

Introduction to Computational Linguistics

PD Dr. Frank Richter

(all slides provided by Prof. Dr. Erhard W. Hinrichs)

fr@sfs.uni-tuebingen.de.

**Seminar für Sprachwissenschaft
Eberhard-Karls-Universität Tübingen
Germany**

A bit of Philosophy of Science

- **Theory:**

A set of statements that determine the format and semantics of descriptions of phenomena in the purview of the theory

- **Methodology:**

An effective theory comes with an explicit methodology for acquiring these descriptions

- **Application:**

A theory associated with a methodology can be applied to tasks for which the methodology is appropriate.

Scientific Strategies

- **Method Oriented Approach:**

devise or import a tool, a procedure or a formalism, apply it to a task and develop it further. Then (optionally) see whether it works for additional tasks

- **Task oriented Approach:**

select a task; devise or import a method or several methods for its solution; integrate the methods as required to improve performance.

What Makes Machine Translation Hard

- Lexical Ambiguity

What Makes Machine Translation Hard

- Lexical Ambiguity
- Lexical Gaps

What Makes Machine Translation Hard

- Lexical Ambiguity
- Lexical Gaps
- Syntactic Divergences between Source and Target Language

Problems: Word-to-Word Translations

English – German

The ticket office in the train station
Der Fahrkartenschalter im Bahnhof

öffnet wieder um ein Uhr.
re-opens at one o'clock.

Lexical Ambiguity: Open (1)

English

in store door

on new building

open door

open golf tourney

open question

open job

open morning

open football player

German

Offen

Neu eröffnet

Tür öffnen

Golfspiel eröffnen

offene Frage

freie Stelle

freier Morgen

freier Fußballspieler

Lexical Ambiguity: Open (2)

English

loose ice

blank endorsement

private firm

unfortified town

blank cheque

to unbutton a coat

German

offenes Eis

offenes Giro

offene Handelsgesellschaft

offene Stadt

offener Wechsel

einen Mantel öffnen

Structural Divergence (1)

English – German

Max likes to swim.

NP VFIN INF

Max schwimmt gerne.

NP VFIN ADV

Structural Divergence (2)

Russian – English

Jego zovut Julian.
Him they call Julian.
They call him Julian.

Japanese – English

Kino ame ga futa.
Yesterday rain fell.
It was raining yesterday.

Differences in Word Order

English – German

Does it make sense to translate

Macht es Sinn

documents automatically ?

Dokumente automatisch zu übersetzen ?

MT: The Weaver Memo (1)

- Translation and Context

If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words.

MT: The Weaver Memo (1)

- Translation and Context

If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words.

But if one lengthens the slit in the opaque mask, until one see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word.

MT: The Weaver Memo (2)

- Translation and Context

The practical question is: “What minimum value of N will, at least, in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?”

MT: The Weaver Memo (2)

- Translation and Context

The practical question is: “What minimum value of N will, at least, in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?”

- Translation and Cryptography

... it is very tempting to say that a book written in Chinese is simply a book written in English which was coded into the “Chinese code”.

MT: The Weaver Memo (3)

- Translation and Language Universals (Invariants)
... there are certain invariant properties which are, again not precisely, but to some statistically useful degree, common to all languages.

MT: The Weaver Memo (3)

- Translation and Language Universals (Invariants)
 - ... there are certain invariant properties which are, again not precisely, but to some statistically useful degree, common to all languages.
 - Thus may it be true that the way to translate Chinese to Arabic or from Russian to Portuguese, is not to attempt the direct route ... but down to the common base of human communication – the real but yet undiscovered universal language – and then to re-emerge by whatever particular route is convenient.

Strategies for Machine Translation

- Word-to-Word (Direct) Translation

Strategies for Machine Translation

- Word-to-Word (Direct) Translation
- Syntactic Transfer

Strategies for Machine Translation

- Word-to-Word (Direct) Translation
- Syntactic Transfer
- Semantic Transfer

Strategies for Machine Translation

- Word-to-Word (Direct) Translation
- Syntactic Transfer
- Semantic Transfer
- Interlingua Approach

Strategies for Machine Translation (2)

- Word-to-Word (Direct) Translation
 - simplest approach:

Strategies for Machine Translation (2)

- Word-to-Word (Direct) Translation
 - simplest approach:
 - may require only an electronic, bi-lingual dictionary

Strategies for Machine Translation (2)

- Word-to-Word (Direct) Translation
 - simplest approach:
 - may require only an electronic, bi-lingual dictionary
 - depending on the source and target languages and the dictionary, minimal morphological analysis and generation may be required.

Strategies for Machine Translation (2)

- Word-to-Word (Direct) Translation
 - simplest approach:
 - may require only an electronic, bi-lingual dictionary
 - depending on the source and target languages and the dictionary, minimal morphological analysis and generation may be required.
 - no use of syntactic or semantic knowledge

Strategies for Machine Translation (3)

- Syntactic Transfer

Strategies for Machine Translation (3)

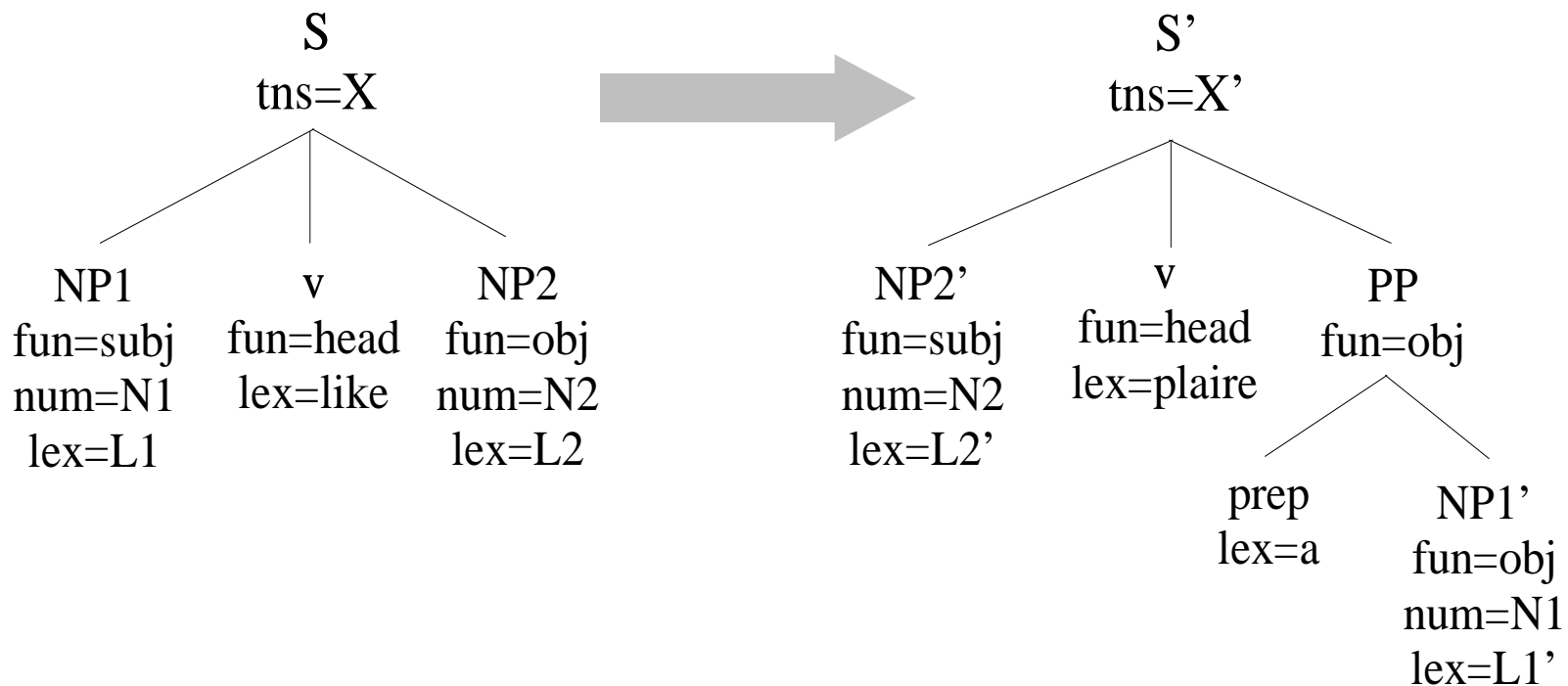
- Syntactic Transfer
 - requires syntactic analysis of the source language

Strategies for Machine Translation (3)

- Syntactic Transfer
 - requires syntactic analysis of the source language
 - requires a syntactic parser

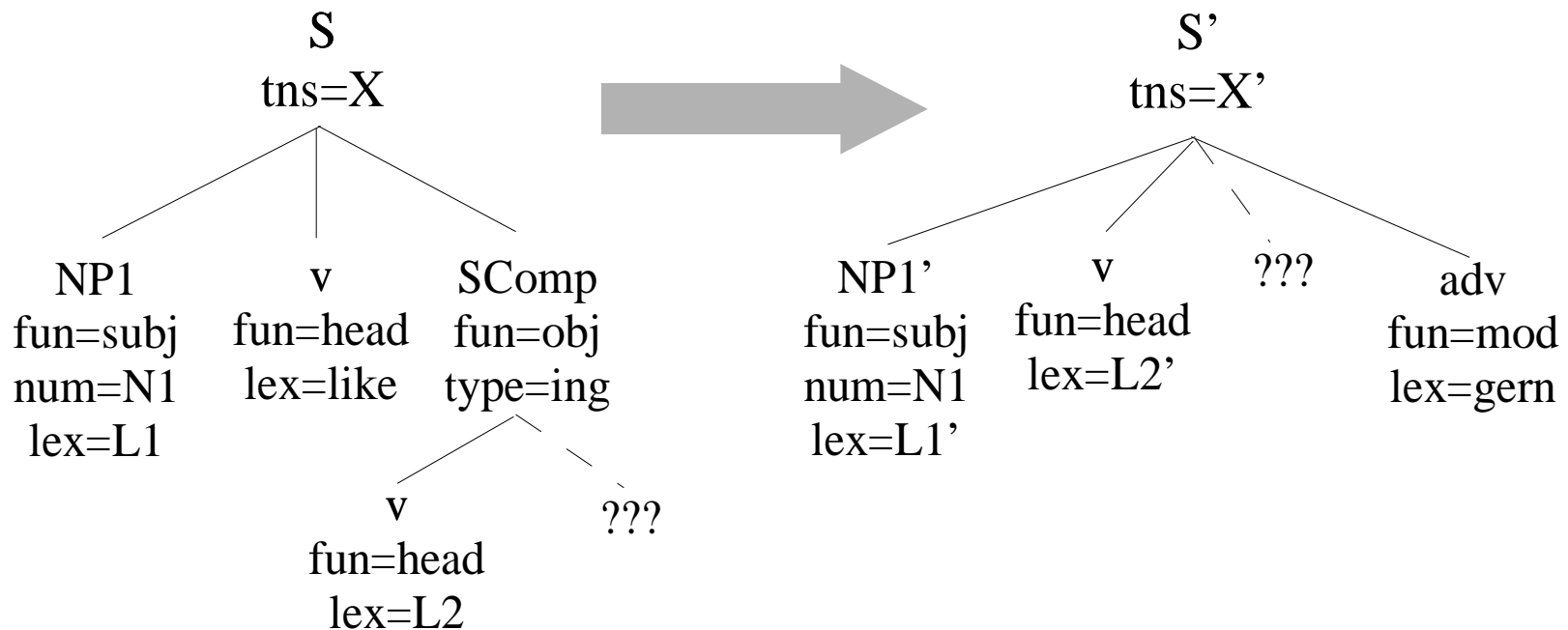
Syntactic Transfer Trees

An Example of a Transfer Tree for English *like* and French *plaire*



Syntactic Transfer Trees (2)

An Example of a Transfer Tree for English *like to* $\langle V \rangle$ and German $\langle V \rangle$ *gern*



Strategies for Machine Translation (4)

- Semantic Transfer
 - requires syntactic and semantic analysis of the source language

Strategies for Machine Translation (4)

• Semantic Transfer

- requires syntactic and semantic analysis of the source language
- requires language-dependent meaning representation language

Strategies for Machine Translation (4)

• Semantic Transfer

- requires syntactic and semantic analysis of the source language
- requires language-dependent meaning representation language
- language-dependent rules that relate source language meaning representations to target language meaning representations

Strategies for Machine Translation (4)

• Semantic Transfer

- requires syntactic and semantic analysis of the source language
- requires language-dependent meaning representation language
- language-dependent rules that relate source language meaning representations to target language meaning representations
- requires language generation component which maps target language meaning representations to output sentences

Strategies for Machine Translation (5)

• Semantic Transfer

- synthesis typically performed in two stages: semantic synthesis (resulting in syntactic trees) and morphological synthesis (resulting in strings of inflected word forms).

Strategies for Machine Translation (5)

- Interlingua Approach
 - source language input is mapped to a language-neutral (quasi-universal) meaning representation language

Strategies for Machine Translation (5)

• Interlingua Approach

- source language input is mapped to a language-neutral (quasi-universal) meaning representation language
- requires syntactic and semantic analysis of the source language into interlingua

Strategies for Machine Translation (5)

• Interlingua Approach

- source language input is mapped to a language-neutral (quasi-universal) meaning representation language
- requires syntactic and semantic analysis of the source language into interlingua
- requires language generation component which maps interlingua to output sentences

Strategies for Machine Translation (5)

● Interlingua Approach

- source language input is mapped to a language-neutral (quasi-universal) meaning representation language
- requires syntactic and semantic analysis of the source language into interlingua
- requires language generation component which maps interlingua to output sentences
- synthesis typically performed in two stages: semantic synthesis from the interlingua (resulting in syntactic trees) and morphological synthesis (resulting in strings of inflected word forms).