

## Document-Level Decoding and Discourse in Statistical Machine Translation

Christian Hardmeier

Uppsala universitet

2016-05-16

## Sentence-by-sentence Translation

Just press one key and in just a few seconds you can switch from Windows Mobile to Android.

**This** is the goal of the American company Vmware, which primarily develops computer virtualisation software.

**This** will let you have two user profiles at once on the same phone.

You can switch between **them** or have **one** for work and **one** for home. Both of **them** will run at the same time, says Srinivas Krishnamurti of VMware in an interview with Computer World magazine.

**It** was presented last November and first demonstrated just a few days ago.

**It** will go on sale in 2012.

## Pronoun translation

*Translating pronouns requires target-language dependencies.*

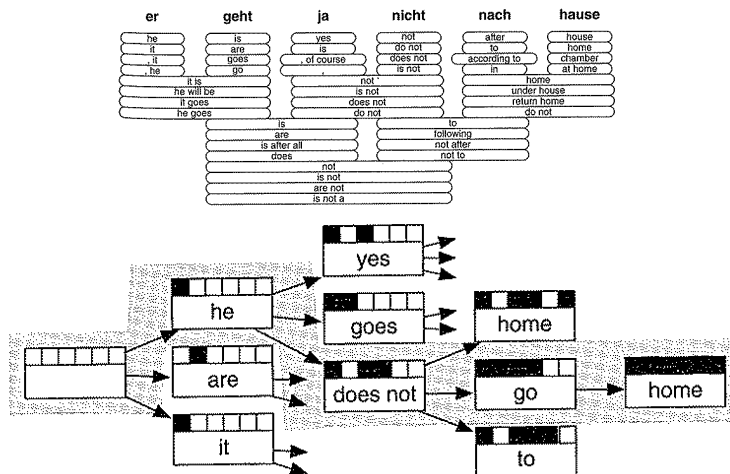
The **funeral** of the Queen Mother will take place on Friday.

**It** will be broadcast live.

Les **funérailles** de la reine-mère auront lieu vendredi.

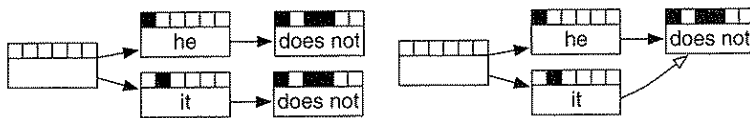
**Elles** seront retransmises en direct.

## SMT decoding: Decoding process



## SMT decoding: Hypothesis recombination

- The search space can be reduced enormously by exploiting the locality of the language model.
- With a trigram model, if two hypotheses match in the last two words, only the better one is retained.
- *Hypothesis recombination* is a form of dynamic programming.
- Search optimality is maintained by recombination.
- Ultimately, we run lossy beam search in the reduced search space.



## Dynamic Programming beam search decoding

### Advantages:

- Very good search results in a huge search space.
- Manageable complexity  
(linear in sentence length with pruning and limited reordering)
- Best efficiency/exactness trade-off with current models.

### Disadvantages:

- Search algorithm incorporates Markov assumption over words.
- Sentence-internal long-distance dependencies increase the search space greatly by inhibiting recombination.
- No obvious way to handle cross-sentence dependencies.

- The standard search algorithm constructs output sentences bit by bit.
- Local search algorithms start with a complete (suboptimal) solution and make small changes to improve it.
- This can be applied to entire documents.
- The initial state can be generated randomly or with beam search.

that prevention this disease

unfortunately , there are is not a miracle cure contribute to preventing get cancer but

in spite to the made progress in ' scientific research there remains the a healthy ways of life the best solution , the if of the risks decrease , in with him disease this .

that prevention of the disease

unfortunately , there are no miracle cure for preventing cancer .

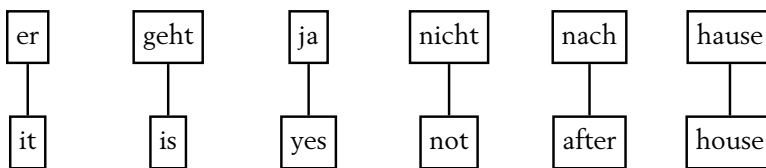
in spite of the progress in research remains the adoption of a healthy lifestyle is the best way , about the risk to reduce , in him on developing them .

prevention of the disease

unfortunately , there is no miracle cure for preventing cancer .

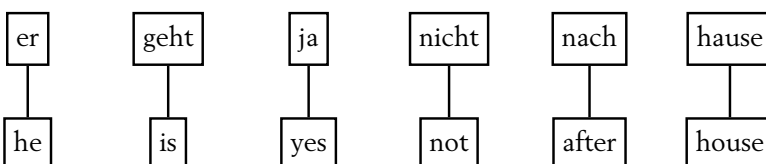
despite the progress in research , the adoption of a healthy lifestyle  
, the best way to minimise the risk of him on ill .

## Local search operations for phrase-based SMT



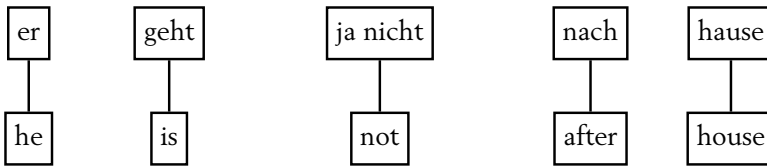
- **Change phrase translation:** Change the translation of one phrase.
- **Resegment:** Find a new phrase segmentation for a sequence of adjacent source words.
- **Swap phrases:** Exchange the positions of two target phrases.

## Local search operations for phrase-based SMT



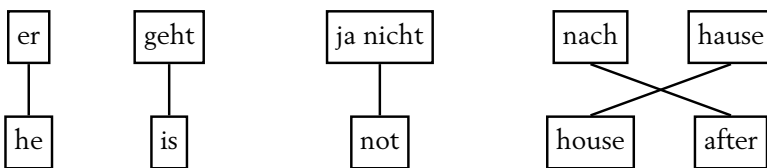
- **Change phrase translation:** Change the translation of one phrase.
- **Resegment:** Find a new phrase segmentation for a sequence of adjacent source words.
- **Swap phrases:** Exchange the positions of two target phrases.

## Local search operations for phrase-based SMT



- **Change phrase translation:** Change the translation of one phrase.
- **Resegment:** Find a new phrase segmentation for a sequence of adjacent source words.
- **Swap phrases:** Exchange the positions of two target phrases.

## Local search operations for phrase-based SMT

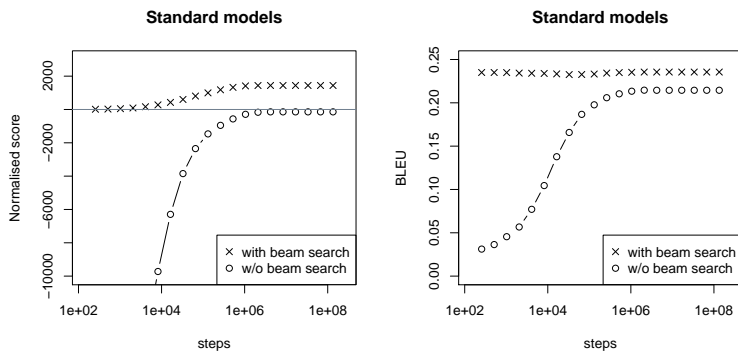


- **Change phrase translation:** Change the translation of one phrase.
- **Resegment:** Find a new phrase segmentation for a sequence of adjacent source words.
- **Swap phrases:** Exchange the positions of two target phrases.

## First-choice hill climbing

- For a given state, generate a successor state by applying a randomly chosen operation at a random location in the document.
- Compute the score of the new state.
- Accept the new state if it is better, else keep the current state.
- Iterate until
  - a maximum number of steps is reached, or
  - there is a large number of rejections in a row.

## English-French Newswire (WMT system)



- Local search is an example of an alternative search algorithm for phrase-based SMT.
- Unlike stack decoding, it can accommodate models with long-range dependencies.
- On the downside, it doesn't benefit from dynamic programming.

- Pronominal anaphora (Hardmeier, 2014)
- Improving output readability (Stymne et al., 2013)
- Repeated compound words
- Text cohesion and coherence coherence

## Why isn't everyone using it?

- Empirically, we've seen the decoder work well, but most problems with long-range dependencies are poorly understood.
- None of the models implemented so far has brought consistent improvements in translation quality.

## Some examples of pronoun translation

It doesn't create the distortion of reality;  
it creates the dissolution of reality.

Elle ne provoque pas une déformation de la réalité;  
mais plutôt une dissolution de la réalité.

## Some examples of pronoun translation

But the thing about tryptamines is they cannot be taken orally because they're denatured by an enzyme found naturally in the human gut.

Par contre les tryptamines ne peuvent pas être consommés par voie orale étant dénaturés par une enzyme se trouvant de façon naturelle dans l'intestin de l'homme.



## Some examples of pronoun translation

Initially, all we did was autograph **it**.

Pour commencer, nous avons juste **mis notre autographe**.



## Some examples of pronoun translation

Most of **them** are ordinary digital camera photos.

La **plupart** sont des photos d'appareils numériques ordinaires.