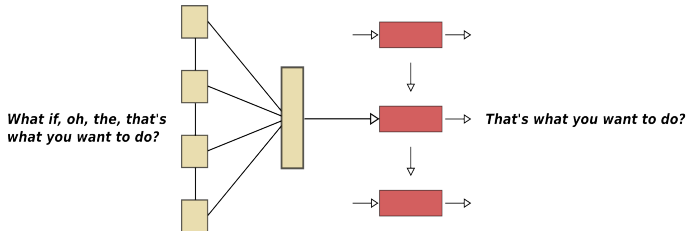


Multilingual Disfluency Removal using NMT

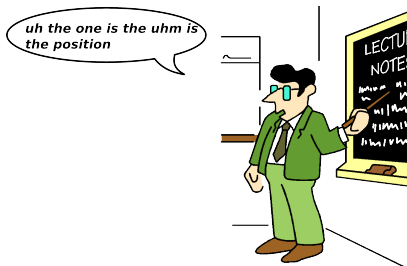
Eunah Cho, Jan Niehues, Thanh-Le Ha, Alex Waibel | December 9, 2016

INSTITUTE FOR ANTHROPOMATICS AND ROBOTICS, KARLSRUHE INSTITUTE OF TECHNOLOGY, KARLSRUHE, GERMANY



- Motivation
- Speech disfluency
 - Impact of disfluency in MT performance
- Previous work
- Model
- Experiments and results
- Conclusion

- Spontaneous speech
- Disfluency: hesitation, stutter, repetition, correction



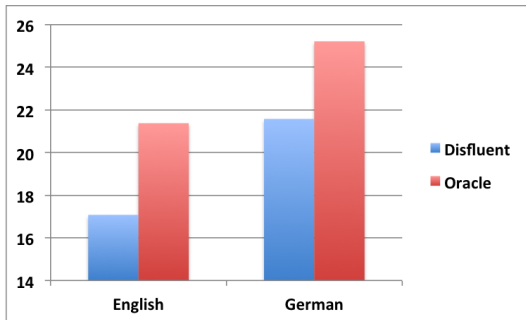
- Disfluency annotation: expensive!
- Limited data
- Are there similarities of disfluency across languages?
- Potential of multilingual disfluency modeling
- Multilingual approach for speech disfluency removal

(reparandum)* <editing term> correction

I'd like to book a flight **to Boston**, sorry, **to Denver**.

- Filler: *uh, uhm*
- Discourse marker: *you know, well*
- Repetition
uhm right, **they don't**, uhm, they don't actually go into ...
- False starts
what if, oh, **the**, that's what you want to do?

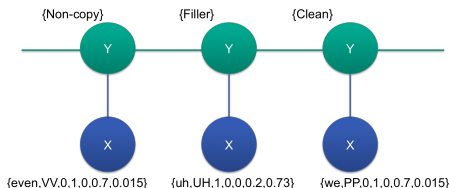
Impact of Disfluency on BLEU



- English meeting data → French
- German lecture data → English
- 3.6 - 4.3 BLEU points of difference when removing disfluencies according to manual annotation

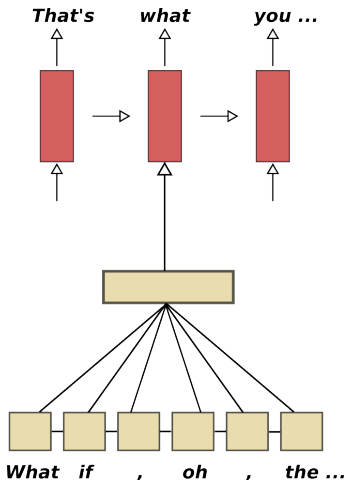
Related work: CRF-based approach

- Given the observed sequence, a hidden label sequence is modeled
 - Label: disfluency classes



- Features
 - Lexical features: word, POS, their patterns, etc
 - Clusters from word vectors
 - Phrase table information

Disfluency removal in NMT framework



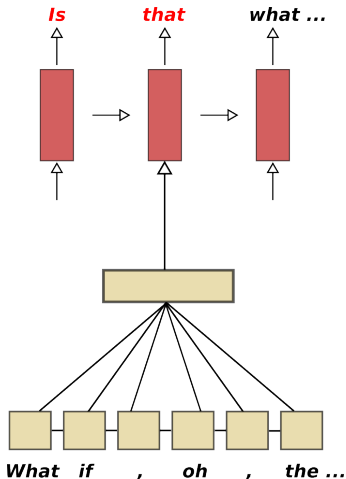
- Disfluency removal as a translation task from disfluent to clean language

what if, oh, the, that's what you want to do?



that's what you want to do?

Disfluency removal in NMT framework



- Possible to extend to further tasks: reconstruction (reordering, replacement of words)

what if, oh, the, that's what you want to do?



Is that what you want to do?

- Colloquial expressions and ungrammatical phrases still remain

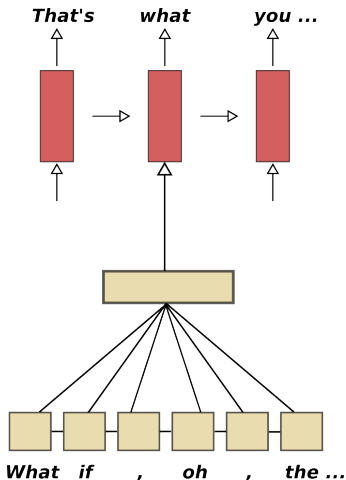


*Surfen sie **mal** ein bisschen '**rum auf den Seiten**, die ich Ihnen gegeben habe, vielleicht fällt Ihnen **was** auf, **was** sie gerne machen wollen.*



*Surfen sie **einmal** ein bisschen **auf den Seiten herum**, die ich Ihnen gegeben habe, vielleicht fällt Ihnen **etwas** auf, **das** sie gerne machen wollen.*

Multilingual disfluency removal



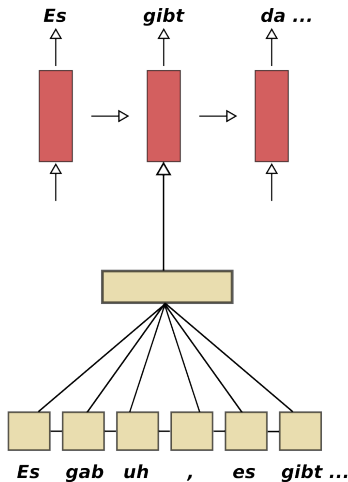
- Source: disfluent language in English and German
- Target: cleaned-up English/German

what if, oh, the, that's what you want to do?



that's what you want to do?

Multilingual disfluency removal



- Source: disfluent language in English and German
- Target: cleaned-up English/German

es gab uh es gibt da drei Prinzipien ...



es gibt da drei Prinzipien ...

- Sub-word operation?

he bro+ ke his arm .

- Rare-word problem of NMT

- Vocabulary sharing, language specification?

*@en_***he** *@en_***bro+** *@en_***ke** *@en_***his** *@en_***arm** *@en_*
*@de_***er** *@de_***ist** *@de_***arm** *@de_*.

- Share parameters
- Disambiguation

- Two single language disfluency removal systems
 - EN-EN
 - DE-DE
- A multilingual disfluency removal systems
 - EN/DE - EN/DE
- Identical architecture
- Output of disfluency removal is translated into another language
 - English→French
 - German→English
 - Evaluate the impact of disfluency removal in a downstream application

- Attention-based encoder-decoder model for NMT `nematus`
 - Source: disfluent transcripts
 - Target: clean transcripts
- Data
 - English: meetings, online lectures
 - German: university lectures
 - Manual annotation on disfluency
 - Almost identical guideline
 - Meeting: *interruption* class
- ~100K tokens of training data for each language
- ~20K, ~30K tokens of test data
- Disfluency rate: 12%

- Evaluation of performance depending on different sub word operation size
- Multilingual system
- BLEU against the human-cleaned transcript

Sub-word operation	No. tokens	Dev
Character	971k	78.97
BPE 150	498k	92.59
BPE 200	465k	92.01
BPE 500	372k	92.37

- 42.4% tokens left unsplit

Sub-word operation

w_hat i_f o_h the th_at_'s w_hat you w_an_t to d_o ?

- Would the language ID be helpful?

en@w_en@**hat** en@i_en@**f** en@o_en@h ...
de@m_de@**it** de@d_de@**em** ...

System	Dev
BPE 150	92.59
+ LangID	91.05
BPE 200	92.01
+ LangID	91.75

- No language specific representation, shared vocabulary

System	English	German
Baseline	74.37	78.03
+ no <i>uh</i>	76.82	84.90
Single language NMT sys.	81.56	89.61
Multilingual NMT sys.	83.57	90.75
CRF-based single language sys.	78.78	-

- Evaluated against human-cleaned test set, in BLEU
- Baseline: all disfluencies kept
- no *uh*: remove *uh* and *uhm*
- Outperforms CRF-based English system

System	English	German
Baseline	17.08	21.58
+ no <i>uh</i>	17.75	23.46
Single language NMT sys.	19.36	24.34
Multilingual NMT sys.	19.59	24.43
CRF-based single language sys.	18.22	-
Oracle	21.38	25.22

- Translate the disfluency-cleaned test data
 - English→French
 - German→English
- 1-1.8 BLEU points of improvement over no *uh* baseline
- Oracle: human-cleaned test data

Example output of disfluency removal

Input	and as we find more groups, we record more groups.
Ref	and as we find more groups, we record more groups.
CRF	as we record more groups.
MultiNMT	and as we find more groups, we record more groups.

Input	yeah but this is really I we we're not uh record a really tight subset of meetings.
Ref	we're not record a really tight subset of meetings.
CRF	we're not record a really tight subset of meetings.
MultiNMT	but this is really tight subset of meetings.

Example output of disfluency removal

- Test on ASR output
- Punctuation inserted using monolingual MT
- German and English mixed input

Input: ja, dann ja, und kann dann auch die Fragen stellen oh, okay, that is cool

MultiNMT: dann auch die Fragen stellen okay, that is cool

Input: oh, okay, okay, I mean, they got like like like federal und wir Informationen heraus suchen

MultiNMT: okay, I mean, they got like federal und wir Informationen heraus suchen

- Model disfluency removal in an NMT framework
- Motivated by data sparsity for disfluency removal
- Multilingual learning for English and German spontaneous data
- Outperformed CRF-based model and single language based models
- Extend to further tasks (reconstruction, reordering, etc)

Thank you!