prediction as Machine Translation

- **Translation Model**
  - Extract from a pseudo-bilingual corpus
  - Take monolingual corpus as source and target text
  - Create monotone alignment
  - Remove punctuation marks from the source text
  - Punctuation marks in the target sentences become unaligned



Modeling Punctuation Prediction as Machine Translation

- **Optimization**
  - **Remove punctuation marks from a development set**
  - **Use the original development set as reference**
  - **Tune scaling factors with MERT [Och 03]**
- **Prediction performance is measured with the F_1-Score**
  - **Use F_α-Score rather than BLEU as a more suitable optimization criterion**

$$F_\alpha = (1+\alpha) \cdot \frac{(precision \cdot recall)}{\alpha \cdot precision + recall}$$

- **By varying α, more emphasis can be put on recall or precision**

Modeling Punctuation Prediction as Machine Translation

- Language model
  - Trained on monolingual corpora with proper punctuation
- Decoding
  - Translate from unpunctuated text to text with punctuation
  - Monotone, no reordering model is necessary
- In this work
  - Perform prediction before the actual translation
  - Final machine translation system has not to be retrained

[Ma & Tinsley+ 08, Peitz & Freitag+ 11]

Hierarchical Phrase-based Translation

- Allow discontinuous phrases with “gaps”
- Obtain phrases from word-aligned bilingual training data
  - Sub-phrases within a phrase are replaced by a generic non-terminal X
  - Maximum of two gaps per rule

$$X \rightarrow \langle \text{über } X_0 \text{ hinausgehen } X_1, \text{go beyond } X_0\ X_1 \rangle$$

- Reordering is modelled implicitly
- Formalized as a synchronous context-free grammar (SCFG)
- Speaking of *rules* rather than phrases

[Chiang 05]

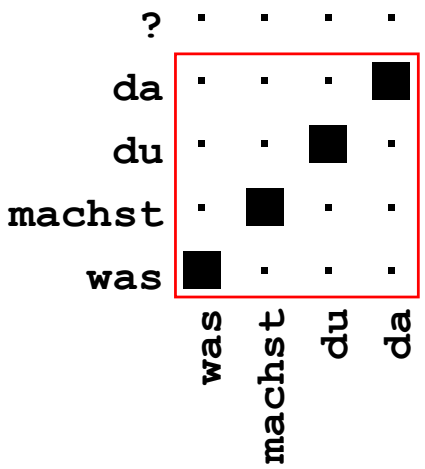
Punctuation Prediction based on Hierarchical Translation

- **Aim: model dependencies between words and punctuation marks**
  - **e.g. relationship between question word (“was”) and question mark**

$$X \rightarrow \langle \textbf{was } X_0, \textbf{was } X_0 \textbf{ ?} \rangle$$
$$X \rightarrow \langle \textbf{machst du } X_0, \textbf{machst du } X_0 \textbf{ ?} \rangle$$

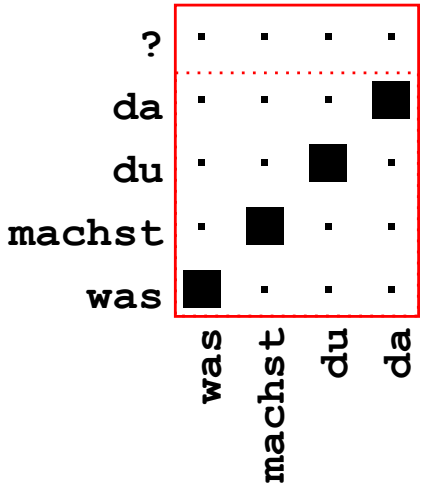
- **Restrictions**
  - **Performing monotone translation**
  - **Reordering is not necessary**
  - **Rules with one non-terminal maximum is sufficient**

Additional Extraction Heuristic



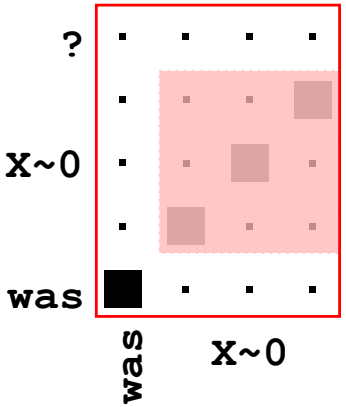
$$X \rightarrow \langle \textbf{was machst du da}, \textbf{was machst du da} \rangle$$

Additional Extraction Heuristic



$$X \rightarrow \langle \textbf{was machst du da, was machst du da ?} \rangle$$

Additional Extraction Heuristic



$$X \rightarrow \langle \text{was machst du da, was machst du da ?} \rangle$$

$$X \rightarrow \langle \text{machst du da, machst du da} \rangle$$

$$X \rightarrow \langle \text{was } X_0, \text{was } X_0 \text{ ?} \rangle$$

Experimental Evaluation

- **Evaluation of prediction performance**
  - **Removed punctuation from provided development and test sets (manual transcriptions, no ASR errors)**
  - **Measurement: Precision, Recall and F_1-Score**
  - **Optimization criteria: BLEU and F_α-Score with $\alpha = \{1, 2, 3, 4\}$**
  - **Phrase-based (PBT) vs. hierarchical translation (HPBT)**
  - **Comparison against HIDDEN-NGRAM [Stolcke 02]**
- **Evaluated on the IWSLT 2014 translation tasks**
  - **German→English and English→French**
- **Translation models were trained on indomain data**
- **Language model was trained on all available data**

Prediction Results

- From unpunctuated German text to German with punctuation marks

system	tuned on	Prec.	Rec.	F_1
PBT	BLEU	82.7	67.5	74.3
	F_1	82.6	67.5	74.3
	F_2	78.3	71.4	74.7
	F_3	76.6	72.2	74.4
	F_4	72.5	73.6	73.0
HPBT	BLEU	86.4	65.5	74.7
	F_1	81.8	71.0	76.0
	F_2	77.0	75.4	76.2
	F_3	75.9	75.2	75.6
	F_4	71.8	73.7	74.2
HIDDEN-NGRAM	-	82.7	69.5	75.5

- HIDDEN-NGRAM outperforms PBT in terms of F_1
- HPBT tuned on F_2 works best

Analysis

- Were hierarchical rules used in the decoding process?

system	tuned on	lexical rules	hierarchical rules
PBT	BLEU	2313	-
PBT	F_2	2549	-
HPBT	F_2	2234	442

- All applied hierarchical rules introduced punctuation marks

Analysis

Input "was machst du nur"

- PBT "was machst du nur ."
- Applied phrases
  - ⟨was machst du, was machst du⟩
  - ⟨nur, nur .⟩

- HPBT "was machst du nur ?"
- Applied rules
  - $X \rightarrow$ ⟨was, was⟩
  - $X \rightarrow$ ⟨machst du $X^{\sim 0}$, machst du $X^{\sim 0}$?⟩
  - $X \rightarrow$ ⟨nur, nur⟩

Impact on Translation Quality

- **Translation tasks:**
  - **English→French**
  - **German→English**
- **Tested on enriched manual and automatic transcription**
- **Applied baseline phrase-based MT systems trained on all available data**
- **Measurement: BLEU**

Impact on Translation Quality

- German→English
- WER of automatic transcription: 21.6%

system	tuned on	Prec.	Rec.	F_1	transcription manual BLEU	transcription automatic BLEU
PBT	BLEU	82.7	67.5	74.3	27.3	18.7
PBT	F_2	78.3	71.4	74.7	27.5	18.6
HPBT	F_2	77.0	75.4	76.2	27.7	19.1
HIDDEN-NGRAM	-	82.7	69.5	75.5	27.2	19.0
correct punctuation					29.4	-

- Prediction using HPBT seems to help
- Only small improvement on automatic transcription

Impact on Translation Quality

- English→French
- WER of automatic transcription: 16.7%

system	tuned on	Prec.	Rec.	F_1	transcription manual BLEU	transcription automatic BLEU
PBT	BLEU	81.2	67.6	73.7	28.4	22.6
PBT	F_2	72.2	75.0	73.6	28.6	22.8
HPBT	F_2	74.8	77.1	75.9	28.9	22.7
HIDDEN-NGRAM	-	82.0	60.2	69.4	27.0	21.7
correct punctuation					31.9	-

- Prediction using monolingual MT systems works best
- Mixed results on automatic transcription

Conclusion

- **Punctuation prediction based hierarchical translation**
  - **Capture long-range dependencies between words punctuation marks**
  - **Improvements in terms of Precision, Recall and F_1-Score**
  - **Small impact on translation quality**
- **Use F_α-Score as optimization criterion**
- **Future work**
  - **Investigate features operating on sentence level**
  - **Enrich grammar with syntactical information**

Thank you for your attention

Stephan Peitz

peitz@cs.rwth-aachen.de

http://www-i6.informatik.rwth-aachen.de/~peitz

References

[Chiang 05] D. Chiang: A Hierarchical Phrase-Based Model for Statistical Machine Translation. pp. 263–270, Ann Arbor, Michigan, June 2005. 8

[Ma & Tinsley+ 08] Y. Ma, J. Tinsley, H. Hassan, J. Du, A. Way: Exploiting Alignment Techniques in MaTrEx: the DCU Machine Translation System for IWSLT08. In *Proc. of the International Workshop on Spoken Language Translation*, pp. 26–33, Hawaii, USA, 2008. 7

[Och 03] F.J. Och: Minimum Error Rate Training in Statistical Machine Translation. pp. 160–167, Sapporo, Japan, July 2003. 6

[Peitz & Freitag+ 11] S. Peitz, M. Freitag, A. Mauser, H. Ney: Modeling Punctuation Prediction as Machine Translation. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, San Francisco, CA, Dec. 2011. 7

[Stolcke 02] A. Stolcke: SRILM-An extensible language modeling toolkit. In *In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*, pp. 901–904, 2002. 13