

# Lecture 25: Machine Translation

Ling 1330/2330 Intro to Computational Linguistics  
Na-Rae Han, 11/30/2023

# Today

---

- ▶ Guest presentation by Tianyi Zheng
  - ◆ Deep-learning language models
  
- ▶ Machine translation
  - ◆ Interlingua
  - ◆ Noisy channel model
  - ◆ Alignment, parallel corpora

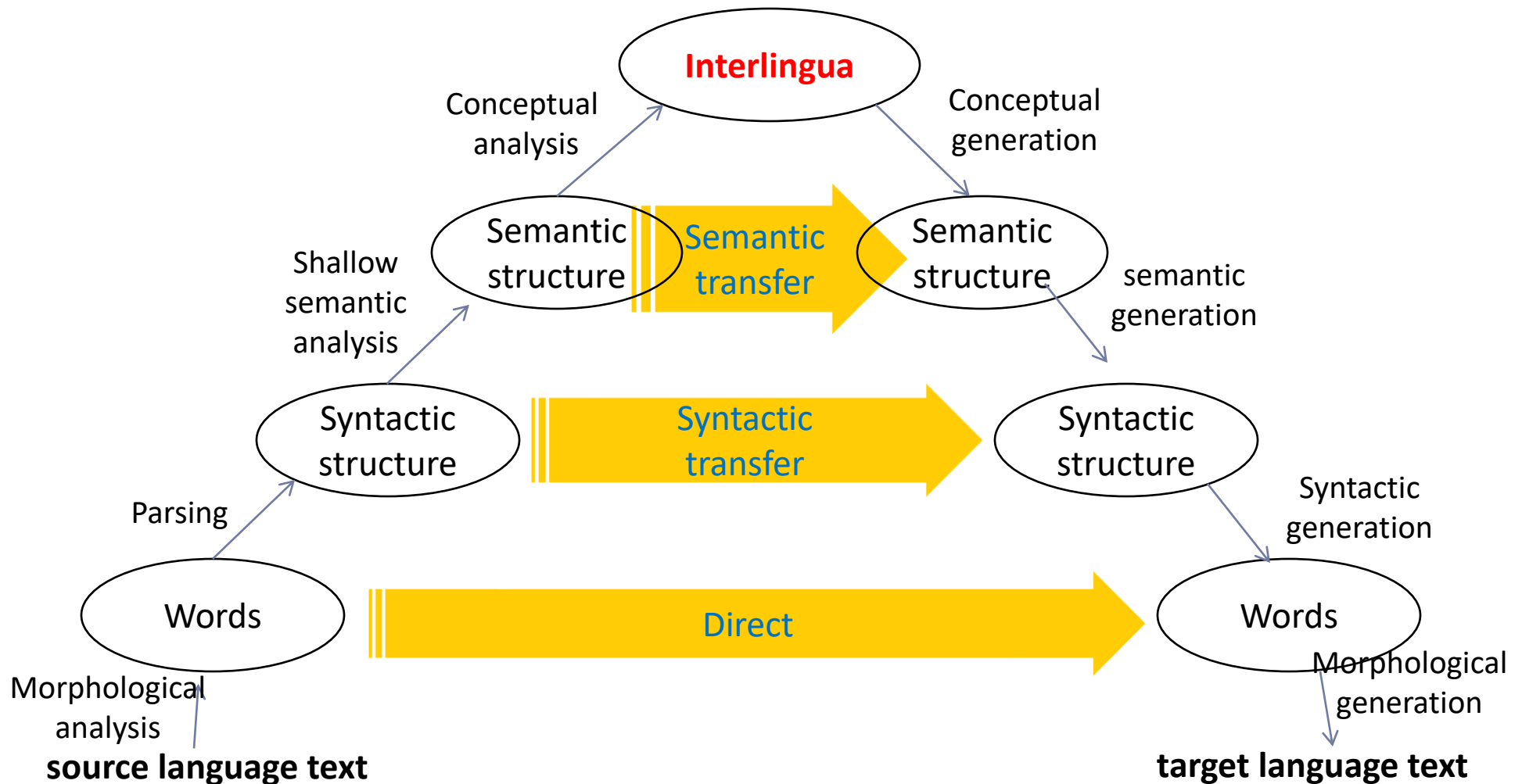
# Classical MT

---

- ▶ Classical, pre-statistical architecture for MT
  - ◆ **Direct translation**
    - ◆ Word-to-word translation
  - ◆ **Transfer approaches**
    - ◆ Uses syntactic and possibly semantic transfers
    - ◆ Based on **transfer rules**: hand-crafted per source-target language pair
  - ◆ **Interlingua approaches**
    - ◆ Source language is mapped to an **abstract meaning representation (interlingua)**. Target language is generated from it.

# The translation triangle & interlingua

## ► The Vauquois (1968) triangle



# Interlingua approaches

---

- ▶ View on translation in direct and transfer approaches:
  - ◆ MT transforms the source text to target text
- ▶ An interlingua view on translation:
  - ◆ MT extracts meaning from the source language and expresses it in target language
- ▶ **Interlingua**
  - ◆ An abstract meaning representation
  - ◆ Obtained from the semantic representation (language-independent) via semantic decomposition



Yeeesssss....  
"language-independent"

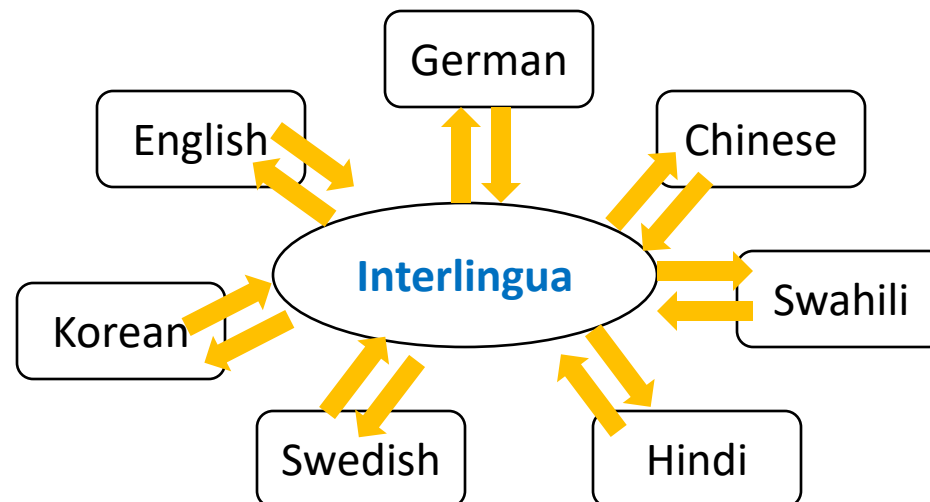
# Interlingua approaches: advantage

---

- ◆ **ADVANTAGE: A single interlingua** intermediates between all language pairs. For  $n$  languages, only  $2*n$  translation models need to be built:

- English → INTERLINGUA → French
- English → INTERLINGUA → Spanish
- English → INTERLINGUA → Japanese
- Japanese → INTERLINGUA → English

- ◆ *Each language only needs to be translated to and from INTERLINGUA:*



# Statistical MT

---

- ▶ The three classic architectures focus on the appropriate representations to use → **symbolic approach**
- ▶ In **statistical MT**, however, the focus is more on the result
  - ◆ JAPANESE: *fukaku hansei shite orimasu.*
  - ◆ English1: *I sincerely apologize.*
  - ◆ English2: *I am deeply regretting.*
- ← E1 is more **natural** English; E2 is more **faithful** to the original meaning
- ← There are two considerations in translation:
  - (1) **Faithfulness** to the original message
  - (2) **Fluency (naturalness)** of the target language text
- ← Successful translation can be schematized as:

A translation T that maximizes the product of the two

$$\text{best-translation } \hat{T} = \operatorname{argmax}_T \text{faithfulness}(T, S) * \text{fluency}(T)$$

# Wrapping up

---

- ▶ Next class:
  - ◆ MT wrap: Neural MT
  - ◆ Formal language theory
- ▶ Homework 10 due next Thu
  - ◆ Submit on Canvas
  - ◆ Also: **extra credit** opportunity
- ▶ Grades, late work forgiveness →
- ▶ Extra credit →
- ▶ Final exam info →



# Your grade: what's ahead

---

## ▶ **Canvas's Grade Center** is being prepped

- ◆ Your exercise score is in
- ◆ Homework 9 and 10 grades are outstanding
- ◆ Attendance & participation records (will post 2<sup>nd</sup> half attendance soon)
  - ◆ **1 missed class exemption → raised to 2**
- ◆ Weighted running total (CAVEAT!!)

## ▶ **Late work forgiveness**

- ◆ Everyone gets one make-up opportunity. Choose from:
  1. Finish up an incomplete homework submission or re-do a part, no penalty.
  2. Up to 3 days of late submission penalty waived.
  3. Missed homework: 25% penalty. Upload on Canvas and email me.
  4. Missed exercise: 5/10 for satisfactory (80+%) work. Email me as attachment.
- ◆ Deadline: **12/15 (Fri) 11:59pm. Email me and let me know of your choice!**
- ◆ If a solution has been published, feel free to look it up. It's fine as long as you don't blindly copy it. (Make sure to demonstrate you are not blindly copying.) There's already a late penalty, and I'd rather you learn.

# Extra credit, round-up

---

- ▶ If you have 100% on Exercises, you are already eligible for a standard round-up, up to 0.4%.
  - ◆ Normally 89.6% is B+; it will be bumped up to A- (90%)
  
- ▶ Extra credit opportunity (1): NLP talk
  - ◆ Attend tomorrow's Linguistics dept colloquium:
    - ◆ Dec 1 (Fri) 3pm, G8 CL. [Lorraine Li](#), "Probabilistic (Commonsense) Knowledge in Language"
    - ◆ If you can't attend this one, find a different CL/NLP talk (CMU, Pitt, online)
  - ◆ Submit a short report on Canvas, earn 0.3% extra credit
  
- ▶ Extra credit opportunity (2): share HW10 essays
  - ◆ On MS Teams, share your HW 10 essay, read 3 classmates' essays and leave comments, earn 0.3% extra credit

Both due 12/15  
(Friday) 11:59pm

# Final exam

---

- ▶ 12/13 (Wed), 4—5:50pm
- ▶ At G17 CL (Language Media Center)
  
- ▶ 150 total points (50% larger than midterm)
- ▶ All pen-and-pencil based.
- ▶ **1 cheat sheet allowed:**
  - ◆ letter-sized, front-and-back, hand-written.
- ▶ Cumulative! 10-20% will be from first half of the semester.
- ▶ Make sure to study book chapters and other linked materials. Post-midterm, my slides are not as "comprehensive".