Natural Language Processing
CSE 325/425

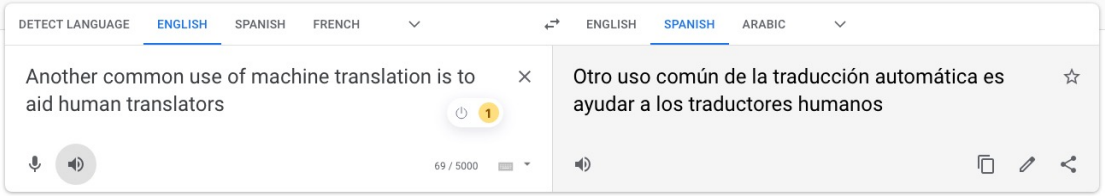
Sihong Xie

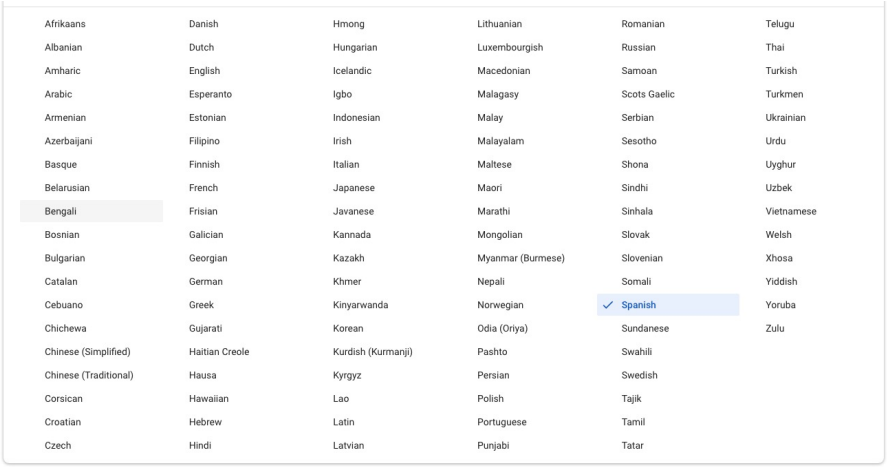
Lecture 22:

- Machine translation
- Direct transfer methods

Machine translation

- Translating sentences in one language to those in another.





  - Still active research area in NLP.
  - Need large-scale training corpora and computing infrastructures.

Why MT is hard

- Many differences between two languages on many levels.
  - lexical level
    - "*bass*" in English can be translated into "*bajo*" (a musical instrument) or "*lubina*" (a fish) in Spanish.
    - "wall" in English can be translated into "*Wand*" (wall inside a building), or "*Mauer*" (wall outside a building) in German.
    - "brother" in English can be translated into "哥哥" (older brother) or "弟弟" (younger brother) in Chinese.
  - syntactic level
    - In English: no gender for adjectives; in French/Spanish, adjectives can have genders.

Why MT is hard

- Many differences between two languages on many levels.
  - syntactic level (word ordering)
    - SVO (Subject-verb-object): English
      - "*He adores listening to music*"
    - SOV (Subject-objective-verb): Japanese
      - word-to-word translation from Japanese

        "*he music to listening adores*"
    - different places for pre-position phrases (PP)

English: *He wrote a letter to a friend*
Japanese: *tomodachi ni tegami-o kaita*
friend to letter wrote
Arabic: *katabt risāla li ṣadq*
wrote letter to friend

Why MT is hard

- Many differences between two languages on many levels.
  - syntactic level (word ordering)
    - SVO (Subject-verb-object): English
      - "*He adores listening to music*"
    - SOV (Subject-objective-verb): Japanese
      - word-to-word translation from Japanese

        "*he music to listening adores*"
    - different places for pre-position phrases (PP)

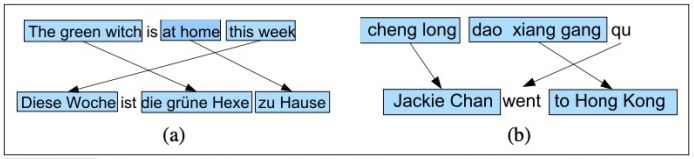


Figure 11.1 Examples of other word order differences: (a) In German, adverbs occur in initial position that in English are more natural later, and tensed verbs occur in second position. (b) In Mandarin, preposition phrases expressing goals often occur pre-verbally, unlike in English.

English: *He wrote a letter to a friend*
Japanese: *tomodachi ni tegami-o kaita*
friend to letter wrote
Arabic: *katabt risāla li ṣadq*
wrote letter to friend

- different adjective-noun ordering

Spanish *bruja verde* **English** *green witch*

When MT is easy

- When the vocabulary and syntactic patterns are limited.
  - weather forecasting:
    - *“There will be rain on Friday”*
  - Instructions (e.g., software manuals, recipes)

Platano en Naranja Para 6 personas

3 Plátanos maduros
1 taza de jugo (zumo) de naranja
1/8 cucharadita de nuez moscada en polvo
1 cucharada de canela en polvo (opcional)
2 cucharadas de mantequilla derretida
5 cucharadas de azúcar morena o blanc
1 cucharada de ralladura de naranja

Pelar los plátanos, cortarlos por la mitad y, luego, a lo largo. Engrasar una fuente o pirex con margarina. Colocar los plátanos y bañarlos con la mantequilla derretida. En un recipiente hondo, mezclar el jugo (zumo) de naranja con el azúcar, jengibre, nuez moscada y ralladura de naranja. Verter sobre los plátanos y hornear a 325 ° F. Los primeros 15 minutos, dejar los pátanos cubiertos, hornear 10 o 15 minutos más destapando los plátanos

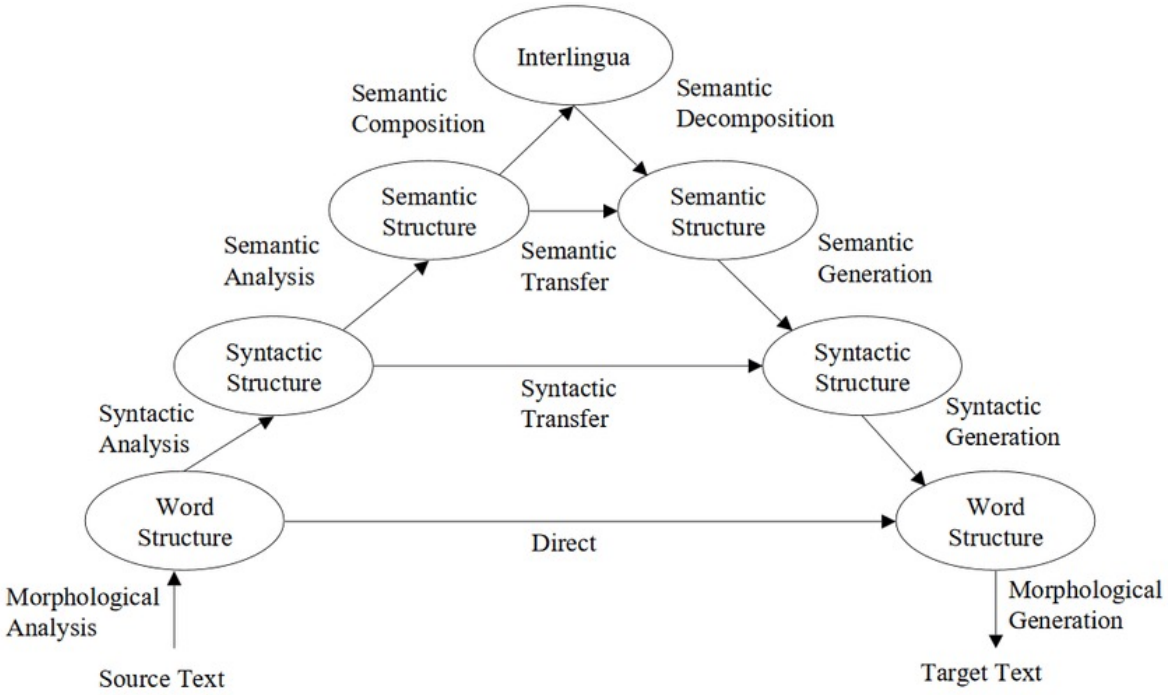
Platano in Orange For 6 people

3 Bananas mature
1 cup juice (juice) orange
1/8 teaspoon nutmeg powder
1 tablespoon cinnamon powder (optional)
2 tablespoon melted butter
5 tablespoons brown sugar or white
1 tablespoon ralladura orange

Peel bananas, cut in half and then along. Grease a source or pirex with margarine. Put bananas and showering them with the melted butter. In a deep bowl, mix the juice (juice) orange with the sugar, ginger, nutmeg and ralladura orange. Pour over bananas and bake to 350° F. The first 15 minutes, leave covered bananas, bake 10 to 15 minutes more uncovering bananas.

Classical (non-statistical) MT

- Vouquois /'va:kua:/ triangle



Classical (non-statistical) MT

- Direct method
  - translating from English to Spanish

Mary didn't slap the green witch

Maria no dió una bofetada a la bruja verde

Mary not gave a slap to the witch green

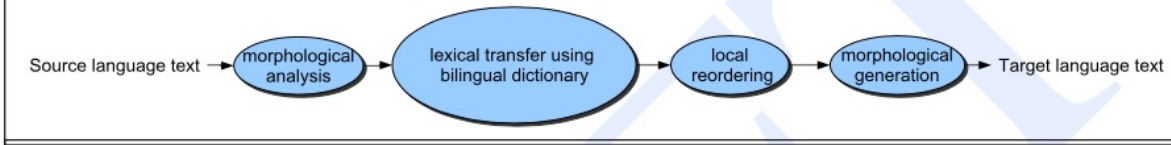


Figure 25.5 Direct machine translation. The major component, indicated by size here, is the bilingual dictionary.

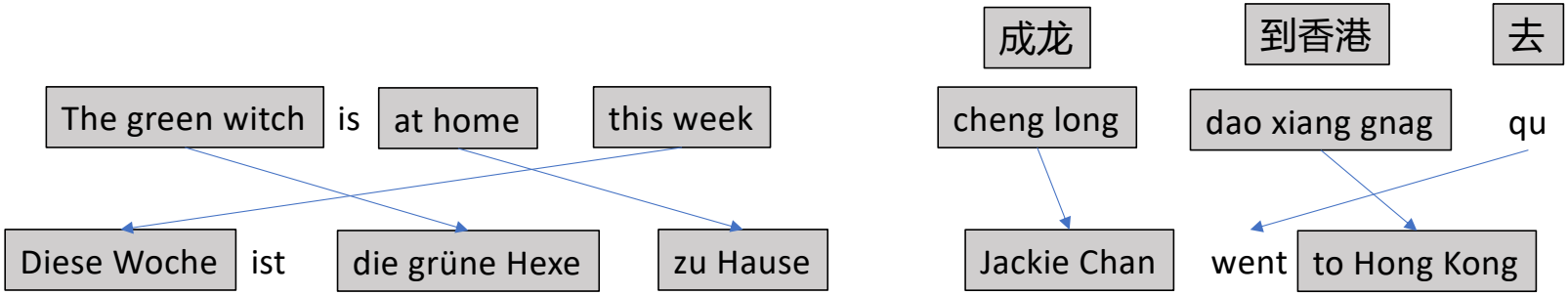
Input:	Mary didn't slap the green witch
After 1: Morphology	Mary DO-PAST not slap the green witch
After 2: Lexical Transfer	Maria PAST no dar una bofetada a la verde bruja
After 3: Local reordering	Maria no dar PAST una bofetada a la bruja verde
After 4: Morphology	Maria no dió una bofetada a la bruja verde

Figure 25.6 An example of processing in a direct system

drawbacks: can't handle different orders of larger linguistic units, such as phrases.

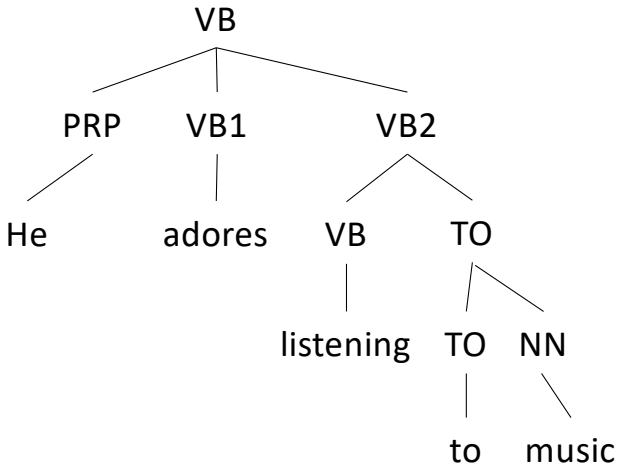
Classical (non-statistical) MT

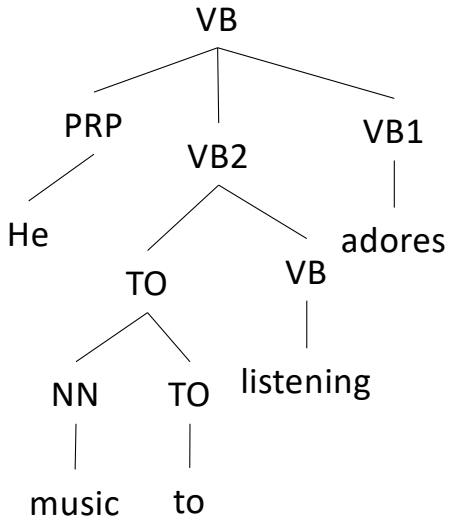
- Direct method
  - drawbacks: can't handle different orders of larger linguistic units, such as phrases.



Classical (non-statistical) MT

- Transfer method
  - syntactic transfer: transforming a parsing tree into another.
  - lexical transfer: translate words to words.





Statistical MT

- Metrics
  - Fluency and faithfulness.
  - Hebrew: *adonai roi* to English "*the Lord is my shepherd*"
  - to some other language: "*the Lord will look after me*" (fluent but not too faithful)
  - or "*the Lord is for me like somebody who looks after animals with cotton-like hair*" (faithful but not fluent)
  - Need some trade-off between the two goals.

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

Statistical MT

- Metrics
  - Fluency and faithfulness.
  - Hebrew: *adonai roi* to English "*the Lord is my shepherd*"
  - to some other language: "*the Lord will look after me*" (fluent but not too faithful)
  - or "*the Lord is for me like somebody who looks after animals with cotton-like hair*" (faithful but not fluent)
  - Need some trade-off between the two goals.

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

  - More formally,

$$\text{best-translation } \hat{T} = \text{argmax}_T \text{ P(T) } P(S|T)$$