Lexical Mismatches & Structural Mismatches (1)

Homework Assignment (1)
- Come up with examples of mismatches
- Explain why and how the mismatch could be a challenge for an MT system
- Propose a solution to the problem
- Solutions may be different when the language is SL or TL

Lexical Mismatches & Structural Mismatches
Examples from MT Class (HW-1)
- Korean
- Chinese
- Telugu
- Kannada
- Vietnamese
- Greek
- Arabic
- Hindi
- French
- German

Lexical and Structural Mismatches between English and Korean

Eric Davis
CMU – LTI
MT
02/19/2007

Lexical Mismatches
- Very
  - 무척 (mucheok) (very...negative)
  - 참 (cham) (very)
  - 아주 (aju) (very)
- Dark
  - 검은 (jiteun) (dark to describe clothes)
  - 어둡다 (eodupda) (dark)
  - 검은 (keomkeomhada) (literally like black)
- Lose
  - 잃다 (iltu) (lose an object)
  - 쓰다 (seuda) (wear glasses or a hat)
  - 쓰다 (sseuda) (wear glasses or a hat)
  - 입다 (sseuda) (wear clothes in general)
  - 신다 (shinda) (wear shoes or slippers)
  - 케다 (kkida) (wear glasses)

Lexical Mismatches (Cont)
- home
  - 집 (jip) (general form)
  -댁 (daek) (polite form)
- rice
  - 밥 (bap) (cooked rice)
  - 밥 (bap) (cooked rice)
  - 밥 (bap) (cooked rice)
- 배 (bae) (stomach)
- 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
  - 배 (bae) (stomach)
Structural Mismatches

- Korean complex NP --> English modal AUX V
  - daehagwonsaengdeureun gondeureun neungnyeokeul kanegimelloneseo chungbunhi ballwhihal su itda.
  - graduate student-PLUR-NOM he-PLUR-POSS potential-ACC CMU-LOC full-MAN show-ABILITY-PRES
  - lit: Graduate students their potential at CMU have the ability to show.
  - gloss: Graduate students can realize their full potential at CMU.

Structural Mismatches (Cont)

- Structures of Relative Clauses (Korean --> no copula or that)
  - daetongnyeoga bushiga ilgeun adongdeosu changgun eoryeonta.
  - president Bush-TOP read-DESC related to children book-NOM difficult-PAST
  - gloss: The children's book that President Bush read was difficult.

- Korean uses lexical causative --> English has main V as causative
  - maekdonaldeu haembeogaga shimjangeul mabishigyeonta.
  - McDonald hamburger-TOP heart-ACC paralyze-CAUSE-PAST
  - lit: The McDonald's hamburger caused the heart to stop.
  - gloss: The McDonald's hamburger paralyzed.

English to Telugu MT Examples

Vamshi
Homework 1: 11-731 MT Course

Lexical Mismatches

- Example 1:
  - Eng: I had an apple. I 'ate' an apple, I 'passed' to apple.
  - Tel: Some 'ete': some eka apilli llauni, Some 'passo': esa dagga eka apple
  - Gloss: Some 'ete': I ate apple, Some 'passo': I (pass marker) near
    an apple

- Example 2:
  - Eng: I have been thinking about it. (have is only an auxiliary verb here)
  - Tel: Nenu daaigutikku aletametamannu.
  - Gloss: I X about thinking

Structural Mismatches

- Example 1:
  - Eng: Linu was going to the market to get an apple for sister.
  - Tel: Nam linu acharu apple teemade-kovalamitki market-ka vellamanu
  - Gloss: Linu sister apple get-to market-to going

- Example 2:
  - Eng: I sat in a less crowded compartment.
  - Tel: Nenu keetam juunai-enni compontem-ko boddumennu.
  - Gloss: I less crowded compartment-in sit.
Lexical Mismatches

<table>
<thead>
<tr>
<th>English</th>
<th>Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>love</td>
<td>αγάπη (agάπη)</td>
</tr>
<tr>
<td>- a feeling or concern towards someone. This form of love is often self centered and does not depend on the beloved possessing desirable qualities.</td>
<td></td>
</tr>
<tr>
<td>- a passionate or romantic love that often has sensual overtones. Typically involves physical attraction.</td>
<td></td>
</tr>
<tr>
<td>- a general love that can be applied to people (such as family), but also situations (playing basketball), or things (like ice cream).</td>
<td></td>
</tr>
<tr>
<td>- a natural affection, almost exclusively used to describe relationships within the family. A love that is the result of familiarity.</td>
<td></td>
</tr>
</tbody>
</table>

Structural Mismatches

Arabic employs a special structure called the Dāʾa (iDaafa) to form a relationship of possession or belonging. In English, possession is shown simply by an “s” or using the word “of”, but in Arabic the nouns are all strung together. These nouns act as a unit and if the phrase is definite only the last noun is marked for definiteness. Here are some examples:

- mktb A lskrtyr: The secretary’s office
- lbp TA albp: A student’s book
Hindi and English

Abhaya Agarwal

French and English

Greg Hanneman

French–English Lexical Mismatch

• Single verb “tourner” can translate to English “to turn” or “to film” depending on context

Le conducteur tourne la voiture.
The driver turns the car.

Le metteur en scène tourne le film.
The director films the movie.

English–French Structural Mismatch

• English “to miss” translates simply to French “manquer” most of the time

Le joueur a manqué le but
The player missed the goal.

Le voyageur a manqué le train.
The traveller missed the train.

English–French Structural Mismatch

• Except when the meaning is “to regret the absence of a person or thing”

Le professeur manque à ses anciens étudiants.
lit. The professor is missing to his former students. His former students miss the professor.

Ses anciens étudiants manquent au professeur.
lit. His former students are missing to the professor. The professor misses his former students.

LTI 11731 Homework 1

Roger Hsiao
wrhsiao@cs.cmu.edu

(* examples about family titles are from http://www.kwanfamily.info/familytitles/familytitle.html)
Lexical Mismatch for Chi and Eng

Family titles

<table>
<thead>
<tr>
<th>Relationship</th>
<th>English (formal)</th>
<th>English (Informal)</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s father</td>
<td>Paternal grandfather</td>
<td>Grandfather</td>
<td>祖父</td>
</tr>
<tr>
<td>Mother’s father</td>
<td>Maternal grandfather</td>
<td>Grandfather</td>
<td>外祖父</td>
</tr>
</tbody>
</table>

Lexical Mismatch for Chi and Eng

Family titles

<table>
<thead>
<tr>
<th>Relationship</th>
<th>English (formal)</th>
<th>English (Informal)</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s older brother</td>
<td>Paternal older uncle</td>
<td>Uncle</td>
<td>伯父</td>
</tr>
<tr>
<td>Father’s younger brother</td>
<td>Paternal younger uncle</td>
<td>Uncle</td>
<td>叔父</td>
</tr>
</tbody>
</table>

Lexical Mismatch for Chi and Eng

Family titles

<table>
<thead>
<tr>
<th>Relationship</th>
<th>English (formal)</th>
<th>English (Informal)</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s older sister</td>
<td>Paternal older aunt</td>
<td>Aunt</td>
<td>姑媽</td>
</tr>
<tr>
<td>Mother’s older sister</td>
<td>Maternal older aunt</td>
<td>Aunt</td>
<td>姨母</td>
</tr>
</tbody>
</table>

Lexical Mismatch for Chi and Eng

Family titles

<table>
<thead>
<tr>
<th>Relationship</th>
<th>English (formal)</th>
<th>English (Informal)</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s brother’s daughter who is younger than you</td>
<td>Paternal younger female cousin</td>
<td>Cousin</td>
<td>堂妹</td>
</tr>
<tr>
<td>Mother’s brother’s daughter who is younger than you</td>
<td>Maternal younger female cousin</td>
<td>Cousin</td>
<td>表妹</td>
</tr>
</tbody>
</table>

Structural Mismatch for Chi and Eng

- tense

- English uses tense to describe the time of an event.
  - He works in a school.
  - He is working in a school.
  - He worked in a school.
  - He has worked in a school.
- But, Chinese does not have tense at all!
  - 他 在 学校 工作
  - 他正在学校工作 ("正在" means current)
  - 他曾经在 学校工作 ("曾经" means past)
  - 他刚刚在 学校工作 ("刚刚" means recent past)
- All four sentences will translate into the same sentence if we do not add additional words.
  - Losing time information.
English-German

Jason Adams

German Lexical Mismatches

• English has one word for *where*
• German has three, depending on context:
  – wo  (at a location)
  – wohin (to a location)
  – woher (from a location)

German Lexical Mismatches (cont.)

• Wo wohnen Sie?
• Wohin gehen Sie heute?
• Woher kommen Sie?

• Where do you live?
• Where are you going today?
• Where do you come from?

German Structural Mismatches

• There are common German verbs corresponding to English copulas (be + adjective).
  • Es *lohnt sich* nicht.
  • It *is not worth* it.

• Er fehlt heute.
• He is *absent* today.

German Structural Mismatches (cont.)

• Cannot just classify all movement verbs as taking wohin or woher:
  – Wohin fahren Sie? (Where are you driving?)
  – Wo fahren Sie Ski? (Where do you ski?)
  – Woher fahren Sie? (Where are you driving from?)

• This problem is compounded by the fact that some English copulas can be translated into German only as a copula, only as a single verb, or as both.
  • Many-to-many correspondence, and often can only make the correct choice based on semantics.
German Structural Mismatches (cont.)

1. The hotel room isn’t worth it.
2. The homework isn’t worth it.

The verb *lohnen* in this type of sentence refers to something of monetary value, while the copula *der Muehe wert* refers to being worth the trouble.

Lexical and Structural Mismatches

Rashmi Gangadharaiah
MLT

One to many possible translations(1)

Eg1: In Kannada (language spoken in Karnataka, India), the word “You” has different translations based on the person being addressed.

Father, can you come in?
thande, niivu ollagade baritiira?
(father, you in come+can)

Rita, can you come in?
rita, niinu ollagade baritiiya?
(rita+nom, you in come+can)

Eg2: The word “rice” in English has different translations in Kannada based on whether it is cooked or uncooked.

“rice” • “anna” if cooked
“rice” • “akki” if uncooked

Complex Predicates

- Nouns, adjectives, and nonfinite forms of verbs can all combine with verbs to form complex predicates.
- The number of arguments, their meanings and case are determined either by N or the V!!
- “LIGHT VERB”: V in a CP
- “HOST NOUN”: The element that the light verb combines with to form a CP.
- Translation of CP is dependent on the host noun and the light verb

Translation of CPs

Problems with CPs,
when the verb in English is a noun in Kannada,
avanu avalameile *barava(N)* itamu
(he+nom her+loc reliance kept) He *refined(V) on her.

when the verb in English takes two arguments where as the verb in Kannada takes three:
avanige *avala meile* *kopa* banthu
(Him-dat her-loc anger-Nom came-Perf (light verb “banthu”) He got angry with her

Without any analysis, the best output that can be produced would be:
Source: avanige avalameile kopa banthu
Target: to him anger came on her \(\rightarrow\) ungrammatical!!
Q2: Derivation: STRUCTURAL PARADOX

- The light verb agrees with the nominal host, the nominal host must be an ARG.
- But the host noun is part of the predicate as it cannot undergo scrambling alone.

nominal host of a light verb is simultaneously an argument and a predicate.

Addition of features

the translation of the verb depends on the entire CP. The translation of “kopa ban’tthu” is “got angry” whereas, the translation of “kopa itthu” is “was angry”.

Summary of Lexical Mismatches

- One to many translations
- No lexical translation in TL
- Contextual information

Summary of Structural Mismatches

- Word order difference: SVO vs. SOV
  - SOV (Japanese, Korean, Malayalam, Tamil, Farsi, Mongolian, etc.)
  - SVO (English, French, Spanish, Chinese, etc.)
- Modifier position
  - Adj + Noun VS. Noun + Adj
- Other structural differences