Language Model Adaptation for ASR of Spoken Translations Using Phrase-based Translation Models and Named Entity Models

Joris Pelemans[1]
Tom Vanallemeersch[2]
Kris Demuynck[3]
Lyan Verwimp[1]
Hugo Van hamme[1]
Patrick Wambacq[1]

1 KU Leuven – ESAT, Belgium
2 KU Leuven – CCL, Belgium
3 Ghent University – ELIS, Belgium

Context: The SCATE Project

= Smart Computer-Aided Translation Environment

Research topics:

1) Translation technology
2) Evaluation of computer-aided translation
3) Terminology extraction from comparable corpora
4) Speech recognition
5) Work flows and personalized user interfaces

Speech Recognition in CAT

MT Pair

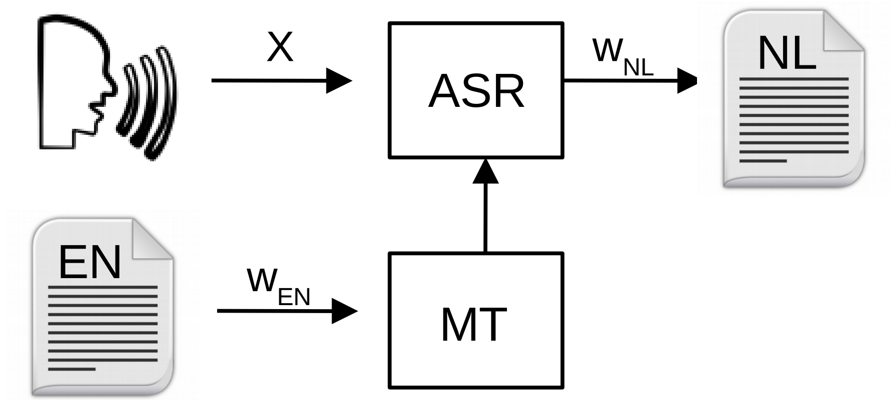
EN: Simulations have shown that in heavy but free flowing traffic, jams can arise spontaneously …

NL: Simulaties hebben aangetoond dat in zwaar maar vlot verkeer, jam kan spontaan ontstaan …

ASR correction: files spontaan kunnen ontstaan

Goal: improve ASR accuracy, hence translator efficiency
Subgoal: speed, limit adaptation overhead

MT-based LM Adaptation



- Use translation model to adapt LM:

$$P(w_{NL}|w_{EN},X) = P(X,w_{EN}|w_{NL})\,P(w_{NL}) / P(w_{EN},X)$$

$$\approx P(X|w_{NL})\,P(w_{EN}|w_{NL})\,P(w_{NL}) / P(w_{EN},X)$$

$\Rightarrow P'(w_{NL}) = P(w_{EN}|w_{NL})\,P(w_{NL})$ is new language model

- Advantages over multi-pass approach:
  - No intermediate storage
  - Maximal information during recognition

Previous Work: focus on speed

- Efficient implementation
  - Update only relevant n-grams using inflation weights:

    $$P'(w_{NL}) = P(w_{NL})\ g(P(w_{EN}|w_{NL}))\ \text{with}\ g(x) = 1 + \alpha\beta^{(1-x)}$$

  - No renormalization (not necessary for ASR)
  - Store update weights instead of full model
  - On-the-fly adaptation
- Lexical translation model

  = one-to-one translations e.g. $(\text{file})_{NL} \rightarrow (\text{jam})_{EN}$
- More info: Pelemans et al., Efficient Language Model Adaptation for ASR of Spoken Translations. In Proc. Interspeech 2015.

Now: focus on accuracy

- Phrase-based TM $P(w_{EN}|w_{NL})$ instead of lexical TM

  = m-to-n translations e.g.

  - $(\text{moeten})_{NL} \rightarrow (\text{have to})_{EN}$
  - $(\text{hou van})_{NL} \rightarrow (\text{love})_{EN}$
  - $(\text{kijkt naar})_{NL} \rightarrow (\text{looks at})_{EN}$
- Named entity model

Phrase-based LM Adaptation

- Phrase-based TM calculates 4 scores:
  - Phrase translation probabilities (relative frequencies):
    - ϕ(EN|NL)
    - ϕ(NL|EN)
  - Lexical weights (average lexical probability):
    - π(EN|NL)
    - π(NL|EN)
- Interpolate scores linearly to adapt LM:

$$P(w_{EN}|w_{NL}) = \lambda_1\,\phi(EN|NL) + \lambda_2\,\phi(NL|EN) + \lambda_3\,\pi(EN|NL) + \lambda_4\,\pi(NL|EN)$$

Named Entity Models

- Problem: word not in TM

  => no $P(w_{EN}|w_{NL})$ => no LM update
- Solution for named entities (NE):
  - In ASR vocabulary and LM (IV)
    - Estimate $P(w_{EN}|w_{NL})$
    - Assume named entities are untranslated e.g. $(\text{Shanghai})_{NL} \rightarrow (\text{Shanghai})_{EN}$

      => $P(w_{EN}|w_{NL}) = \alpha \approx 1$

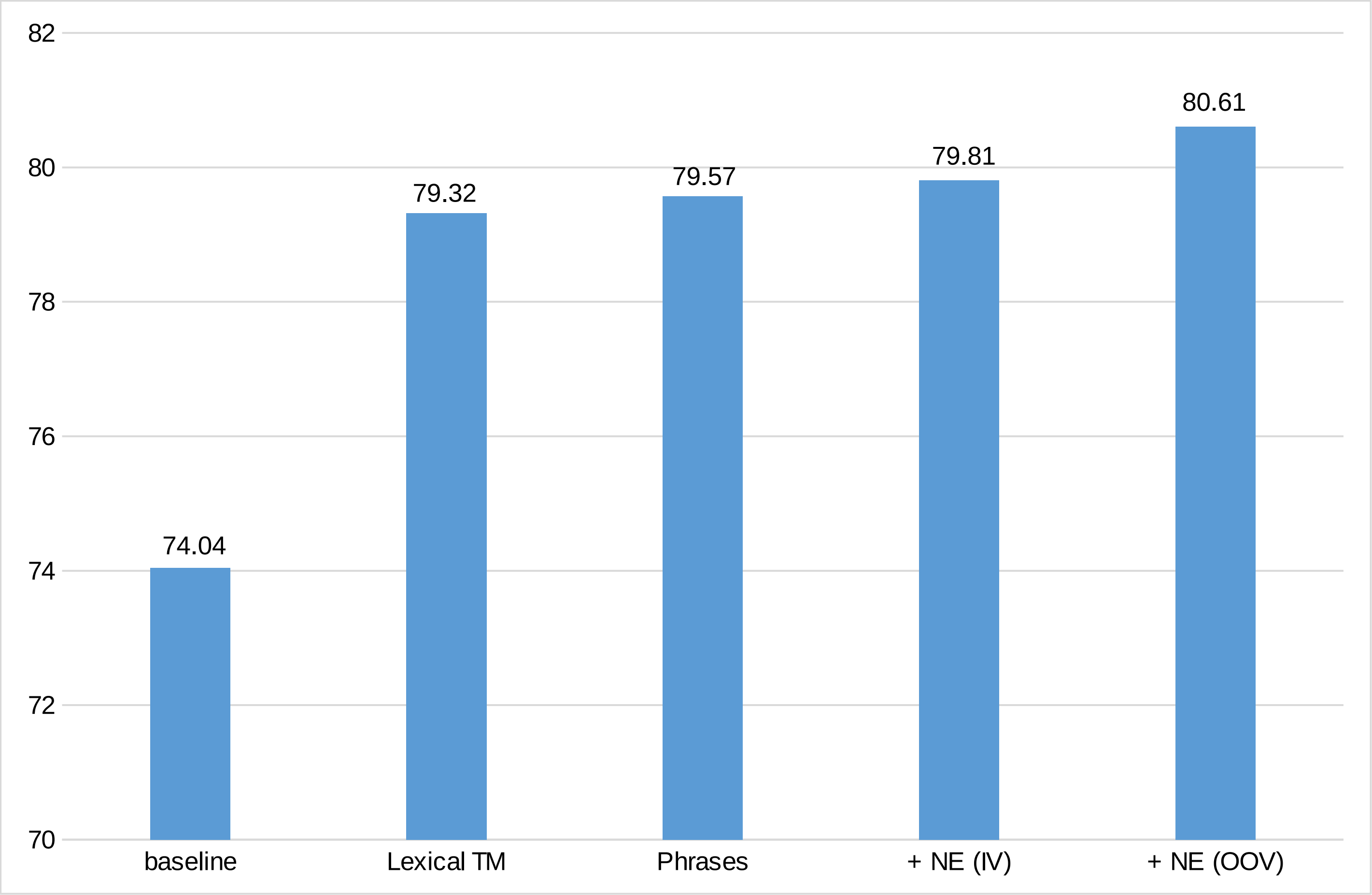
      => $P'(w_{NL}) = P(w_{NL})\, g(\alpha)$
  - Not in ASR vocabulary and LM (OOV)
    - Add to pronunciation lexicon, using g2p
    - Estimate $P'(w_{NL})$ directly, based on OOV statistics

      => $P'(w_{NL}) = h_{NE}\, P(OOV_{NL})$

Experiments

- No post-editing, but ASR on translated English literature from Corpus Spoken Dutch (CGN), component “o”
- Nbest recognizer with:
  - 100k words
  - 3-gram LM (mod KN)
- TM created by GIZA++ on 1M EN-NL sentence pairs from European Parliament
- Timings averaged over 100 executions on single core Intel i5-2400 processor
- Demo available on our YouTube page

Results: accuracy



Results: storage & speed

- Efficient on-the-fly adaptation, independent of n-gram order:
  - Disk storage of <250KB per sentence (vs 0.5 GB with existing implementation)
  - Virtually no overhead: 0.21s per sentence (vs 4m33s)

Conclusions and Future Work

- Extending MT-based LM adaptation with phrase-based TMs and named entity models yields:
  - Increase of 6.2% recognition accuracy
  - No noticeable overhead

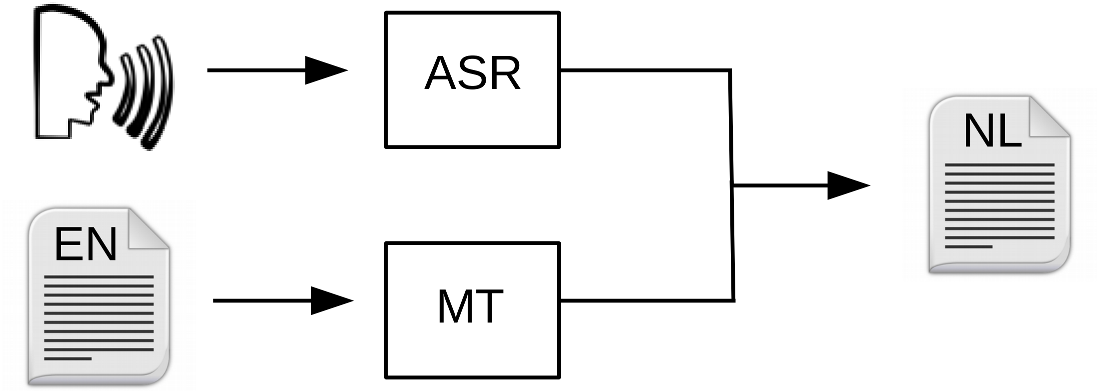
    => can be readily used in CAT software
- Further improvements are expected when:
  - MT model domain matches task
  - Many-to-many translations are used

$(\text{Dank u})_{NL}$
$=$
$(\text{Thank you})_{EN}$

More information:

- joris.pelemans@esat.kuleuven.be
- Pelemans et al., Efficient Language Model Adaptation for ASR of Spoken Translations. In Proc. Interspeech 2015.
- Demo: http://www.esat.kuleuven.be/psi/spraak/demo/
- Twitter: #SpeechAtKULeuven

EXTRA: Multi-pass Approach



- Use translation model to rescore ASR output:
  - N-best list [Brousseaux et al., 1995] [Khadivi et al., 2005] [Paulik et al., 2005]
  - Lattice/confusion network [Khadivi and Ney, 2008] [Reddy and Rose, 2010]
- Disadvantages:
  - Valuable hypotheses might already be lost in ASR output
  - Time-consuming:
    - Multi-pass
    - Storage of intermediate results