INF5820/INF9820

LANGUAGE TECHNOLOGICAL APPLICATIONS

Jan Tore Lønning, Lecture 9, 19 Oct. 2016
jtl@ifi.uio.no

Today

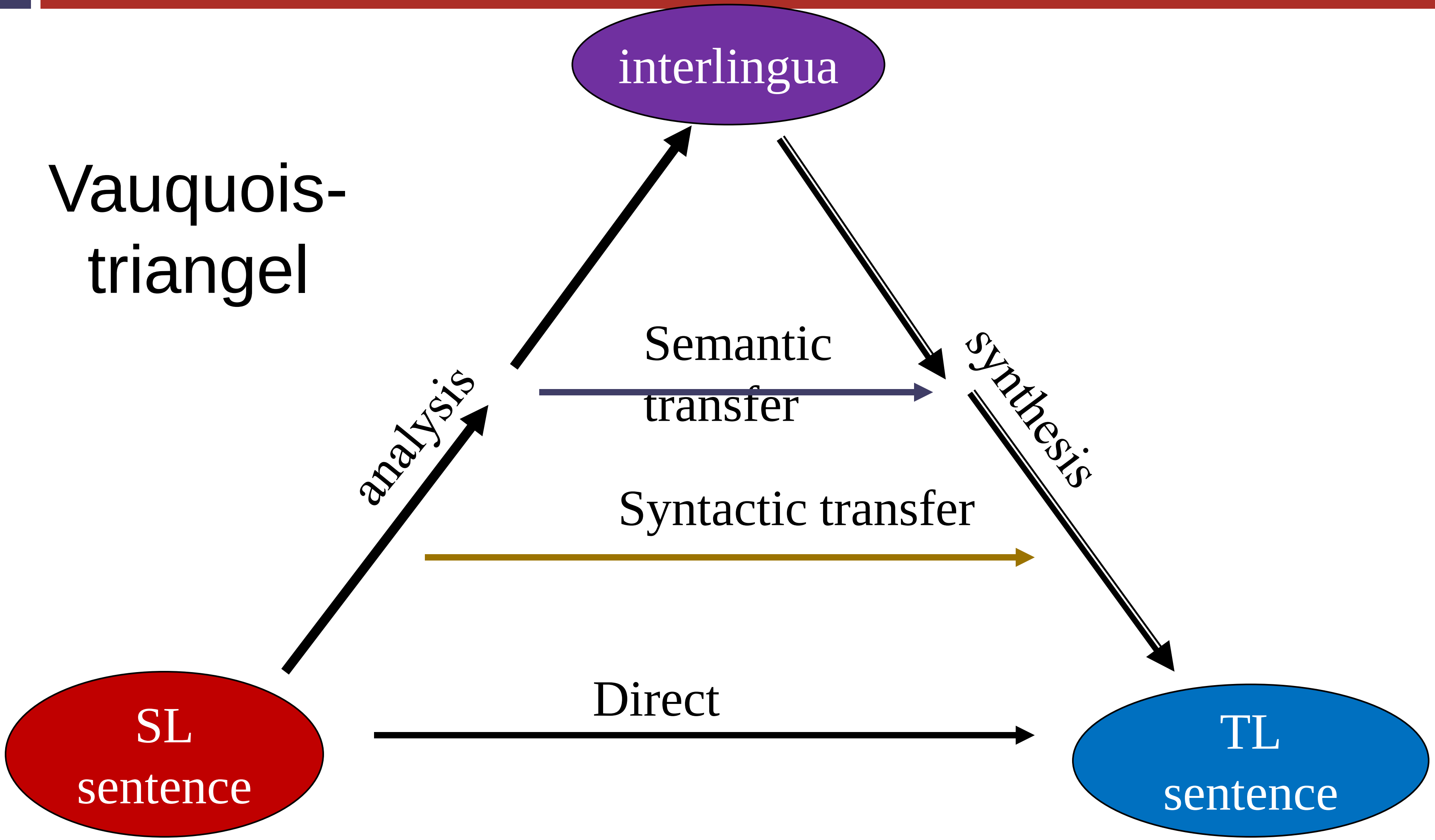
- Hybrid translation:
  - Linguistic rule-based
  - + probability ranking
- Linguistic information in STATMT
  - Morphology
  - Word/order - syntax
- State of the art: alternatives
  - Tree-based translation
  - Neural networks

The LOGON project

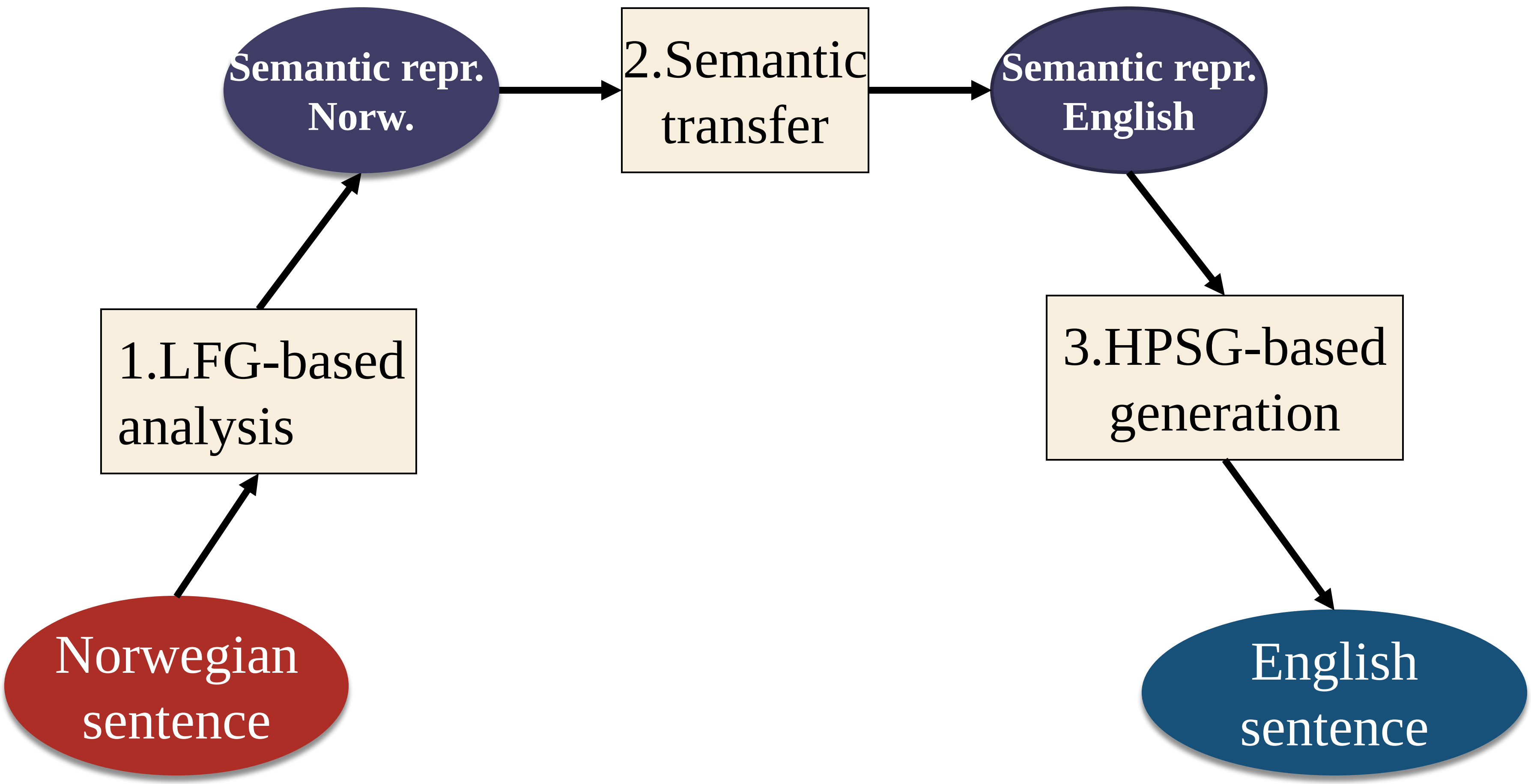
- MT: Norwegian → English
- Tourist texts – hiking descriptions
- High quality – limited recall
- 2003-2007
- Strategy
  - Mainly rule-based:
    - Semantic transfer
  - Probability ranking

Alternative strategies

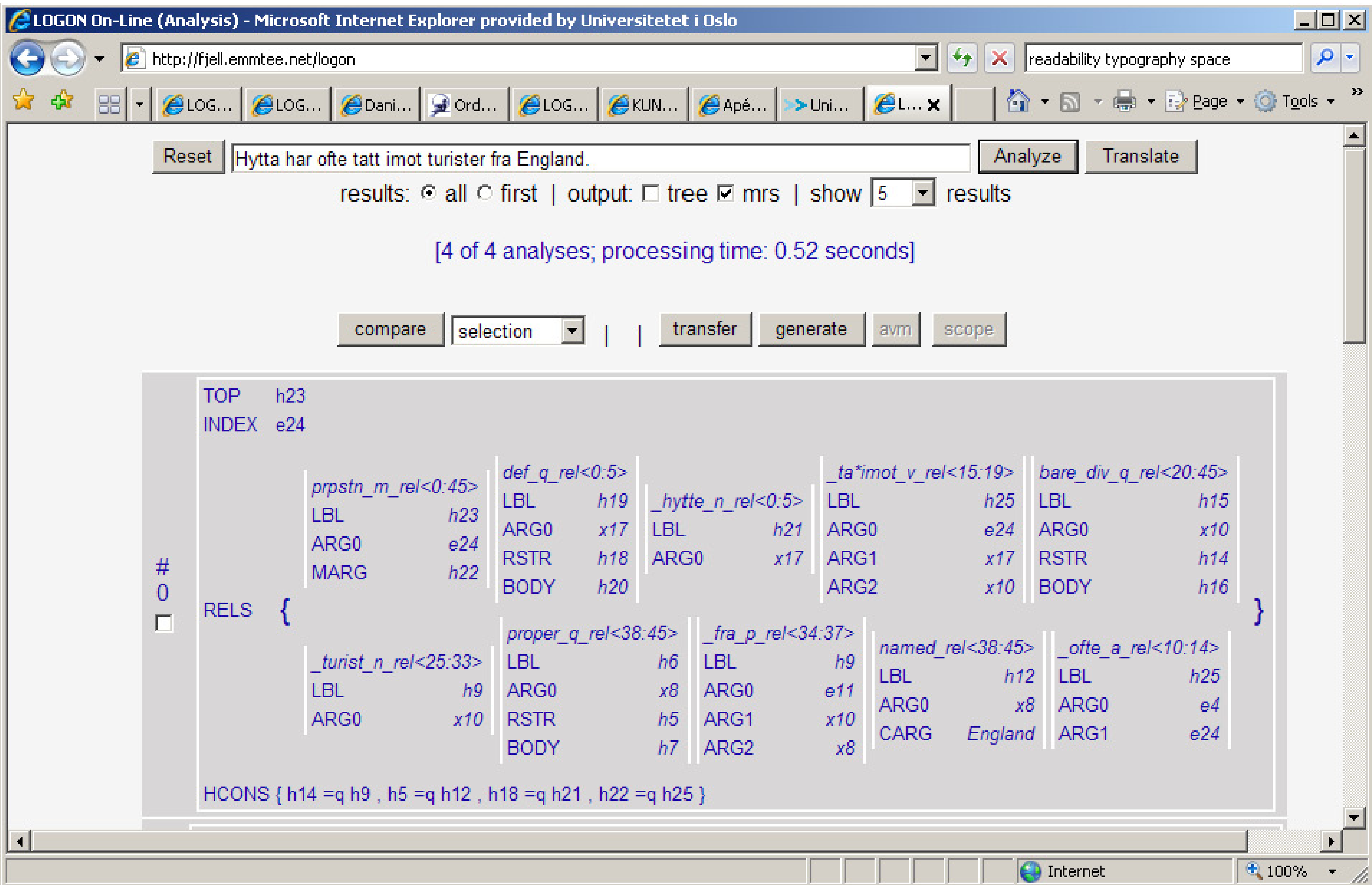
Vauquois-
triangel



Back bone: Semantic transfer



Minimal Recursion Semantics



Analysis of Norwegian

- Grammar: NorGram,
  - A multipurpose computational grammar based on LFG
    - Developed at UiB since 1998
  - LOGON
    - extended grammatical coverage
    - equipped it with an MRS semantics module
  - Currently developed further in the INESS-prosject
    - http://clarino.uib.no/iness/xle-web

- Processing
  - The XLE system from PARC
  - Morphological processing developed at UiB on top of earlier projects (tagging, UiB & UiO & NTNU)
  - Compositional analysis of compounds

Generation

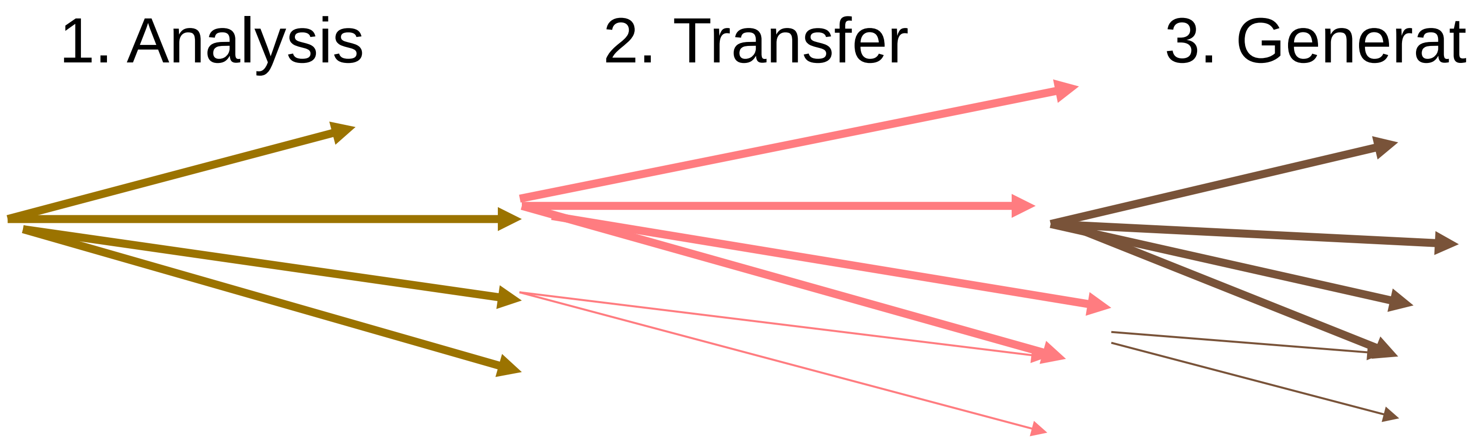
- Grammar
  - The English Resource Grammar (ERG)
  - A multipurpose computational grammar based on HPSG
  - Continuously developed since 1994 (CSLI Stanford)
  - Refined, domain-adapted, and extended by LOGON
  - Open source, used in other ongoing projects
- Processing
  - Adapted technology from DELPH-IN consortium
  - LOGON: forty times faster generation algorithms

Transfer

- Grammar
  - Hand-coded transfer rules (7000 rules)
  - Semi-automatic acquisition of transfer correspondences
    - for open class words
    - from a dictionary (Kunnskapsforlagets store No-En)
    - (ca 10 000)
- Processing
  - Typed unification-based formalism for rewriting of MRSs
  - Design and implementation from scratch
  - Non-deterministic rewriting of MRS-fragments

Today

- Hybrid translation:
  - Linguistic rule-based
  - + probability ranking
- Linguistic information in STATMT
  - Morphology
  - Word/order - syntax
- State of the art: alternatives
  - Tree-based translation
  - Neural networks



- Challenge: Each step generates many different hypotheses
- Approach:
  - Stochastic models score the alternative outcomes of each component: Parsing, Transfer, Generation
  - The per-component scores are calculated together and the final outcomes are ranked.
  - Component models are trained on corpora and treebanks.

- |< |Toppen er luftig, og har en utrolig utsikt!| (83) --- 2 x 24 x 12 = 12
- |> |the top is airy and has an incredible view| [85.9] <0.70> (1:0:0).
- |> |the summit is airy and has an incredible view| [87.4] <1.00> (1:4:0).
- |> |the top is breezy and has an incredible view| [87.7] <0.46> (1:6:0).
- |> |the top is airy and has an unbelievable view| [88.9] <0.70> (1:1:0).
- |> |the peak is airy and has an incredible view| [89.1] <0.96> (1:2:0).
- |> |the summit is breezy and has an incredible view| [89.1] <0.66> (1:10:0).
- |> |the summit is airy and has an unbelievable view| [90.3] <1.00> (1:5:0).
- |> |the top is breezy and has an unbelievable view| [90.7] <0.46> (1:7:0).
- |> |the peak is breezy and has an incredible view| [90.8] <0.66> (1:8:0).
- |> |the peak is airy and has an unbelievable view| [92.0] <0.96> (1:3:0).
- |> |the summit is breezy and has an unbelievable view| [92.1] <0.66> (1:11:0).
- |> |the peak is breezy and has an unbelievable view| [93.8] <0.66> (1:9:0).
- |= 64:19 of 83 {77.1+22.9}; 58:9 of 64:19 {90.6 47.4}; 55:9 of 58:9 {94.8 100.0} @ 64 of 83 {77.1} <0.51 0.67>.

Parse ranking

- First build a parse bank
  - Demo on http://erg.delph-in.net/logon
- Then use this for building a discriminator to select/rank between candidates
- Choices:
  - Features
  - Learning algorithm

Generation ranker

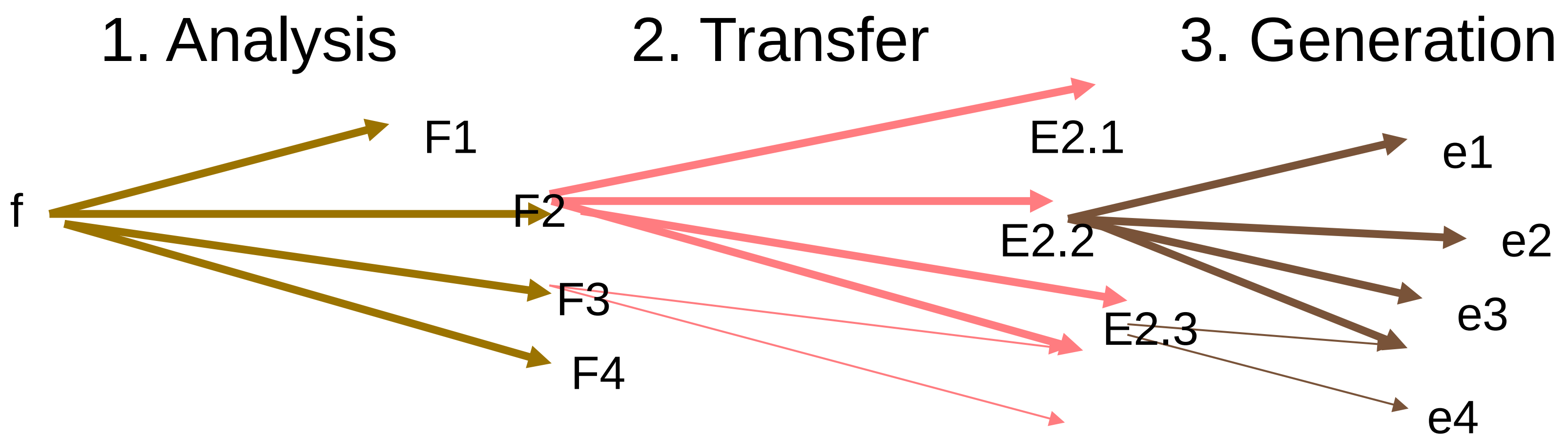
- Roughly 30 realizations per MRS
- First attempt:
  - N-gram language model
- Better:
  - Inspired by parse ranking
  - Developed on the basis of a parse bank
  - Extract features
  - Max-ent learning
  - Better results!

model	exact match	five-best	WA
BNC LM	53.24	78.81	0.882
Log-Linear	72.28	84.59	0.927

Transfer

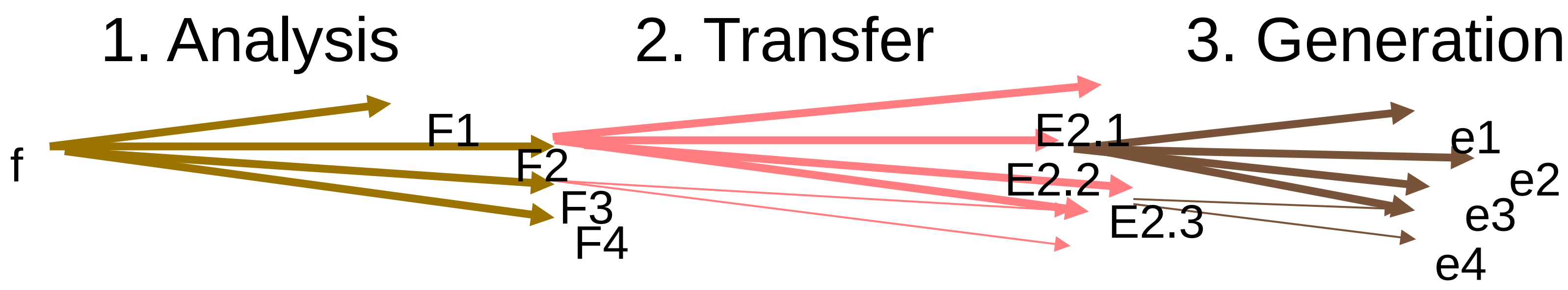
- Should have been **conditional** probabilities:
  - The probability of an English MRS given a Norwegian MRS:
- Only included absolute probabilities:
  - The probability of an English MRS

Putting the 3 together



- Alternatives
  1. First $\arg\max_i P(F_i \mid f)$, say F_2, then $\arg\max_j P(E_j \mid F_2)$ etc
  2. The most likely path $\arg\max_{i,j,k} P(e_k \mid E_j) P(E_j \mid F_i) P(F_i \mid f)$
  3. The most likely translation $\arg\max_e \sum_{F_i} \sum_{E_j} P(e_k \mid E_j) P(E_j \mid F_i) P(F_i \mid f)$

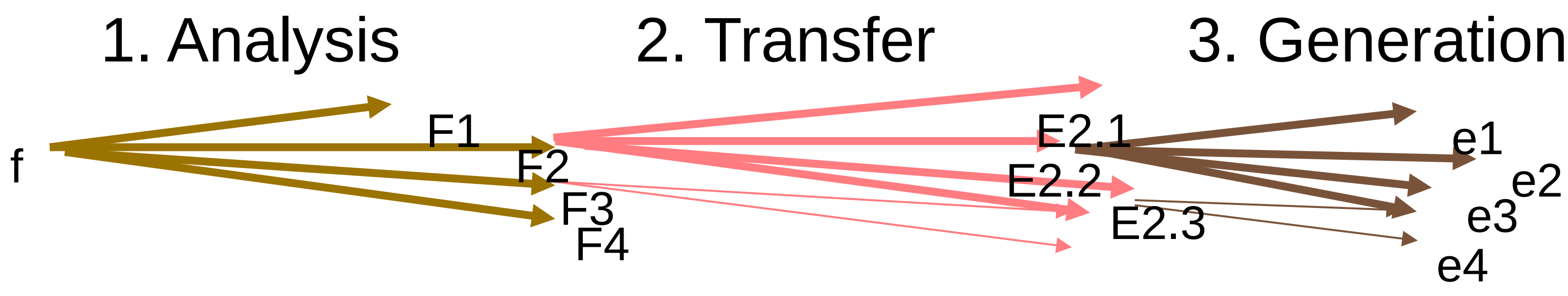
Putting the 3 together



1. First $\arg\max_i P(F_i \mid f)$, say F_2, then $\arg\max_j P(E_j \mid F_2)$ etc

- Theoretically sound:
  - The best parse is in principal independent of the translation, etc.

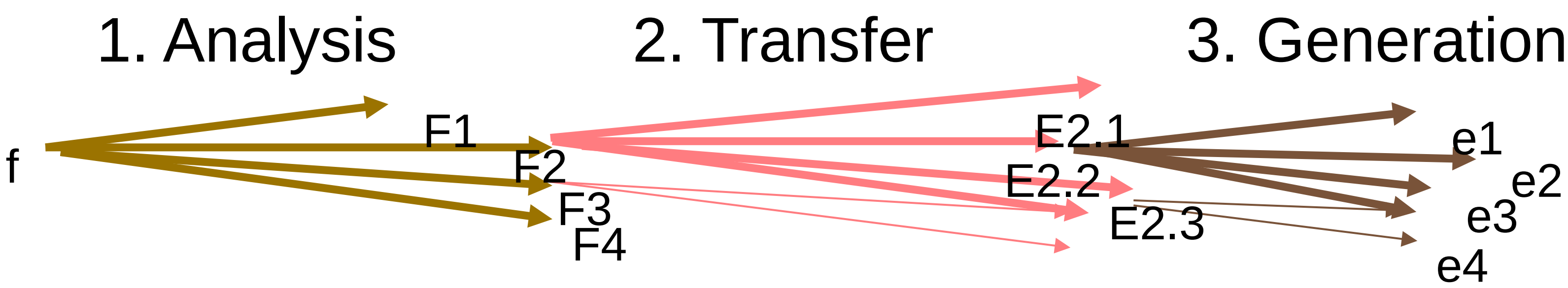
Putting the 3 together



2. The most likely path $\arg\max_{i,j,k} P(e_k \mid E_j)P(E_j \mid F_i)P(F_i \mid f)$

- Might yield better results:
  - When we see that the translation is unlikely, we may detect mistakes earlier in the process

Putting the 3 together



3. The most likely translation $\arg\max_e \sum_{F_i} \sum_{E_j} P(e_k \mid E_j) P(E_j \mid F_i) P(F_i \mid f)$

- Might yield better results:
  - Ambiguities in source language may be the same in target language, e.g. PP-attachement
    - Jeg så mannen i parken med kikkerten
    - I saw the man in the park with the binoculars
    - The same 5 way ambiguity in Norw. and English

End-to-end reranking

- Adding an end-to-end-reranker
  - Goal: rank all the candidates end-to-end towards a modified, sentence-based BLEU-score
- Why?
  - Possibly correct the individual modules
  - Include more information than the three modules e.g.
    - Lexical trans. probabilities
    - Word order etc.
  - Can be considered a refinement/extension of the model 3 on last slide

Results

set	#	chance	first	LL	top	judge
JH_d	1391	34.18	40.95	44.10	49.89	–
JH_t	115	30.84	35.67	38.92	45.74	46.32

Table 4: BLEU scores for various re-ranking configurations, computed over only those cases actually translated by LOGON (second column). For all configurations, BLEU results on the training corpus are higher by about four points.

- 'first' is the first strategy
- LL is the end-to-end reranker, strategy 3+
- Top/judge is human selection of best from all alternatives

Today

- Hybrid translation:
  - Linguistic rule-based
  - + probability ranking
- Linguistic information in STATMT
  - Morphology
  - Word/order - syntax
- State of the art: alternatives
  - Tree-based translation
  - Neural networks

STATMT vs linguisitcs

- The STATMT model works best if there is
  - A 1-1 relationship between words in source sentence and target sentence
  - Same word order
- Not always the case!

STATMT vs linguisitcs

- Linguistic challenges for STATMT
  - Morphology:
    - One source word – many alternative translations
      - STATMT is particularly designed to handle that one word may have alternative translations, but
      - Different forms of the same lexeme is a challenge
    - Not a word-to-word relationship
      - Phrase-based STATMT is designed to meet this, but
      - Synthetic languages (many morphemes in a word) a challenge
  - Syntax:
    - Larger differences in word order is a problem

Different forms of the same lexeme

- English has a poor morphology
- Other languages:
  - Inflection of verbs in person and number
  - Inflection in case and gender: nouns, relative pronouns, determiners, …
- Problems:
  - Sparse training data: a form may not have been seen
  - Challenge to choose the corret form

Morphology

- One possibility:
  - Analyze the training data, replace a fullform with the lemma form and morphological information
  - Learn translation probabilities on lemma pairs
  - Process morphology information separately

f			e
bil	bil+SG+IND	car+SG	car
bilen	bil+SG+DEF	car+SG	car
biler	bil+PL+IND	car+PL	cars
bil	bil+PL+DEF	car+PL	cars

Translating the morphology

f			e
bilen	bil+SG+DEF	car+SG	car

- Some features should be translated:
  - Number
- Other features are ignored:
  - Norw: definiteness (into english)
  - German: case (into Norw. Or english)
- Or determined by the source language (model)

A statistical model

$$p\,(s_e, m_e | s_f, m_f) = p\,(s_e | s_f, m_f)\; p\,(m_e | s_e, s_f, m_f)$$
$$\simeq p\,(s_e | s_f)\; p\,(m_e | m_f)$$

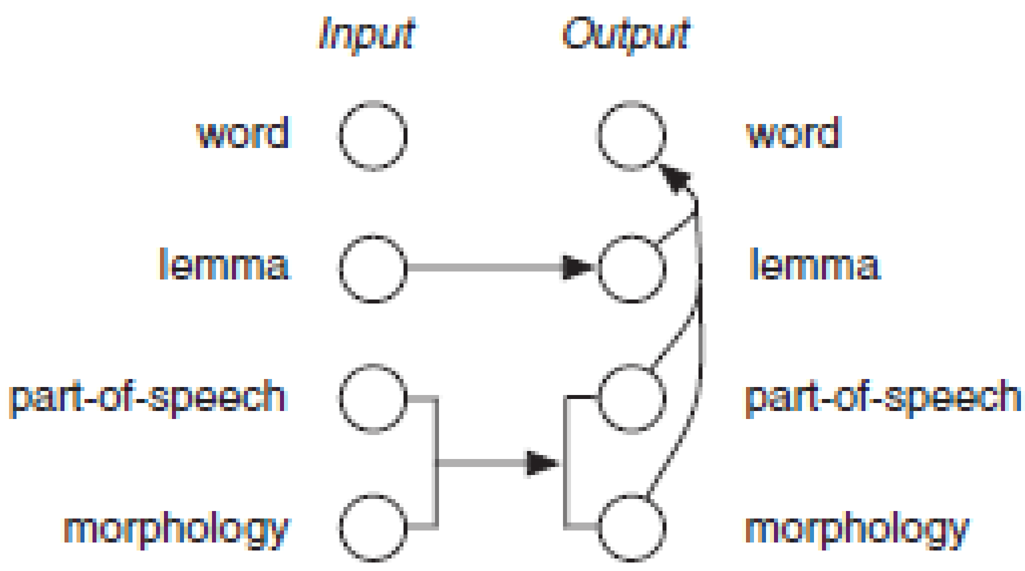
- (s_e is stem of e, m_e is morpholgoy of e, similarly for f)
- But a word may have more than one analysis

$$p(e|f) = \sum_{(s_e, m_e)} p(e|s_e, m_e) \sum_{(s_f, m_f)} p(s_e, m_e|s_f, m_f)\, p(s_f, m_f|f)$$

- Not in use in this form in SMT, but
- motivating factored translation

Factored translation

- Consider a source language word a set of features
- Factor out what should depend on what



häuser

1. **Translation:** Mapping lemmas
   - *haus* → *house, home, building, shell*
2. **Translation:** Mapping morphology
   - *NN|plural-nominative-neutral* → *NN|plural, NN|singular*
3. **Generation:** Generating surface forms
   - *house|NN|plural* → *houses*
   - *house|NN|singular* → *house*
   - *home|NN|plural* → *homes*
   - ...

häuser

1. **Translation:** Mapping lemmas

   { *?|house|?|?,   ?|home|?|?,   ?|building|?|?,   ?|shell|?|?* }

2. **Translation:** Mapping morphology

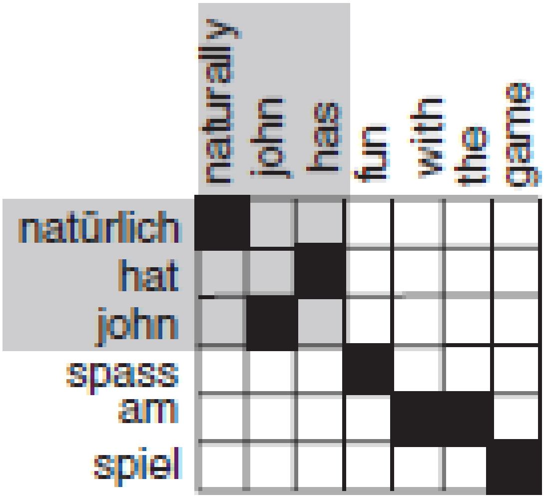
   { *?|house|NN|plural,   ?|home|NN|plural,   ?|building|NN|plural,   ?|shell|NN|plural,   ?|house|NN|singular,* ... }

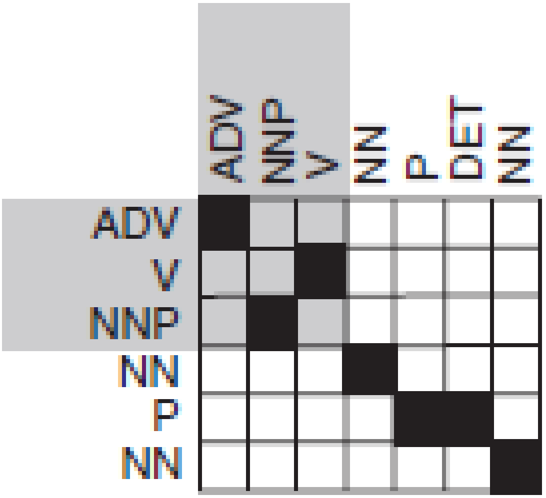
3. **Generation:** Generating surface forms

   { *houses|house|NN|plural,   homes|home|NN|plural,   buildings|building|NN|plural,   shells|shell|NN|plural,   house|house|NN|singular,* ... }

Learning factored model

- Try to learn on the basis of bitext:
  1. Word/phrase-align
  2. Parse/tag both languages separately
  3. (1)+(2) yields:
     1. category/tag alignment
     2. morphology alignment





Decoding factored models

- The book is sparse on details
- Basically the same algorithm as for phrase-based translation

Today

- Hybrid translation:
  - Linguistic rule-based
  - + probability ranking
- Linguistic information in STATMT
  - Morphology
  - Word/order - syntax
- State of the art: alternatives
  - Tree-based translation
  - Neural networks

Word order

- How to handle word-order better?
- Alt 1: Preprocessing
  - Reorder the source sentences in the corpus before word-alignment
- Alt 2: Postprocessing
  - Add rules that reorder the output of the STATMT-system

Syntactic restructuring

- Approach:
  1. Analyze f sentence
  2. Restructure f-sentence to e word order
  3. Use SMT (phrase trans prob.s+LM+dist.)
- Example (German→ English):
  1. Move head verb first
  2. Move subject in front of head verb
  3. etc.

Reordering

- Hand-written rules, or
- Try to learn on the basis of bitext:
  1. Word/phrase-align
  2. Parse/tag both languages separately
  3. (1)+(2) yields category/tag alignment
  4. Try to extract rules
  5. Test the reliability of rules

Tag or parse?

- Tagger
  - Always succeeds
  - Rules like:
    - **V VINF VMFIN → VMFIN V VINF**

    - **VAFIN X* VVFIN → VAFIN VVFIN X***

Parser

- The X*-s are hard to match
  - Many possible candidates
  - Time consuming
- Want to locate HEADVERB, SUBJ, …
- **SUBJ VAINF OBJ* VVFIN →**
  SUBJ VAINF VVFIN OBJ*
- Reorders a local tree
  (daughters of the same mother)
- Try to keep the alternatives

Syntactic post-editing

- Use syntactic features in the post-editing reranking
- E.g.
  - Number agreement source – target
  - Agreement Verb – Subject
- Use a parser to rerank:
  - Grammatical output better than ungrammatical

Today

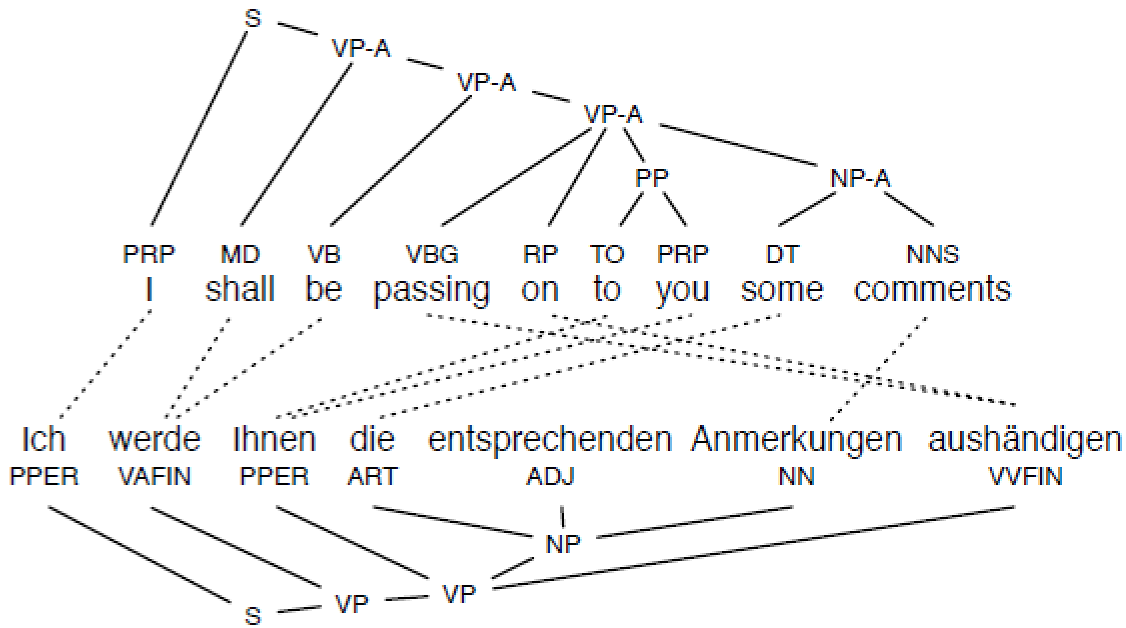
- Hybrid translation:
  - Linguistic rule-based
  - + probability ranking
- Linguistic information in STATMT
  - Morphology
  - Word/order - syntax
- State of the art: alternatives
  - Tree-based translation
  - Neural networks

Tree-based models

- A different approach to statistical MT.
  - Instead of aligning words or phrases
  - Aligning trees

- Conceiving the difference:
  - Word-based STATMT can be considered a combination of traditional direct approach + probabilities
  - Tree-based STATMT can be considered a combination of syntactic transfer + probabilities

Aligned Tree Pair



Phrase structure grammar trees with word alignment
(German–English sentence pair.)

Tree-based

- We will not consider the tree-based models
  - Too much
  - In flux

What Works Best?

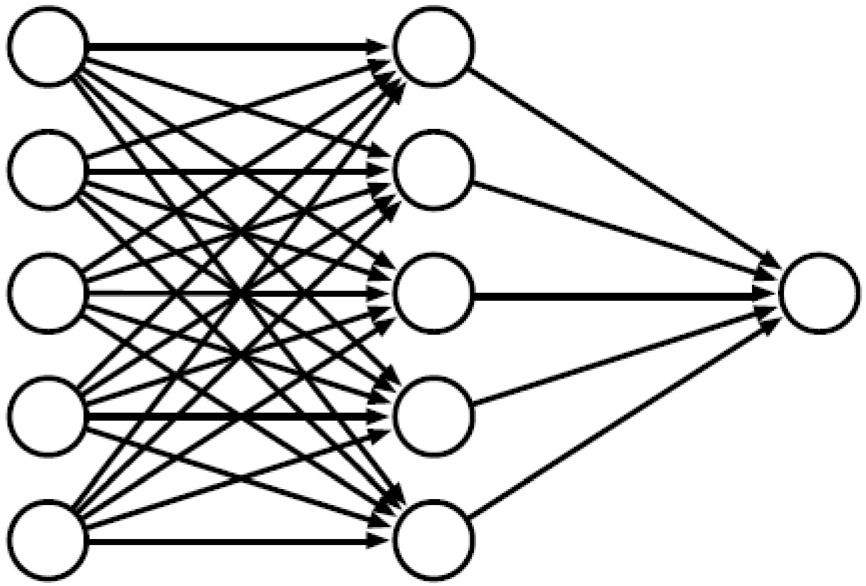
- WMT evaluation campaign

- Winner English–German (with official ties)

System	2008	2009	2010	2011	2012	2013	2014	2015	2016
rule	X	X		X	X	X			
phrase			X	X	X	X	X		
syntax							X	X	
neural								X	X

- For other language pairs, phrase-based systems dominated longer

Neural Networks



- Real valued vector representations
- Multiple layers of computation
- Non-linear functions

$$\vec{h} = \text{sigmoid}(W\vec{x})$$

$$\vec{y} = \text{sigmoid}(V\vec{h})$$

Deep learning: neural nets

- A large shift towards nural network models in the 2010s
- Great success:
  - Image reconition
  - Speech recognition
- Tested for all types of NLP tasks
  - Including MT
  - Will probably have to be included in future curriculum