

# Improving translation of verbal and sentential objects with Its2 MT system

Alexis Kauffmann

Kyoto University, Kurohashi-Kawahara Lab

Geneva University, LATL

12/10/2010

# Contents

- Intro: Its2: a Linguistics Based MT system
- Japanese verb classification:
  - a- subcategorisation
  - b- auxiliaries
  - c-generation
- Translation of English modals:
- Translation of English semi-modals
- Implementation:
  - a- description
  - b- tests
- Experimentation and Results
- Next Research

# Its2: a Linguistics Based MT system

- Used here on English-to-Japanese translation
- With core modules, and language-specific parsing, transfer and generation modules
- With language-specific lexical databases
- Coded in Component Pascal language
- (see Werhli, Scherrer, Nerima 2009)

# Japanese verb classification: subcategorisation

- Subcategorisation frames added in Japanese lexicon
- Bilingual correspondences with most likely corresponding English subcat frame in bilingual lexicon

- Verbs with **と** argument take sentential object

**彼は来ると**思います。

- Some verbs with **を** or **が** take nominalised verbal/sentential objects

**あそこに行ったのを**覚えてる。

# Japanese verb classification: auxiliaries

- Japanese auxiliary and light verbs can take verbal objects:
    - conjugated in ㇿ form
    - or conjugated in base form
    - or conjugated in base form + に
- object conjugation form updated in the lexicon

# Japanese verb classification: generation

- Specific rules implemented
- In order to generate verbal/sentential object with the adequate conjugation form
- Post-verbal particles (**のを**, **ことが**, **に**) added when necessary

# Translation of English modals

- “can”, “may”, “must”...
- Linguistic rule specifications based on **望月くん** and **中澤さん**'s work on modality, and on some observation on bicorpus.
- Can → -える  
Must → -なければならない  
May → かもしれない  
...

# Translation of English semi-modals

- Linguistic rule specifications based on **望月くん** and **中澤さん**'s work on modality.
- English semi-modals express modality, but their conjugation has tense variation like a normal verb: “want to”, “have to”, “need to”...
- Have to → **-なければならない**  
Do not have to → **必要がない**  
Want to → **-たい**  
...



# Implementation

- Generation rules for verbal and sentential objects: all listed cases implemented.
- Transfer rules for modals: 75 % of studied cases implemented. However, some ruled have been oversimplified at implementation.
- Transfer rules for semi-modals: 50% of studied cases implemented.

# Implementation: tests

- Some ambiguity on modals, such as “could”, can lead to false translation
- Bad results on nominalised sentential objects
- Problem in choosing the tense of some verbal objects
- But good results for most of implemented rules
- However, overall effect on global translation result seem not as high as expected, due to other problems with the MT system (syntactic parsing mistakes, missing vocabulary in the lexicon...)

# Experimentation and Results

- 1 test set with travel domain sentences
- 1 test set with paper abstract sentences
- BLEU score without modality and verbal object handling:
- BLEU score with modality and verbal object handling:

# Experimentation and Results

- 1 test set with travel domain sentences
- 1 test set with paper abstract sentences
- BLEU score without modality and verbal object handling:  
ZERO??
- BLEU score with modality and verbal object handling:  
ZERO??

mistake: forgot Japanese word segmentation

→ TO BE CONTINUED