MACHINE TRANSLATION

MACHINE TRANSLATION (MT)

- Machine translation is the automatic translation of text from one natural language (the source) to another (the target).
- **Translation is difficult** - it requires in-depth understanding of the text.
- Consider the word "Open" on the door of a store.
- It communicates the idea that the store is accepting customers at the moment.[
- Now consider the same word "Open" on a large banner outside a newly constructed store.
- It means that the store is now in daily operation, but readers of this sign would not feel misled if the store closed at night without removing the banner.
- The two signs use the identical word to convey different meanings.

MACHINE TRANSLATION (MT)

- Machine translation is the automatic translation of text from one natural language (the source) to another (the target).
- A translator (human or machine) often needs to understand the actual situation described in the source, not just the individual words.

- Three main applications of machine translation.
- *Rough translation*, as provided by free online services, gives the “gist” of a foreign sentence or document, but contains errors.
- *Pre-edited translation* is used by companies to publish their documentation and sales materials in multiple languages.
- The original source text is written in a constrained language that is easier to translate automatically, and the results are usually edited by a human to correct any errors.
- *Restricted-source translation* works fully automatically, but only on highly stereotypical language, such as a weather report.

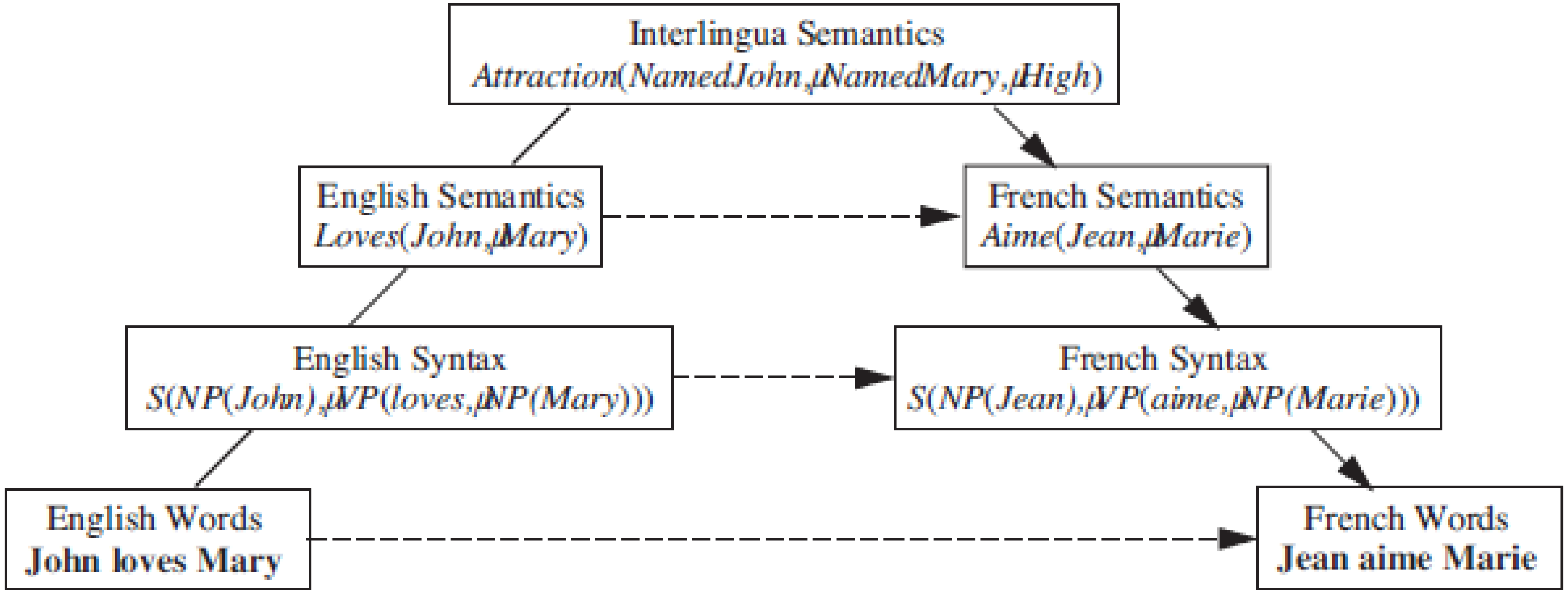
- **Machine translation systems**
- **Statistical machine translation**

Machine Translation Systems

- All translation systems must model the source and target languages, but systems vary in the type of models they use.
- Some systems attempt to analyze the source language text all the way into an interlingua knowledge representation and then
- generate sentences in the target language from that representation.
- This is difficult because it involves three unsolved problems:
    - creating a complete knowledge representation of everything;
    - parsing into that representation;
    - generating sentences from that representation.

Machine Translation Systems - Transfer Model

- Keep a **database of translation rules (or examples),** and whenever the rule (or example) matches, they translate directly.
- Transfer can occur at the lexical, syntactic, or semantic level.
- For example, transfer English to French
- **A strictly syntactic rule maps**
- English [*Adjective Noun*] to French [*Noun Adjective*].
- **A mixed syntactic and lexical rule maps**
- French [S1 “et puis” S2] to English [S1 “and then” S2].



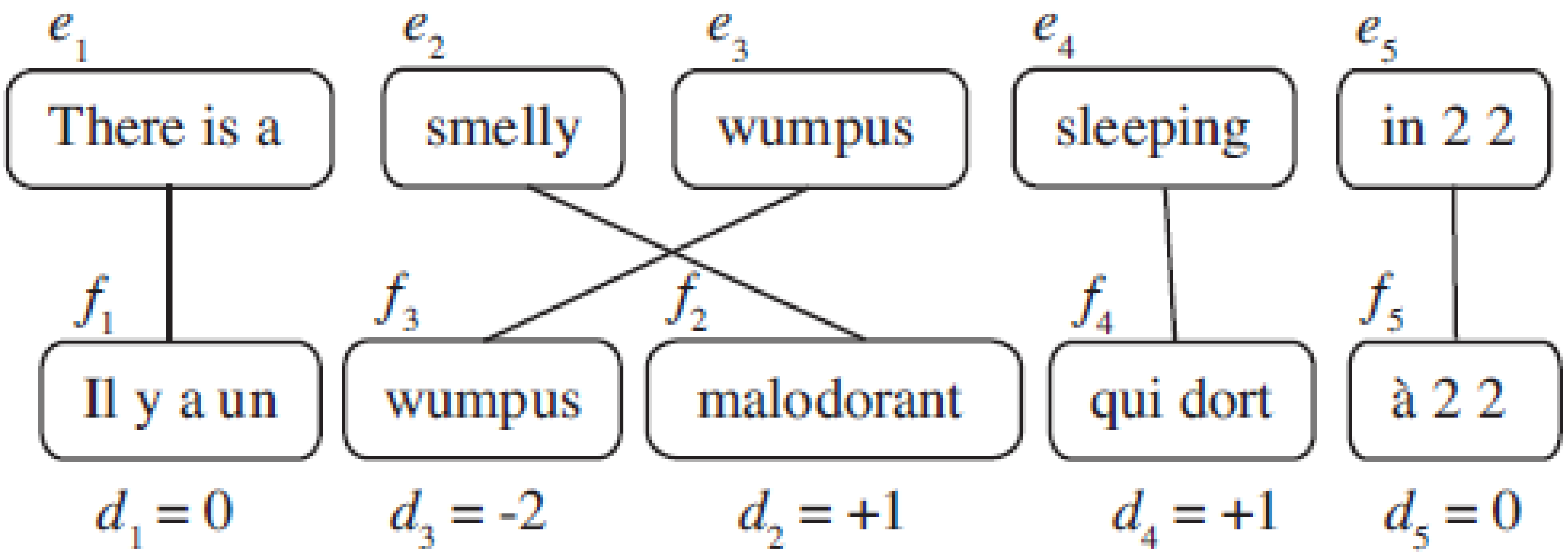
- The Vauquois triangle: schematic diagram of the choices for a machine translation system (Vauquois, 1968).
- Start with English text at the top.
- An interlingua based system follows the solid lines, parsing English first into a syntactic form, then into a semantic representation and an interlingua representation, and then through generation to a semantic, syntactic, and lexical form in French.
- A transfer-based system uses the dashed lines as a shortcut.
- Different systems make the transfer at different points; some make it at multiple points.

Statistical Machine Translation

- English sentence, e, finding a French translation f is a matter of three steps:
- 1. **Break the English sentence into phrases** e1, . . . , en.
- 2. **For each phrase ei, choose a corresponding French phrase fi.**
- We use the notation P(fi | ei) for the phrasal probability that fi is a translation of ei.
- 3. **Choose a permutation of the phrases f1, . . . , fn**.
- For each fi, we choose a **distortion** di, (misrepresentation) which is the number of words that phrase fi has moved with respect to fi−1;
- positive for moving to the right, negative for moving to the left, and zero if fi immediately follows fi−1.

Statistical Machine Translation - The procedure

- 1. **Find parallel texts**:
- 2. **Segment into sentences**: The unit of translation is a sentence, so we will have to break the corpus into sentences. Periods are strong indicators of the end of a sentence,
- 3. **Align sentences**: For each sentence in the English version, determine what sentence(s) it corresponds to in the French version.
- 4. **Align phrases**: Within a sentence, phrases can be aligned by a process, that is similar to that used for sentence alignment.
- 5. **Extract distortions (misrepresentation)**: count how often distortion occurs in the corpus for each distance $d = 0, \pm 1, \pm 2, \ldots$ and apply smoothing.
- 6. **Improve estimates with EM (**expectation–maximization**):** used to improve the estimates of $P(f \mid e)$ and $P(d)$ values.
- E Step: compute the best alignments with the current values of these parameters,
- M step : update the estimates and iterate the process until convergence.



- Candidate French phrases for each phrase of an English sentence, with distortion (d) values for each French phrase.

- the procedure
- 1. **Find parallel texts**:
- 2. **Segment into sentences**:
- 3. **Align sentences**:
- 4. **Align phrases**:
- 5. **Extract distortions**:
- 6. **Improve estimates with EM (**expectation–maximization**)**:

SPEECH RECOGNITION

- **Speech recognition** is the task of identifying a sequence of words spoken by a speaker, given the acoustic (audio) signal.
- It has become one of the mainstream applications of AI—millions of people interact with speech recognition systems every day
  - to navigate voice mail systems,
  - search the Web from mobile phones, and
  - Voice-text conversion and other applications.
- Speech is an attractive option when hands-free operation is necessary, as when operating machinery.

- Speech recognition is difficult because the sounds made by a speaker are ambiguous and sometimes noisy.
- As a well-known example, the phrase “recognize speech” sounds almost the same as “wreck a nice beach” when spoken quickly.

The issues in Speech recognition

- **Segmentation**: written words in English have spaces between them, but in fast speech there are no pauses.
- In “wreck a nice” that would distinguish it as a multiword phrase as opposed to the single word “recognize.”
- **Coarticulation**: when speaking quickly the “s” sound at the end of “nice” merges with the “b” sound at the beginning of “beach,” yielding something that is close to a “sp.”
- **Homophones**—words like “to,” “too,” and “two” that sound the same but differ in meaning.

- Sequential process, sequence of state variables, **x**1:t, given a sequence of observations **e**1:t.
- In this case the state variables are the words, and the observations are sounds.
- More precisely, an observation is a vector of features extracted from the audio signal.
- the most likely sequence can be computed with the help of Bayes' rule to be:

$$\underset{word_{1:t}}{\operatorname{argmax}} P(word_{1:t} \mid sound_{1:t}) = \underset{word_{1:t}}{\operatorname{argmax}} P(sound_{1:t} \mid word_{1:t}) P(word_{1:t})$$

$$\underset{word_{1:t}}{\operatorname{argmax}} P(word_{1:t} \mid sound_{1:t}) = \underset{word_{1:t}}{\operatorname{argmax}} P(sound_{1:t} \mid word_{1:t}) P(word_{1:t})$$

- $P(sound_{1:t} | word_{1:t})$ is the **acoustic model**.
- It describes the sounds of words—
- “ceiling” begins with a soft “c” and sounds the same as “sealing.”
- P(word 1:t) is known as the **language model.**
- It specifies the prior probability of each utterance—

The **Noisy Channel Model**

- "ceiling fan" is about 500 times more likely as a word sequence than "sealing fan", this approach was named the **noisy channel model.**
- **a situation in which an original message (the *words* in our example) is transmitted over a noisy channel (such as a telephone line) such that a corrupted message (the *sounds* in our example) are received at the other end.**
- it is possible to recover the original message with arbitrarily small error, if we encode the original message in a redundant enough way.
- Applications - speech recognition, machine translation, spelling correction, and other tasks.

Acoustic models

- Sound waves are periodic changes in pressure that propagate through the air.
- When these waves strike the diaphragm of a microphone, the back-and-forth movement generates an electric current.
- An analog-to-digital converter measures the size of the current—which approximates the amplitude of the sound wave—at discrete intervals called the **sampling rate**.
- Speech sounds, which are mostly in the range of 100 Hz (100 cycles per second) to 1000 Hz, are typically sampled at a rate of 8 kHz. (CDs and mp3 files are sampled at 44.1 kHz.)
- The precision of each measurement is determined by the **quantization factor**; speech recognizers typically keep 8 to 12 bits.
- That means that a low-end system, sampling at 8 kHz with 8-bit quantization, would require nearly half a megabyte per minute of speech.

- Since we only want to know what words were spoken, not exactly what they sounded like, we don’t need to keep all that information.
- We only need to distinguish between different speech sounds.
- Linguists have identified about 100 speech sounds, or **phones**, that can be composed to form all the words in all known human languages.
- Roughly speaking, a phone is the sound that corresponds to a single vowel or consonant, but there are some complications:
- combinations of letters, such as “th” and “ng” produce single phones, and some letters produce different phones in different contexts

- The ARPA (Advanced Research Projects Agency) phonetic alphabet, or **ARPAbet**, listing all the phones used in American English.
- There are several alternative notations, including an International Phonetic Alphabet (IPA), which contains the phones in all known languages.

Vowels		Consonants B–N		Consonants P–Z	
Phone	Example	Phone	Example	Phone	Example
[iy]	beat	[b]	bet	[p]	pet
[ih]	bit	[ch]	Chet	[r]	rat
[eh]	bet	[d]	debt	[s]	set
[æ]	bat	[f]	fat	[sh]	shoe
[ah]	but	[g]	get	[t]	ten
[ao]	bought	[hh]	hat	[th]	thick
[ow]	boat	[hv]	high	[dh]	that
[uh]	book	[jh]	jet	[dx]	butter
[ey]	bait	[k]	kick	[v]	vet
[er]	Bert	[l]	let	[w]	wet
[ay]	buy	[el]	bottle	[wh]	which
[oy]	boy	[m]	met	[y]	yet
[axr]	diner	[em]	bottom	[z]	zoo
[aw]	down	[n]	net	[zh]	measure
[ax]	about	[en]	button		
[ix]	roses	[ng]	sing		
[aa]	cot	[eng]	washing	[-]	*silence*

- To represent spoken English we want a representation that can distinguish between different phonemes, but one that need not distinguish the nonphonemic variations in sound: loud or soft, fast or slow, male or female voice, etc.
- speech systems summarize the properties of the signal over time slices called **frames**.
- short-duration phenomena will be missed
- Overlapping frames are used to make sure that we don’t miss a signal because it happens to fall on a frame boundary.

Thank You