Language and Computers (Ling 384)

Topic 5: Machine Translation

Adriane Boyd*
Department of Linguistics, OSU
Autumn 2005

The course was created by Markus Dickinson, Detmar Meurers and Chris Brew.

What is Machine Translation?

Translation is the process of:
- moving texts from one (human) language (source language) to another (target language),
- in a way that preserves meaning.

Machine translation (MT) automates (part of) the process:
- Fully automatic translation
- Computer-aided (human) translation

What is MT good for?

- When you need the gist of something and there are no human translators around:
  - translating e-mails & webpages
  - obtaining information from sources in multiple languages (e.g., search engines)
- If you have a limited vocabulary and a small range of sentence types:
  - translating weather reports
  - translating technical manuals
  - translating terms in scientific meetings
  - determining if certain words or ideas appear in suspected terrorist documents — help pin down which documents need to be looked at closely
- If you want your human translators to focus on interesting/difficult sentences while avoiding lookup of unknown words and translation of mundane sentences.

What is MT not good for?

- Things that require subtle knowledge of the world and/or a high degree of (literary) skill:
  - translating Shakespeare into Navajo
  - diplomatic negotiations
  - court proceedings
  - ...  
- Things that may be a life or death situation:
  - Pharmaceutical business
  - Automatically translating frantic 911 calls for a caller who speaks only Spanish

What goes into a translation?

Some things to note about these examples and thus what we might need to know to translate:
- Words have to be translated. → dictionaries
- Words are grouped into meaningful units. (cf., our discussion of syntax for grammar checkers).
- Word order can differ from language to language.
- The forms of words within a sentence are systematic, e.g., verbs have to be conjugated, etc.

Example translations

The simple case

- It will help to look at a few examples of real translation before talking about how a machine does it.
- Take the simple Spanish sentence and its English translation below:

  (1) Yo hablo español.
  I speak Spanish.

- Words in this example pretty much translate one-for-one
- But we have to make sure hablo matches with Yo, i.e., that the subject agrees with the form of the verb.

Example translations

A slightly more complex case

The order and number of words can differ:

(2) a. Tu hablas española?
   You speak Spanish?

   b. Hablas español?
   Do you speak Spanish?

References
Different approaches to MT

- Transformer systems
- Systems based on linguistic knowledge
  - Direct transfer systems
  - Interlinguas
- Machine learning approaches

Most of these use dictionaries in one form or another, so we will start by looking at dictionaries.

What dictionary entries might look like

- **WORD**: knob
  - **PART OF SPEECH**: noun
  - **HUMAN**: no
  - **CONCRETE**: yes
  - **GERMAN**: Knopf

- **WORD**: knowledge
  - **PART OF SPEECH**: noun
  - **HUMAN**: no
  - **CONCRETE**: no
  - **GERMAN**: Wissen, Kenntnisse
    - There can be extra rules which tell you whether to choose Wissen or Kenntnisse.

Dictionaries

An MT **dictionary** is differs from a “paper” dictionary:

- must be computer-usable (electronic form, indexed)
- contain the inherent properties (meaning) of a word
- need to be able to handle various word inflections
- have is the dictionary entry, but we want the entry to specify how to conjugate this verb.

A dictionary entry with frequency

- **WORD**: knowledge
  - **PART OF SPEECH**: noun
  - **HUMAN**: no
  - **CONCRETE**: no
  - **GERMAN**: Wissen: 80%, Kenntnisse: 20%
- **Probabilities can be derived from various machine learning techniques → to be discussed later.**

Dictionaries (cont.)

- contain (syntactic and semantic) restrictions it places on other words
  - e.g., Subcategorization information: give needs a giver, a person given to, and an object that is given
  - e.g., Selectional restrictions: if X is eating, then X must be animate.
- may also contain frequency information
- can be hierarchically organized, e.g.:
  - all nouns have person, number, and gender
  - verbs (unless irregular) conjugate in the past tense by adding ed.

Transformer approaches

- **Transformer** architectures transform example sentences from one language into another.
- They consist of
  - a grammar for the source/input language
  - a source-to-target language dictionary
  - source-to-target language rules
- Note that there is no grammar for the target language, only mappings from the source language.

Transformers: Less than meets the eye

- By their very nature, transformer systems are non-reversible because they lack a target language grammar.
- If we have a German to English translation system, for example, we are incapable of translating from English to German.
- However, as these systems do not require sophisticated knowledge of the target language, they are usually very robust = they will return a result for nearly any input sentence.

An example for the transformer approach

We’ll work through a German-to-English example.

(3) a. Drehen Sie den Knopf eine Position zurück.
    b. Turn the knob back one position.

1. Using the grammar, assign parts-of-speech:

(4) Drehen Sie den Knopf eine Position zurück.
    verb pron. article noun article noun prep.

2. Using the grammar, give the sentence a (basic) structure

(5) Drehen Sie [den Knopf] [eine Position] zurück.
Linguistic knowledge-based systems

- Linguistic knowledge-based systems include knowledge of both the source and the target languages.
- We will look at direct transfer systems and then the more specific instance of interlinguas.
  - Direct transfer systems
  - Interlinguas

Steps in a transfer system

1. source language grammar analyzes the input and puts it into an underlying representation (UR).
   Der Tisch gefällt Paul → Der Tisch gefallen Paul (source UR)
2. The transfer component relates this source language UR (German UR) to a target language UR (English UR).
   German UR → English UR
   X gefallen Y ↔ Eng(Y) like Eng(X) (where Eng(X) means the English translation of X)
   Der Tisch gefallen Paul (source UR) → Paul like the table, (target UR)
3. target language grammar translates the target language UR into an actual target language sentence.
   Paul like the table → Paul likes the table

Levels of abstraction

- There are differing levels of abstraction at which transfer can take place. So far we have looked at URs that represent only word information.
- We can do a full syntactic analysis, which helps us to know how the words in a sentence relate.
- Or we can do only a partial syntactic analysis, such as representing the dependencies between words.

Direct transfer systems

A direct transfer system consists of:

- A source language grammar
- A target language grammar
- Rules relating source language underlying representation to target language underlying representation

Things to note about transfer systems

- The transfer mechanism is essentially reversible; e.g., the *fallen* rule works in both directions (at least in theory)
- Because we have a separate target language grammar, we are able to ensure that the rules of English apply:
  like → likes.
- Word order is handled differently than with transformers; the URs are essentially unordered.
- The underlying representation can be of various levels of abstraction – words, syntactic trees, meaning representations, etc.; we will talk about this with the translation triangle.

Czech-English example

(8) Kaufman & Broad odmítla institucionální investory jménovat.

Kaufman & Broad refused the institutional investors to name/identify.

Example taken from Čmejrek, Cuřín, and Havelka (2003).

- They find the base forms of words (e.g., obmíditu ‘to decline’ instead of odmíta ‘declined’)
- They find which words depend on which other words and represent this in a tree (e.g., the noun *investory* depends on the verb *odmita*).
- This dependency tree is then converted to English (comparative grammar) and re-ordered as appropriate.

Direct transfer systems (cont.)

- A direct transfer system has a transfer component which relates a source language representation with a target language representation.
- This can also be called a comparative grammar.
- We’ll walk through the following French to English example:

(7) Der Tisch gefällt Paul.

The table is pleasing to Paul.

‘Paul likes the table.’

Caveat about reversibility

- It seems like reversible rules are highly desirable—and in general they are—but we may not always want reversible rules.
  - e.g., Dutch aanvangen should be translated into English as begin, but English begin should be translated into Dutch as beginnen.

Dependency tree for Czech-English example
We are not told what each word is translated as: use a sentence alignment. But how can we tell what a word is being translated as? For this, we need a way of learning from that data. Ideally, we could use a word alignment. But if it has already been done for us, it gives us good information about what a word's translation equivalent is.

A fine-grained interlingua can require extra (unnecessary) work: e.g., Japanese distinguishes older brother from younger brother, so we have to disambiguate English brother to put it into the interlingua. Then, if we translate into French, we have to ignore the disambiguation and simply translate it as frère, which simply means 'brother'.

Instead of trying to tell the MT system how we're going to translate, we might try a machine learning approach = the computer will learn how to translate based on example translations.

For this, we need examples of translations as training data, and a way of learning from that data.

We can look at how often a source language word is translated as a target language word, i.e., the frequency of a given translation, and choose the most frequent translation. But how can we tell what a word is being translated as? There are two different cases:

- We are told what each word is translated as: text alignment
- We are not told what each word is translated as: use a bag of words
Different word alignments

- One word can map to one word or to multiple words. Likewise, sometimes it is best for multiple words to align with multiple words.
- English-Hungarian examples:
  - one-to-one: well = jól
  - one-to-many: round – kör alakú
  - many-to-one: to play the guitar – gitárzik
  - many-to-many: even though = még ha ... is (‘even if ... also’)

Word alignment difficulties (cont.)

- Sometimes it is not even clear that word alignment is possible.
  - (9) Kati fotózó.
    Kati photographer
    ‘Kati is a photographer.’
- What does it align with?
  - In cases like this, a word can be mapped to a “null” element in the other language.

Calculating probabilities

- With word alignments, it is relatively easy to calculate probabilities.
  - e.g., What is the probability that run translates as rennen in German?
  1. Count up how many times run appears in the English part of your bi-text, e.g., 500 times.
  2. Out of all those times, count up how many times it was translated as (i.e., aligns with) rennen, e.g., 275 (out of 500) times.
  3. Divide to get a probability: 275/500 = 0.55, or 55%.

The “bag of words” method

- What if we’re not given word alignments?
  - How can we tell which English words are translated as which German words if we are only given an English text and a corresponding German text?
    - We can treat each sentence as a bag of words – unordered collection of words.
    - It word A appears in a sentence, then we will record all of the words in the corresponding sentence in the other language as appearing with it.

Example for bag of words method

<table>
<thead>
<tr>
<th>English</th>
<th>Hungarian</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>Ó</td>
</tr>
<tr>
<td>jól</td>
<td>beszél</td>
</tr>
<tr>
<td>speaks</td>
<td>magyarul</td>
</tr>
</tbody>
</table>

The idea is that, over thousands, or even millions, of sentences, He will tend to appear more often with Ó, speaks will appear with beszél, and so on.

Example for bag of words method (cont.)

So, for He in He speaks Hungarian well/Ö jól beszél magyarul, we do the following:

1. Count up the number of Hungarian words: 4.
2. Assign each word equal probability of translation: 1/4 = 0.25, or 25%.

If we also have He is a photographer.Ő fotóz., then for He, we do the following:

1. Count up the number of possible translation words: 4 from the first sentence, 2 from the second = 6 total.
2. Count up the number of times Ó is the translation = 2 times out of 6 = 1/3 = 0.33, or 33%.

Every other word has the probability 1/6 = 0.17, or 17%, so Ön is clearly the best translation for Ó.

What makes MT hard?

We’ve seen how MT systems can work, but MT is a very difficult task because languages are vastly different. They differ:

- Lexically: In the words they use
- Syntactically: In the constructions they allow
- Semantically: In the way meanings work
- Pragmatically: In what speakers take from a sentence.

In addition, there is a good deal of real-world knowledge that goes into a translation.

References
Lexical ambiguity

Words can be **lexically ambiguous** = have multiple meanings.

- bank can be a financial institution or a place along a river,
- can be a cylindrical object, as well as the act of putting something into that cylinder (e.g., John cans tuna), as well as being a word like must, might, or should.

⇒ We have to know which meaning before we translate.

### Hyponyms and Hypernyms

- **English hypernyms** = words that are more general in English than in their counterparts in other languages
  - English know is rendered by the French savoir (‘to know a fact’) and connaître (‘to know a thing’)
  - English library is German Bücherei if it is open to the public, but Bibliothek if it is intended for scholarly work.
- **English hyponyms** = words that are more specific in English than in their foreign language counterparts
  - The German word Berg can mean either hill or mountain in English.
  - The Hungarian word lab can mean either leg or foot.

### Lexical gaps

Sometimes there is no simple equivalent for a word in a language, and the word has to be translated with a more complex phrase. We call this a **lexical gap or lexical hole**.

- French gratiner means something like ‘to cook with a coating of bread crumbs and cheese’
- Hebrew stam means something like ‘I’m just kidding’ or ‘Nothing special.’

### How words divide up the world (lexical issues)

Words don’t line up exactly between languages. Within a language, we have synonyms, hyponyms, and hypernyms.

- *sofa* and *couch* are synonyms (mean the same thing)
- *sofa* is a hyponym (more specific term) of *furniture*
- *furniture* is a hypernym (more general term) of *sofa*

### Semantic overlap

And then there’s just fuzziness, as in the following English and French correspondences

- *leg* = etape (journey), jambe (human), pied (chair), patte (animal)
- *foot* = pied (human), patte (bird)
- *paw* = patte (animal)

### Light verbs

Some verbs carry little meaning, so-called light verbs

- French faire une promenade is literally ‘make a walk;’ but it has the meaning of the English take a walk
- Dutch een poging doen ‘do an attempt’ means the same as the English make an attempt

### Synonyms

Often we find **synonyms** between two languages (as much as there are synonyms within a language):

- English *book* = Hungarian *könyv*
- English *music* = German *Musik*

But words don’t always line up exactly between languages.

### Idioms

And we often face **idioms** = expressions whose meaning is not made up of the meanings of the individual words.

- e.g., English kick the bucket
  - approximately equivalent to the German ins Gras beießen (‘bite into the grass’)
  - but we might want to translate it as sterben (‘die’)
  - and we want to treat it differently than kick the table
There are idiosyncratic choices among languages, e.g.:

- English *heavy smoker*
- French *grand fumeur* (‘large smoker’)
- German *starker Raucher* (‘strong smoker’)

How fast is the MT system?

How fast is the MT system (Quality)?

http://www.aija.org.au/online/

e.g., The word ‘skin’ is taboo in a Western Australian aboriginal language.

We know what it refers to only because we know that printers, not paper, can be switched on.

Language and Computers
Topic 5: Machine Translation
Introduction
Examples for Translations
Structure and word order differences

There are taboo words = words which are “forbidden” in some way or in some circumstances (i.e., swear/curse words)

- You, of course, know several English examples. Note that the literal meanings of these words lack the emotive impact of the actual words.
- Other languages/cultures have different taboos: often revolving around death, body parts, bodily functions, disease, and religion.
  - e.g., The word ‘skin’ is taboo in a Western Australian (Aboriginal) language (http://www.aija.org.au/online/ICA/Benchbook/BenchbookChapter5.pdf)
  - Imagine encountering the word ‘skin’ in English and translating it without knowing this.

How language is used (Pragmatics)

Translation becomes even more difficult when we try to translate something in context.

- Thank you is usually translated as *merci* in French, but it is translated as *s’il vous plaît* ‘please’ when responding to an offer.
- Can you drive a stick-shift? could be a request for you to drive my manual transmission automobile, or it could simply be a request for information about your driving abilities.

Evaluating MT systems

We’ve seen some translation systems and we know that translation is hard.

- The question now is: How do we evaluate MT systems, in particular for use in large corporations as likely users?
  - How much change in the current setup will the MT system force?
  - Translator tasks will change from translation to updating the MT dictionaries and post-editing the results.
  - How will it fit in with word processors and other software?
  - Will the company selling the MT system be around in the next few years for support and updates?
  - How fast is the MT system?
  - How good is the MT system (Quality)?
Intelligibility

Intelligibility Scale (from Arnold et al., 1994)

1. The sentence is perfectly clear and intelligible. It is grammatical and reads like ordinary text.

2. The sentence is generally clear and intelligible. Despite some inaccuracies or infelicities of the sentence, one can understand (almost) immediately what it means.

3. The general idea of the sentence is intelligible only after considerable study. The sentence contains grammatical errors and/or poor word choices.

4. The sentence is unintelligible. Studying the meaning of the sentence is hopeless; even allowing for context, one feels that guessing would be too unreliable.

Further reading

Some of the examples are adapted from the following books:
