Transfer Methods for Machine Translation

11-731
Machine Translation
Alon Lavie
February 5, 2007

Direct Approaches

- No intermediate representation stages in the translation
- First MT systems developed in the 1950's-60's (assembly code programs)
  - Morphology, bi-lingual dictionary lookup, local reordering rules
  - "Word-for-word, with some local word-order adjustments"
- Modern Approaches: EBMT and SMT

Analysis and Generation

Main Steps

- **Analysis:**
  - Morphological analysis (word-level) and POS tagging
  - Syntactic analysis and disambiguation (produce syntactic parse-tree)
  - Semantic analysis and disambiguation (produce symbolic frames or logical form representation)
  - Map to language-independent Interlingua
- **Generation:**
  - Generate semantic representation in TL
  - Sentence Planning: generate syntactic structure and lexical selections for concepts
  - Surface-form realization: generate correct forms of words

Transfer Approaches

- **Syntactic Transfer:**
  - Analyze SL input sentence to its syntactic structure (parse tree)
  - Transfer SL parse-tree to TL parse-tree (various formalisms for specifying mappings)
  - Generate TL sentence from the TL parse-tree
- **Semantic Transfer:**
  - Analyze SL input to a language-specific semantic representation (i.e., Case Frames, Logical Form)
  - Transfer SL semantic representation to TL semantic representation
  - Generate syntactic structure and then surface sentence in the TL

Interlingua versus Transfer

- With interlingua, need only N parsers/generators instead of N² transfer systems.
Transfer Approach

- **Language-dependent** intermediate representations
- Disadvantage: costly as number of languages grows
  - \( O(n^2) \); specifically, \( n \times (n - 1) \)
- \( n \) Analyzers, \( n \) Generators, plus \( n \times (n - 1) \) transfer components

Transfer System Dataflow

From Hutchins & Somers

![Transfer System Dataflow Diagram](image)

Advantages of Interlingua

- Add a new language easily
  - get all-ways translation to all previous languages by adding one module for analysis and one module for generation
- Mono-lingual development teams.
- Paraphrase
  - Generate a new source language sentence from the interlingua so that the user can confirm the meaning

Disadvantages of Interlingua

- "Meaning" is arbitrarily deep.
  - What level of detail do you stop at?
- If it is too simple, meaning will be lost in translation.
- If it is too complex, analysis and generation will be too difficult.
- Should be applicable to all languages
  - how do we ensure that?
- Human development time.

Transfer Approaches

Main Advantages and Disadvantages:

- **Syntactic Transfer**:
  - No need for semantic analysis and generation
  - Syntactic structures are general, not domain specific
  - Less domain dependent, can handle open domains
  - Requires word translation lexicon

- **Semantic Transfer**:
  - Requires deeper analysis and generation, symbolic representation of concepts and predicates
  - difficult to construct for open or unlimited domains
  - Can better handle non-compositional meaning structures
  - can be more accurate
  - No word translation lexicon – generate in TL from symbolic concepts
Major Sources of Translation Problems and Divergences

- **Lexical Differences:**
  - Multiple possible translations for SL word, or difficulties expressing SL word meaning in a single TL word

- **Structural Differences:**
  - Syntax of SL is different than syntax of the TL: word order, sentence and constituent structure

- **Differences in Mappings of Syntax to Semantics:**
  - Meaning in TL is conveyed using a different syntactic structure than in the SL

- **Idioms and Constructions**

Lexical Differences

- **Lexical Differences:**
  - SL word has several different meanings, that translate differently into TL
    - Ex: financial bank vs. river bank
  - **Lexical Gaps:** SL word reflects a unique meaning that cannot be expressed by a single word in TL
    - Ex: English *snub* doesn’t have a corresponding verb in French or German
  - TL has finer distinctions than SL.
    - SL word should be translated differently in different contexts
    - Ex: English *wall* can be German *wand* (internal), *mauer* (external)

Lexical Differences

- **Lexical Gaps:**
  - Examples: these have no direct equivalent in English:
    - *gratiner* (v., French, “to cook with a cheese coating”)
    - *ōtosanrin* (n., Japanese, “three-wheeled truck or van”)

[From Hutchins & Somers]

MT Handling of Lexical Differences

- **Direct MT and Syntactic Transfer:**
  - Lexical Transfer stage uses bilingual lexicon
  - SL word can have multiple translation entries, possibly augmented with disambiguation features or probabilities
  - Lexical Transfer can involve use of limited context (on SL side, TL side, or both)
  - Lexical Gaps can partly be addressed via phrasal lexicons

- **Semantic Transfer:**
  - Ambiguity of SL word must be resolved during analysis to get the correct symbolic representation at semantic level
  - TL Generator must select appropriate word or structure for correctly conveying the concept in TL

Semantic Transfer:

- Antiquity of SL word must be resolved during analysis to get the correct symbolic representation at semantic level

Structural Differences

- **Syntax of SL is different than syntax of the TL:**
  - Word order within constituents:
    - English NPs: *art adj n* (the big boy)
    - Hebrew NPs: *art n adj * (he yeled ha gadol)
  - Constituent structure:
    - English is SVO: Subj Verb Obj (I saw the man)
    - Modern Arabic is VSO: Verb Subj Obj (I saw the man)
  - Different verb syntax:
    - Verb complexes in English vs. in German
      - I can eat the apple vs. Du kannst die Apfel essen
    - Case marking and free constituent order
    - German and other languages that mark case:
      - *drei apfel* (three apples) vs. *drei apple* (three apples)
MT Handling of Structural Differences

- **Direct MT Approaches:**
  - No explicit treatment: Phrasal Lexicons and sentence level matches or templates
- **Syntactic Transfer:**
  - Structural Transfer Grammars
    - Trigger rule by matching against syntactic structure on SL side
    - Rule specifies how to reorder and re-structure the syntactic constituents to reflect syntax of TL side
  - Semantic Transfer:
    - SL Semantic Representation abstracts away from SL syntax to functional roles done during analysis
    - TL Generation maps semantic structures to correct TL syntax

Syntax-to-Semantics Differences

- **Structure-change example:**
  
  I like swimming
  "Ich schwimme gern"
  I swim gladly

- **Verb-argument example:**
  
  Jones likes the film.
  "Le film plait à Jones."
  (lit: “the film pleases to Jones”)
  - Use of case roles can eliminate the need for this type of transfer
    - Jones = Experiencer
    - film = Theme

- **Passive Constructions**

  Example: French reflexive passives:
  
  Ces livres se lisent facilement
  “These books read themselves easily”
  These books are easily read

Same intention, different syntax

- rigly bitiwgacy
  my leg hurts
- candy wagac fe rigly
  I have pain in my leg
- rigly bitiClimny
  my leg hurts
- fe wagac fe rigly
  there is pain in my leg
- rigly bitiqaH calya
  my leg bothers on me

Romanization of Arabic from CallHome Egypt.
MT Handling of Syntax-to-Semantics Differences

- Direct MT Approaches:
  - No Explicit treatment: phrasal lexicons and sentence level matches or templates
- Syntactic Transfer:
  - "Lexicalized" Structural Transfer Grammars
  - Triggers rule by matching against "lexicalized" syntactic structure on SL side: **lexical and functional features**
  - Rule specifies how to reorder and re-structure the syntactic constituents to reflect syntax of TL side
- Semantic Transfer:
  - SL Semantic Representation abstracts away from SL syntax to functional roles done during analysis
  - TL Generation maps semantic structures to correct TL syntax

Example of Structural Transfer Rule (verb-argument) [From Hutchins & Somers]

Semantic Transfer: Theta Structure (case roles) [From Hutchins & Somers]

- Abstracts away from grammatical functions
- Looks more like a "semantic F-structure"
- The basis for "semantic transfer"

Idioms and Constructions

- Main Distinction: meaning of whole is not directly compositional from meaning of its sub-parts \(\rightarrow\) no compositional translation
- Examples:
  - George is a bull in a china shop
  - He kicked the bucket
  - Can you please open the window?

Formulaic Utterances

- Good night.
- tisbah cala x Er
- waking up on good
- Romanization of Arabic from CallHome Egypt

Constructions

- Identifying speaker intention rather than literal meaning for formulaic and task-oriented sentences.
  - How about ... suggestion
  - Why don’t you... suggestion
  - Could you tell me... request info.
  - I was wondering... request info.
MT Handling of Constructions and Idioms

- Direct MT Approaches:
  - No Explicit treatment: Phrasal Lexicons and sentence level matches or templates

- Syntactic Transfer:
  - No effective treatment
  - "Highly Lexicalized" Structural Transfer rules can handle some constructions
  - Trigger rule by matching against entire construction, including structure on TL side
  - Rule specifies how to generate the correct construction on the TL side

- Semantic Transfer:
  - Analysis must capture non-compositional representation of the idiom or construction → specialized rules
  - TL Generation maps construction semantic structures to correct TL syntax and lexical words

Transfer-based MT Systems

- Primarily Syntactic-transfer, based on large manually developed transfer grammars
- Most notable systems:
  - SYSTRAN translation engines
  - PAHO system (Spanish/English)
  - EUROTRA
  - VERBMOBIL
- Main Issues:
  - Large volume and complexity of transfer grammars
  - Interaction between "general" and "exception" rules
  - Interaction between transfer grammar and lexicon

New Transfer-based Approaches

- Syntax-based approaches suitable for automatic transfer-rule acquisition from parallel data are at forefront of current MT research
- Combine ideas from syntactic transfer with statistical modeling techniques
- Synchronous Context-Free Grammars (S-CFGs):
  - Formalism for simultaneously describing the syntactic structure of two languages and the correspondence between them
  - Common set of categories (non-terminals)
  - Rules specify how to derive the category in each language
  - Example: NP: \[DET \text{ adj } n\] → \[DET n \text{ adj}\]

Example: The CMU XFER MT Framework

- Transfer-based formalism designed to support both manual rule development as well as automatic acquisition of rules from data
- Formalism is an extended SCFG, augmented with feature-structure unification constraints
- Runtime XFER engine performs interleaved bottom-up parsing, transfer and generation with constituents augmented with feature-structures
- XFER engine constructs a translation lattice
- Decoder constructs complete translations from the lattice

Transfer Rule Formalism

- Type information: NP, VP, etc.
- Part-of-speech/constituent information: \[DET \text{ adj } n\], \[DET n \text{ adj}\]
- Alignments: \[X1::Y1], \[X1::Y3], \[X2::Y4], \[X3::Y2\]
- Value constraints:
  - Agreement constraints: ((Y1 AGR) = *3-SING), ((Y1 DEF) = *DEF), ((Y2 AGR) = *3-SING), ((Y2 GENDER) = (Y4 GENDER))

Transfer Rule Formalism (II)

- Type information: NP, VP, etc.
- Part-of-speech/constituent information: \[DET \text{ adj } n\], \[DET n \text{ adj}\]
- Alignments: \[X1::Y1], \[X1::Y3], \[X2::Y4], \[X3::Y2\]
- Value constraints:
  - Agreement constraints: ((Y1 AGR) = *3-SING), ((Y1 DEF) = *DEF), ((Y2 AGR) = *3-SING), ((Y2 GENDER) = (Y4 GENDER))
The Transfer Engine

• Some Unique Features:
  – Works with either learned or manually-developed transfer grammars
  – Handles rules with or without unification constraints
  – Supports interfacing with servers for morphological analysis and generation
  – Can handle ambiguous source-word analyses and/or SL segmentations
    represented in the form of lattice structures

The Lattice Decoder

• Simple Stack Decoder, similar in principle to SMT/EBMT decoders
• Searches for best-scoring path of non-overlapping lattice arcs
• Scoring based on log-linear combination of scoring components (no MER training yet)
• Scoring components:
  – Standard trigram LM
  – Fragmentation: how many arcs to cover the entire translation?
  – Length Penalty
  – Rule Scores (not fully integrated yet)
Decoder Output: Hebrew Example

0 4 REMNANTS OF WALLS FROM THE ANCIENT BRONZE PERIOD

Overall: -3.23212, Prob: -23.857, Rules: 0, Frag: 0.125, Length: 0, Words: 8,8

143 < 0 4 -10.8107: $RID QIR (NNP,1 (NP0,0 (N,0 'REMNANTS'))) (LITERAL 'OF') (NP,0 (NP,2 (NP,1 (NP0,0 (N,9 'WALLS'))))))>

291 < 4 12 -13.8587: M TQWH H BRWNZH H QDWM (PP,0 (PREP,0 'FROM') (NP,2 (LITERAL 'THE')) (NP,0 (NP,1 (ADJ,1 (ADJ,4 'ANCIENT'))) (NP1,0 (NN,3 (NP0,1 (N,28 'BRONZE'))) (NP0,0 (N,15 'PERIOD')))))>

Summary

- Distinction between syntactic and semantic transfer approaches
- Syntactic transfer can best handle lexical and syntactic divergences, and partially handle other types
- Transfer-based approaches are having a "come-back" – automatic acquisition of transfer rules from parallel data, in conjunction with modern statistical modeling techniques

Questions?

Dataflow in Direct MT

Figure 4.1 Direct MT system

[From Hutchins & Somers]

Morphological Structure

[From Hutchins & Somers]

Interlingua-based MT

- The "natural" deep Artificial Intelligence approach:
  - Analyze the source language into a language independent detailed symbolic representation of its meaning
  - Generate this meaning in the target language
- "Interlingua": one single meaning representation for all languages
  - Nice in theory, but extremely difficult in practice
What is an Interlingua?

- Representation of meaning or speaker intention.
- Sentences that are equivalent for the translation task should have the same interlingua representation.

The room costs 100 Euros per night.
The room is 100 Euros per night.
The price of the room is 100 Euros per night.

Examples of Learned Rules (Hindi-to-English)

- [NP-14444]; Score:0.0040
  NP::NP [ADJ CONJ ADJ N] -> [ADJ CONJ ADJ N]
  (X1::Y1) (X2::Y2) (X3::Y3) (X4::Y4)

- [NP-14448]; Score:0.0429
  NP::NP [N] -> [DET N]
  (X1::Y1)

- [PP-4894]; Score:0.0470
  PP::PP [NP POSTP] -> [PREP NP]
  (X2::Y2) (X1::Y1)