Machine Translation: Summer Term 2021

Ex1: Introduction; Why is MT hard?; RBMT

Sangeet Sagar (7009050)
sasa00001@stud.uni-saarland.de

May 1, 2021

1. **In your own words, why is MT hard?**
   Following points constitute to the complexities in MT-

   - Words with multiple meaning.
   - Sentences with multiple grammatical structures.
   - Uncertainty about what a pronoun refers to.
   - Grammar problem.

   E.g *American history teacher*

2. **Please give at least two interpretations for each of the following (examples from Chris Manning)**

   - Minister Accused Of Having 8 Wives In Jail
     - The minister was accused of having a relationship with 8 women in the prison.
     - The minister was accused of reproducing (used as a verb of "having") 8 women in the prison.
   - Juvenile Court to Try Shooting Defendant
     Structural ambiguity.
     - Juvenile court will put the shooting defendant on trial.
     - Juvenile court will attempt to encounter or shoot the defendant (and probably kill them)
   - Teacher Strikes Idle Kids
     A lexical ambiguity.
     - Walkout of teachers has left the kids idle, with no one to take classes.
     - Teaches hits (or strike) idle (used as a verb) kids.
   - Miners refuse to work after death
     - Miners refused to work after cases of accidents resulting to death was reported.
     - Miners refuse to work after being dead (they can not work in afterlife).
   - Local High School Dropouts Cut in Half
     - Local high school students chopped in half (probably dead).
     - The proportion of local high school student dropouts reduced to half (may be in comparison to last year)
   - Red Tape Holds Up New Bridges
     - Red tape refers to the government rules and regulations that make has made the task of building of bridges difficult.
     - Red tape or a thin red coloured paper is holding up or supporting a bridge.
   - Clinton Wins on Budget, but More Lies Ahead
     - Clinton was successful on the budget, but more challenges lies ahead that need to be dealt with.
     - Clinton was successful on the budget, but more deceit or lie yet to come.

- Hospitals Are Sued by 7 Foot Doctors
  - Hospitals are sued by doctors of height 7 foot.
  - Hospitals are sued by 7 doctors that specialize in foot or simply Podiatrist.
- Police: Crack Found in Man's Buttocks
  - Drugs found in man's buttocks.
  - Crack between butt-cheeks in the man's buttocks.

3. **Give an example of lexical ambiguity and an example of structural ambiguity**
   Lexical ambiguity occurs when a single word poses multiple meanings. While structural ambiguity is the presence of two or more possible meanings within a single sentence.
   - **Lexical ambiguity**: *Red Tape Holds Up New Bridges*
   - **Structural ambiguity**: *Juvenile Court to Try Shooting Defendant*
4. **Can you bracket the following in different ways to show how they are ambiguous**
   - Chinese history teacher
     - Chinese [history teacher]: history teacher who is Chinese.
     - [Chinese history] teacher: A teacher who teaches Chinese history.
   - Unlockable
     - [Un]lockable: not able to be locked
     - [Unlock]able: able to be unlocked
   - She saw the boy with the telescope
     - She saw the boy with the [telescope]: She saw the boy using a telescope.
     - She saw the [boy with the telescope]: She saw the boy holding a telescope.
   - Visiting relatives can be dangerous
     - [Visiting] relatives can be dangerous
     - [Visiting relatives] can be dangerous
5. **What is "culture translation"?**
   Cultural translation is the practice of translation while respecting and showing cultural differences. This kind of translation solves some issues linked to cultures, such as dialects, food, or architecture. [Source: Wikipedia]
6. **What is a pro-drop language?**
   Languages like Spanish, Chinese, Japanese etc. are pro-drop languages because they can drop pronouns e.g. *he*, *she*, *it* to form grammatically meaningful sentences. These languages are allowed to drop the pronouns but to a different extent.
7. **What are SVO, SOV, and VSO languages? How many possible orderings of S,V, O are there?**
   SVO (subject-verb-object) languages have subject followed by a verb followed by an object and these make use of pre-positions e.g. English. SOV languages have modifiers before heads and use postpositions. VSO languages tend to place modifiers after their heads and use prepositions. [Source: Wikipedia]
8. **In your own words, describe the "Vauquois Triangle" for rule-based MT (RBMT) below:**

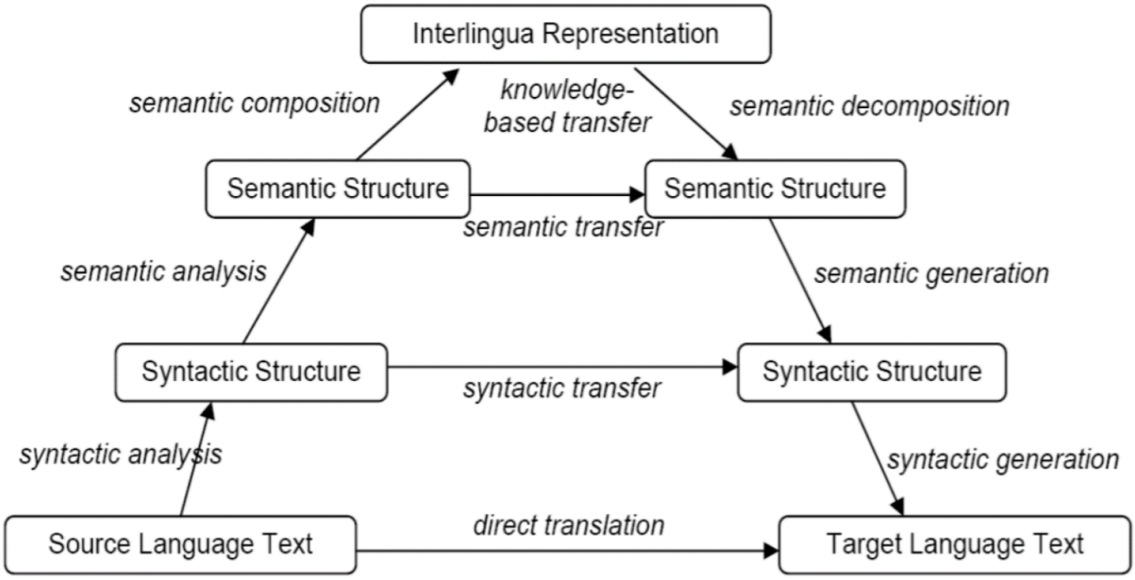


Figure 1: The Vauquois triangle

Vauquois conceptualized the space of Rule-based Machine Translation (RBMT was a dominant paradigm until the late 1908s-1990s) in the above Vauquois triangle. The translation is always from some **Source Language Text** and it goes to **Target Language text** which is called as *direct translation*. It focuses on word-to-word translation with very little focus on word reordering and the context.

To handle more complicated text, **Syntactic structure** gained popularity. The source text is automatically parsed into syntactic representation. We need rules for parsing source text into a syntactic structure. And then we need another rule set that takes such a representation of **Syntactic Structure** of the source text and performs a *syntactic transfer* into the corresponding syntactic structure of the target text such that the meaning of the source is preserved. The leaves of the resulting syntactic structure are used to get the sequence of translated words into a sentence in the target language.

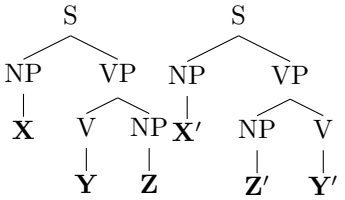
Going up the Vauquois triangle, we have **Semantic Structure**. We parse the source language to its semantic structure via syntactic structure. For this, we have *semantic transfer* rule to transfer the semantic structure for source language into the target language with the meaning preserved. Then we use the target language grammars and form a semantic structure representation. Step by step we further generate the target-language text.

We also note that the semantic transfer rules are simpler than the syntactic transfer rules. Higher up we go onto the Vauquois triangle, closer to the meaning we get, less work we do for transfer, and more closer the semantics get. Theoretically, the higher the triangle, the less cost the analysis and synthesis becomes.

9. **Design a set of direct translation rules that translate:**

   - an English garden
   - ein englischer Garten
   - an English garden
   - un jardin anglais

10. **Design a set of analysis, transfer, and generation rules (incl. lexical rules) based on CFG representations that translate. It may help to draw them as CFG trees.**



- Rule set I
  - $S \rightarrow NP \quad VP$
  - $VP \rightarrow V \quad NP$
- Rule set II
  - $S \rightarrow NP \quad VP$
  - $VP \rightarrow NP \quad V$

11. **How many transfer-based MT systems do you need to translate between n languages?** $n(n-1)$

12. **What is interlingua-based MT? Interlingua Representation** lies at the top of the Vauquois triangle (Figure 1). In interlingua-based MT, the source language is translated into a language-independent abstract meaning representation such that it is the same for all people. The target language is further generated from the interlingua representation.

13. **In your own words, what are the pros and cons of RBMT?**

    - **Pros**:
      - Can achieve high accuracy on small dataset.
      - Excellent performance over closed dataset.
      - No bilingual texts are required.
    - **Cons**:
      - Large number of rules leading to a complex system.
      - Requires expert linguistic knowledge to frame such rules
      - Very local view of translation i.e takes context of current word and may be nearby words.
      - Long range context hard to capture