

NLP

Machine Translation

Basic Techniques

Translation as Decoding

- “One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.' “
- Warren Weaver, “Translation (1955)”

Question For The Audience

- If you know (some) French
- Look at the next two slides
- They contain a recipe in English
 - and its automatic translation to French by Google
- Can you identify the problems in the translation?

Automatic Translation

- Systran
- Google Translate

Ingredients

Original recipe makes 4 - 6 servings

2 cups heavy whipping cream

1 tablespoon chopped fresh basil

1 tablespoon chopped fresh thyme

2 teaspoons salt

2 teaspoons ground black pepper

1 1/2 teaspoons crushed red pepper flakes

1 teaspoon ground white pepper

1 cup chopped green onions

1 cup chopped parsley

1/2 pound shrimp, peeled and deveined

1/2 pound scallops

1/2 cup shredded Swiss cheese

1/2 cup grated Parmesan cheese

1 pound dry fettuccine pasta

Directions

Cook pasta in a large pot of boiling salted water until al dente. Meanwhile, pour cream into large skillet. Cook over medium heat, stirring constantly, until just about boiling. Reduce heat, and add herbs, salt, peppers, onions, and parsley. Simmer 7 to 8 minutes, or until thickened. Stir in seafood, cooking until shrimp is no longer transparent. Stir in cheeses, blending well. Drain pasta. Serve sauce over noodles.

<http://translate.google.com/>

ingrédients

Recette originale fait 4 - 6 portions

2 tasses de crème épaisse à fouetter

1 cuillère à soupe de basilic frais haché

De thym frais 1 cuillère à soupe

2 cuillères à café de sel

2 cuillères à café de poivre noir moulu

1 1/2 cuillères à café écrasés de flocons de piment rouge

Sol 1 cuillère à café de poivre blanc

1 tasse oignons verts hachés

1 tasse de persil haché

1/2 livre de crevettes, décortiquées et déveinées

1/2 escalopes *de livres*

1/2 tasse de fromage râpé suisse

1/2 tasse parmesan râpé

£ 1 sec pâtes fettucine

instructions

Cuire les pâtes dans une grande casserole d'eau bouillante salée jusqu'à ce qu'elles soient al dente. Pendant ce temps, versez la crème dans une grande poêle. Cuire à feu moyen, en remuant constamment, jusqu'à ce que presque ébullition. Réduire le feu et ajoutez les herbes, sel, poivrons, oignons et persil. Laisser mijoter 7 à 8 minutes, ou jusqu'à ce que le mélange épaississe. Incorporer les fruits de mer, cuisson jusqu'à ce que les crevettes ne soit plus transparente. Incorporer le fromage en mélangeant bien. Égoutter les pâtes. Servir la sauce sur les nouilles.

Answer

Cuire les pâtes dans une grande casserole d'eau bouillante salée jusqu'à ce qu'elles soient al dente. Pendant ce temps, versez la crème dans une grande poêle. Cuire à feu moyen, en remuant constamment, *jusqu'à ce que presque ébullition*. Réduire le feu et *ajoutez* les herbes, sel, poivrons, oignons et persil. Laisser mijoter 7 à 8 minutes, ou jusqu'à ce que le mélange épaississe. *Incorporer* les fruits de mer, *cuisson* jusqu'à ce que les crevettes ne soit plus *transparente*. Incorporer le fromage en mélangeant bien. Égoutter les pâtes. Servir la sauce sur les nouilles.

Language Differences (1/2)

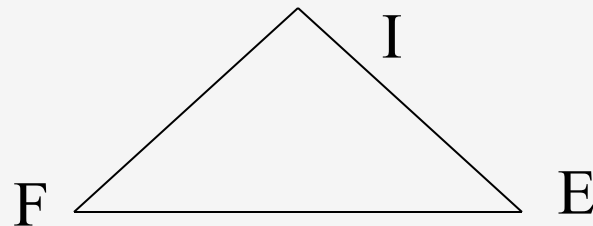
- Word order
 - SVO: English, Mandarin
 - VSO: Irish, Classical Arabic
 - SOV: Hindi, Japanese
- Prepositions (Jap.)
 - to Mariko, Mariko-ni
- Inflection (Sp.)
 - have: tengo, tienes, tenemos, tienen, tener
- Lexical distinctions (Sp.):
 - the bottle floated out – la botella salió flotando
- Brother (Jap.)
 - oototo (younger), oniisan (older)
- They (Fr.)
 - elles (feminine), ils (masculine)

Language Differences (2/2)

- Word order in phrases (Fr.)
 - la maison bleue, the blue house
- Word order in sentences (Jap.)
 - I like to drink coffee
 - watashi wa kohii o nomu no ga suki desu
 - I-subj coffee-obj drink-dat-rheme like
- vocabulary (Sp.)
 - wall
 - pared, muro
- phrases (Fr.)
 - play
 - pièce de théâtre

Basic Strategies of MT

- **Direct Approach**
 - 50's, 60's
 - naïve
 - the flesh is weak, but the spirit is strong
 - out of sight, out of mind
- **Indirect: Transfer**
- **Indirect: Interlingua**
 - No looking back
 - Language-neutral
 - No influence on the target language



Basic Strategies of MT

- This is a blue house.
- Direct Approach
 - translate each word separately
- Transfer
 - Eng (adj noun) \rightarrow Fr (noun adj)
- Interlingua
 - $\text{House}(H) \wedge \text{Blue}(H)$

NLP