



Smart Computer-Aided Translation Environment: highlights

Vincent Vandeghinste, Tom Vanallemeersch, Bram Bulté, Liesbeth Augustinus, Frank Van Eynde, Joris Pelemans, Lyan Verwimp, Patrick Wambacq, Geert Heyman, Marie-Francine Moens, Iulianna van der Lek-Ciudin, Frieda Steurs, Ayla Rigouts Terryn, Els Lefever, Arda Tezcan, Lieve Macken, Véronique Hoste, Sven Coppers, Jens Brulmans, Jan Van den Bergh, Kris Luyten, Karin Coninx

Intelligible Translator Interfaces

Demo: <http://scate.edm.uhasselt.be/>

(A) source text segment (B) the edit box for target text segment (C) suggested hybrid machine translation (D) translation alternatives (E) fuzzy matches from the translation memory (F) autocompletion activated as soon as a new segment is opened; combines results of machine translation, translation memory and translation alternatives

The screenshot shows a multi-paneled interface. At the top, source text segments (G, H) are listed. Below, a target text segment (B) is shown in an edit box. A suggested hybrid machine translation (C) is displayed below the edit box. A list of translation alternatives (D) is shown on the left. Fuzzy matches from the translation memory (E) are shown on the right. The interface includes various icons for editing and navigation.

Improved fuzzy matching

Adaptations of non-linguistic metrics

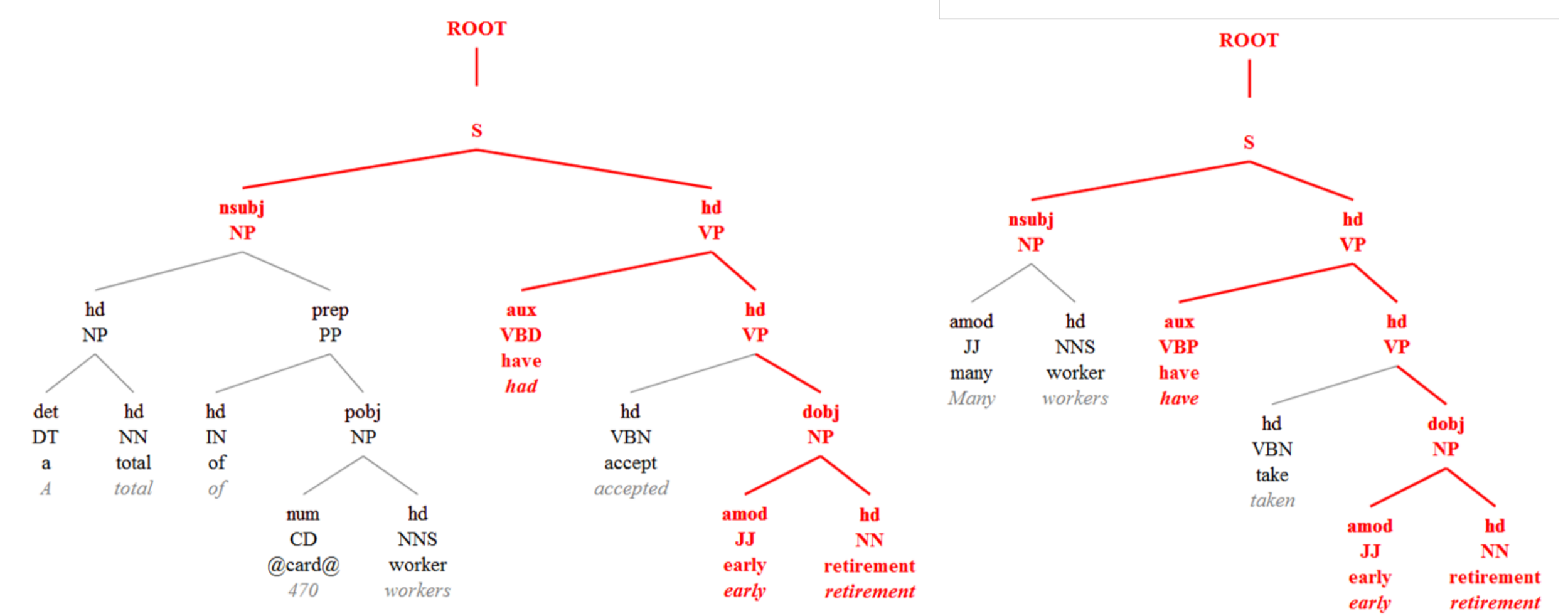
Levenshtein, percent match & TER on sequences of lemmas

Linguistic metrics

Meteor

Shared partial subtree matching

Ngram precision for head word chains



Integration TM and MT

Components

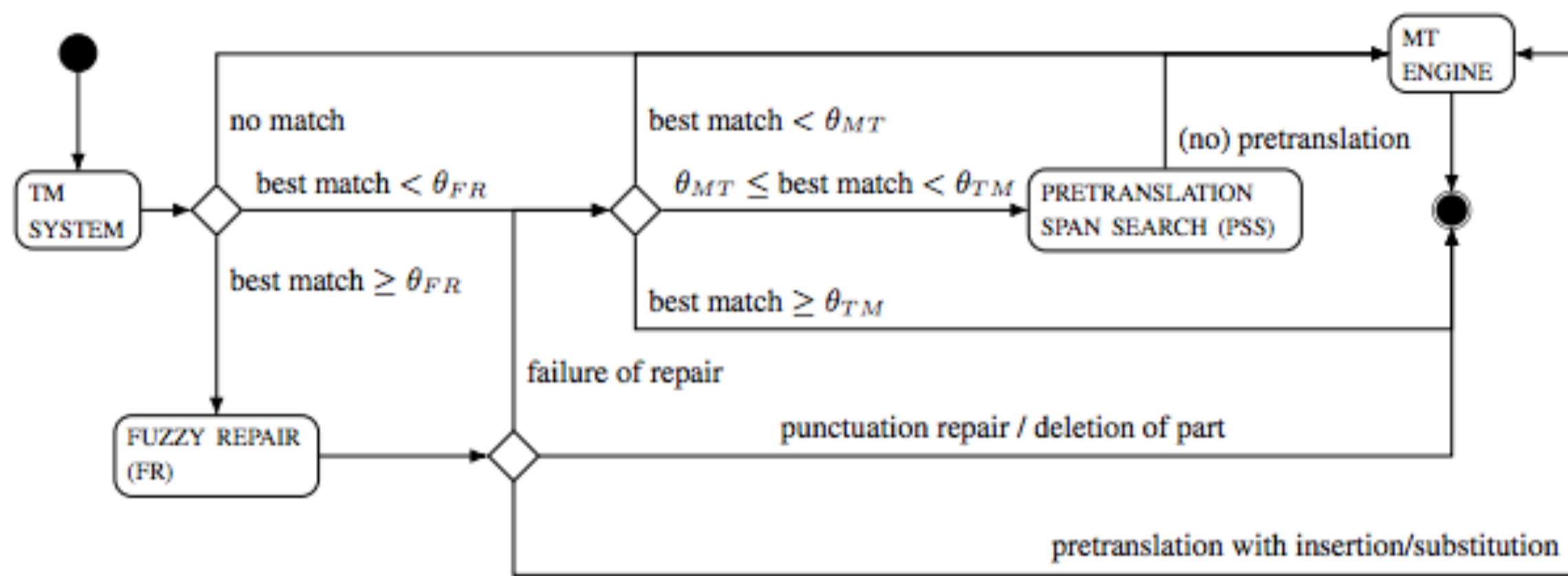
- TM system
- PBSMT engine (Moses)
- Fuzzy repair system
- Pretranslation span search

Different translation routes

Based on thresholds for (fuzzy) match score
Success/failure of certain operations

Novelty

Combination of different components
Tunable thresholds / parameters / criteria



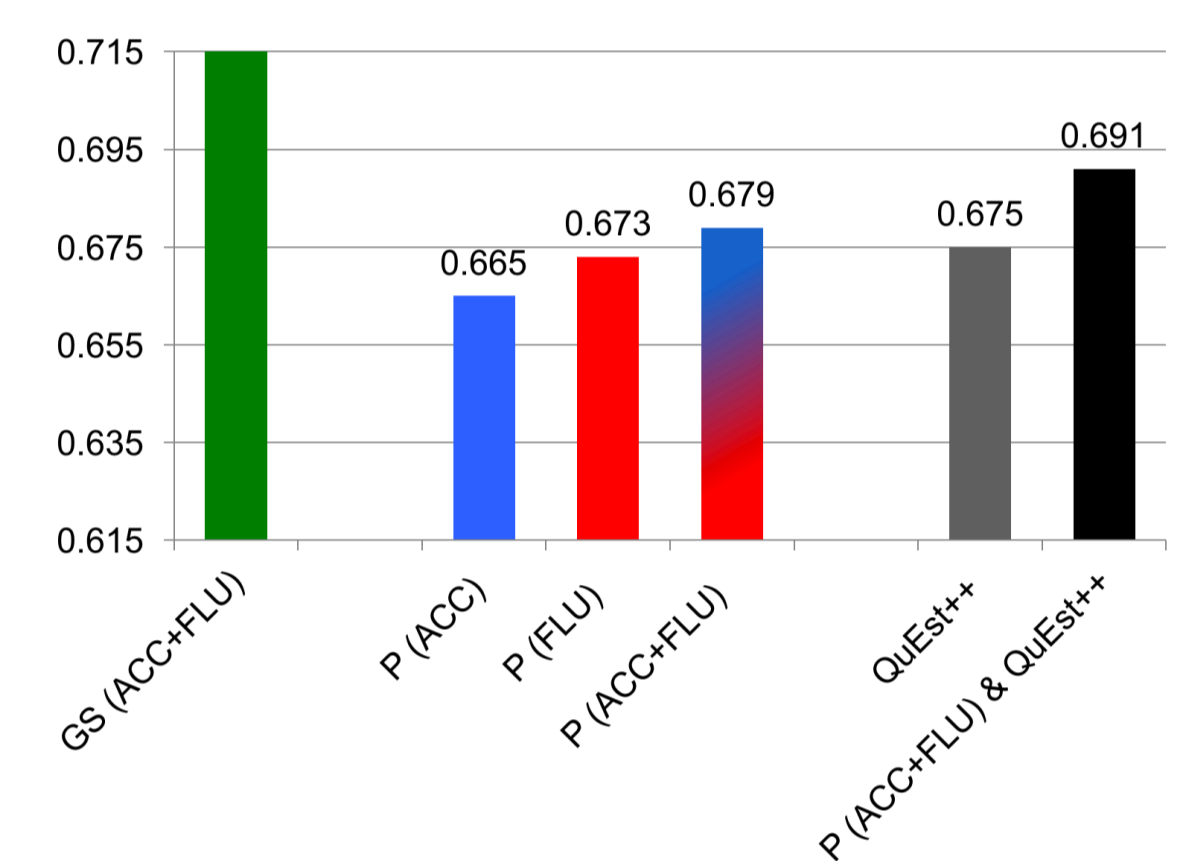
Informed quality estimation

1. Detect different types of MT errors at word level

➤ Fluency / Accuracy

2. Use word-level predictions to estimate quality at sentence-level

➤ Post-editing time



Editing Details

2. Estimate Post-Editing Time

Segment ID	Words	Time-to-edit (TTE)	Secs/Word	Post-editing effort (PEE)
659878671	7	00m:16s	2.3s	37%
Segment	What are the real chances of success?			
Suggestion (MT)	Wat zijn de realistische kans op succes?			
Translation	What are the real chances of success?			
DIFF View	Wat zijn de realistische kans echte kansen op succes?			
QA Issues				

1. Detect different types of MT errors

Post-Editing via Automated Speech Recognition

Use translation model to improve the accuracy of the speech recognizer

Phrase-based translation models instead of word-based translation models

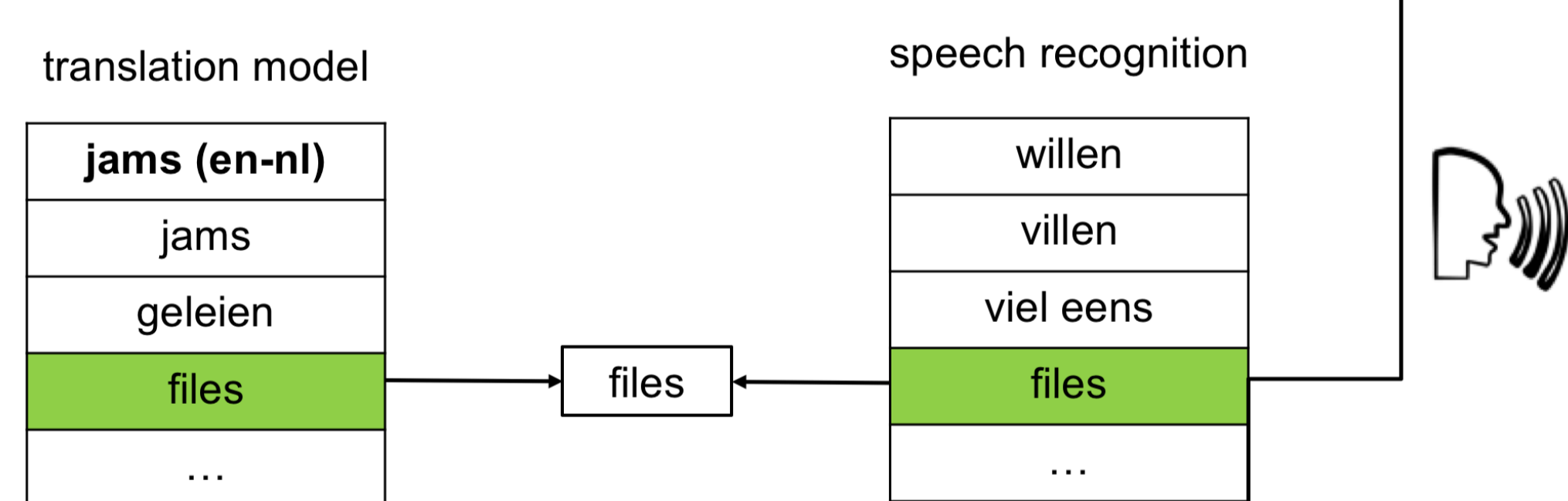
➤ grey horse - schimmel, screen resolution - schermresolutie

Named entities can often be 'translated' as is

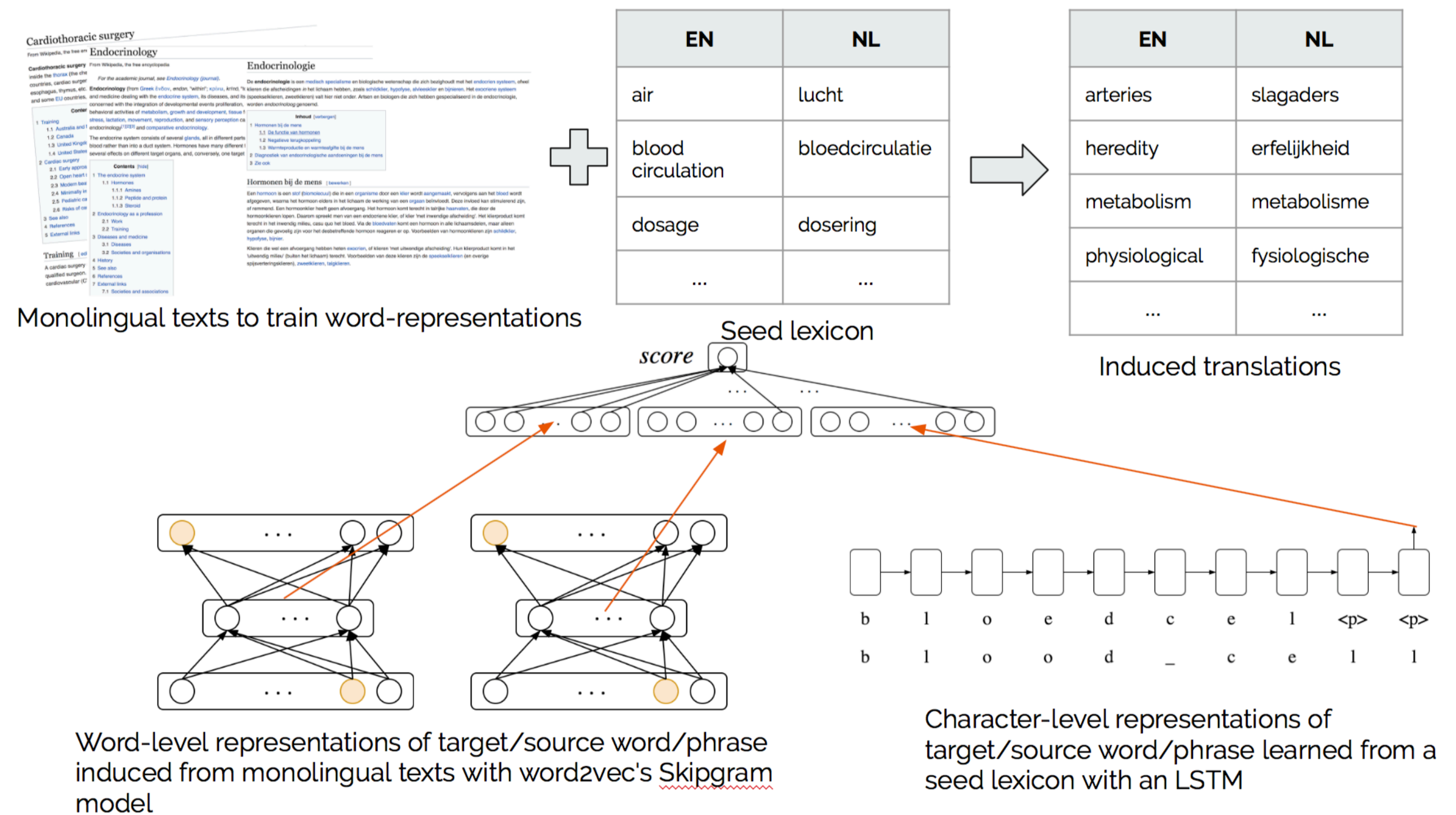
➤ add the untranslated named entities to the ASR model

In heavy but free flowing traffic, jams can arise spontaneously, triggered by an abrupt steering maneuver by a single motorist.

In zwaar maar vlot verkeer kunnen jams spontaan ontstaan, veroorzaakt door een abrupt stuurmanoeuvre door een automobilist.



Identifying bilingual terms in comparable texts



References:

Coppers, S., Van den Berg, J., Luyten, K., van de Lek-Ciudin, J., Vanallemeersch T. & V. Vandeghinste. 2018. Intellingo: An Intelligible Translation Environment, Proceedings of CHI 2018

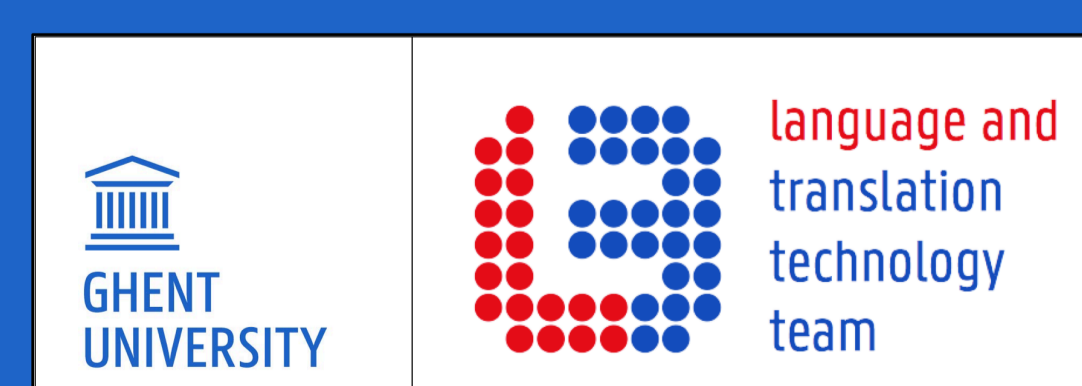
Vanallemeersch, T. & Vandeghinste, V. (2015). Assessing linguistically aware fuzzy matching in translation memories. Proceedings of EAMT 2015

Bulté, B., Vanallemeersch, T. & Vandeghinste, V. (2018). M3TRA: integrating TM and MT for professional translators. Proceedings of EAMT 2018

Pelemans J., Vanallemeersch T., Demuyck K., Verwimp L., Van hamme H., Wambacq P. 2016. Language model adaptation for ASR of spoken translations using phrase-based translation models and named entity models. In Proceedings ICASSP 2016, pp. 5985-5989.

Tezcan, A. (2018). Informative quality estimation of machine translation output (Doctoral dissertation)

Heyman G., Vulić I., Moens M. (2017). Bilingual Lexicon Induction by Learning to Combine Word-Level and Character-Level Representations. Proceedings of EACL 2017



Industrial Advisory Committee: Aktor Knowledge Technology, CommArt International, CrossLang, ITP Europe, Mastervoice, Nuance, OneLiner, VRT Onderzoek en Innovatie, Xplanation, Yamagata-Europe, Yazzoom, Televic