Reference in Japanese-English Machine Translation

Francis Bond and Kentaro Ogura

Presented by: Manuel Medina González. Department of Artificial Intelligence
Professor Nomura's Laboratory of Natural Language Processing

Pattern Recognition 2006 December 4th, 2006

Outline

1. Introduction
   - Understanding the problem
   - NP reference
   - Referentiality in Natural Language Processing
2. Determination of Reference
   - Rules
3. Application to Translation
4. Evaluation
   - ALT-J/E
   - Result
   - Discussion

Outline

1. Introduction
   - Understanding the problem
   - NP reference
   - Referentiality in Natural Language Processing
2. Determination of Reference
   - Rules
3. Application to Translation
4. Evaluation
   - ALT-J/E
   - Result
   - Discussion

The problem

- Determining the reference of NPs helps to generate correct articles and Noun's number when translating from Japanese to English

Example

私はりんごが好きです。

- I like apple.
- I like apples.
- I like the apple.
- I like the apples
- ...

The problem

- Determining the reference of NPs helps to generate correct articles and Noun's number when translating from Japanese to English

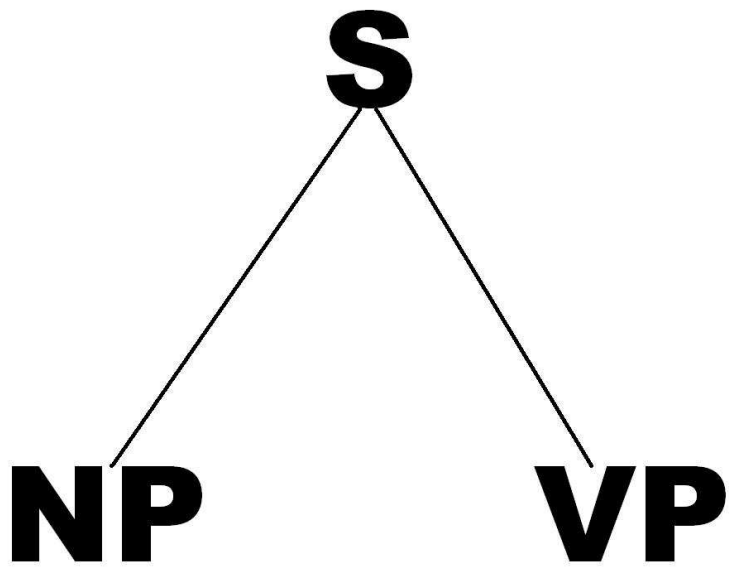
Example

私はりんごが好きです。

- I like apple.
- I like apples.
- I like the apple.
- I like the apples
- ...

What is NP?

NP = Noun Phrase



- Noun is the main element.
  - Articles
  - Number (Singular/Plural)

What is NP reference?

Definition

To understand the way a noun is being used within a sentence.

The authors consider 3 types of reference:

- `Referential`
- `Generic`
- `Ascriptive`

Referential NPs

- Refers to some entity in the discourse world.
- Introduce a potential identity.
- Can be pronominalized.
- Can be modified by all determiners.
- Can be plural or singular.

Example

私はマンモスを探している。I'm looking for a mammoth

Default interpretation for all NPs.

Referential NPs

- Refers to some entity in the discourse world.
- Introduce a potential identity.
- Can be pronominalized.
- Can be modified by all determiners.
- Can be plural or singular.

Example

私はマンモスを探している。I'm looking for a mammoth

Default interpretation for all NPs.

Referential NPs

- Refers to some entity in the discourse world.
- Introduce a potential identity.
- Can be pronominalized.
- Can be modified by all determiners.
- Can be plural or singular.

Example

私はマンモスを探している。I'm looking for a mammoth

`Default interpretation for all NPs.`

Generic NPs

- Refers to kinds, not single members.
- Can't be modified by all determiners.
- Grammatical number is fixed.
- Take 3 forms in English:
  - Indefinite singular *"a"*
  - Definite singular *"the"*
  - Bare plural and generic ϕ

Example

a. A mammoth is a mammal.

b. The mammoth is a mammal.

c. Mammoths are mammals.

d. Furniture is expensive.

Generic NPs

- Refers to kinds, not single members.
- Can't be modified by all determiners.
- Grammatical number is fixed.
- Take 3 forms in English:
  - Indefinite singular *"a"*
  - Definite singular *"the"*
  - Bare plural and generic ϕ

Example

a. A mammoth is a mammal.

b. The mammoth is a mammal.

c. Mammoths are mammals.

d. Furniture is expensive.

Ascriptive NPs

- Ascribe an attribute to another NP.
- Complement of a *copular* verb.
- Agree with their target in number.

Example

a. A computer is a tool.

b. Computers are tools.

c. The students are a mob

Ascriptive NPs

Not all complements of copular verbs are ascriptive:

- Predicational

Example

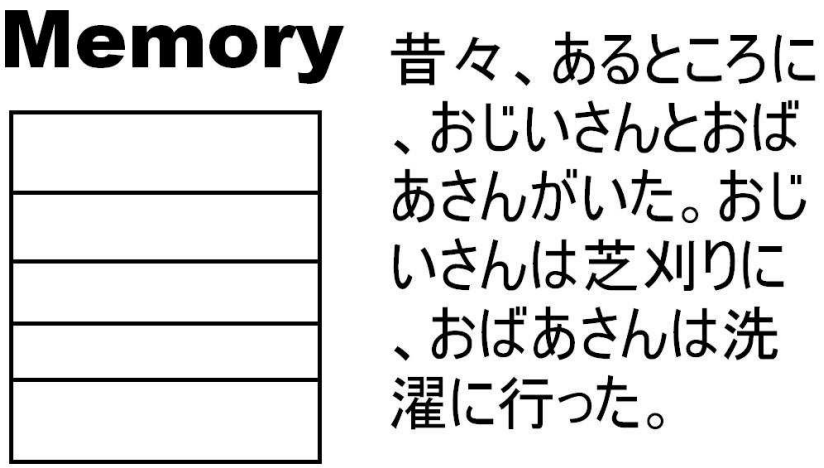
私の指導者はいい人です。

- Specificational

Example

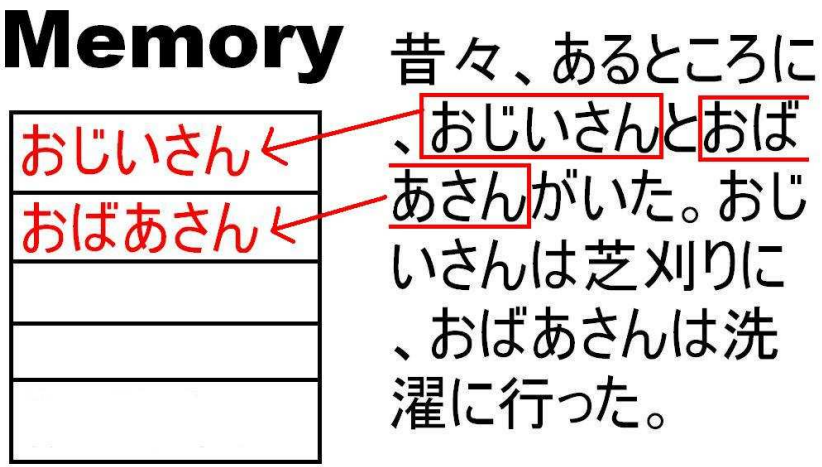
私の指導者はXさんです。
Xさんは私の指導者です。

Why is referentiality necessary?



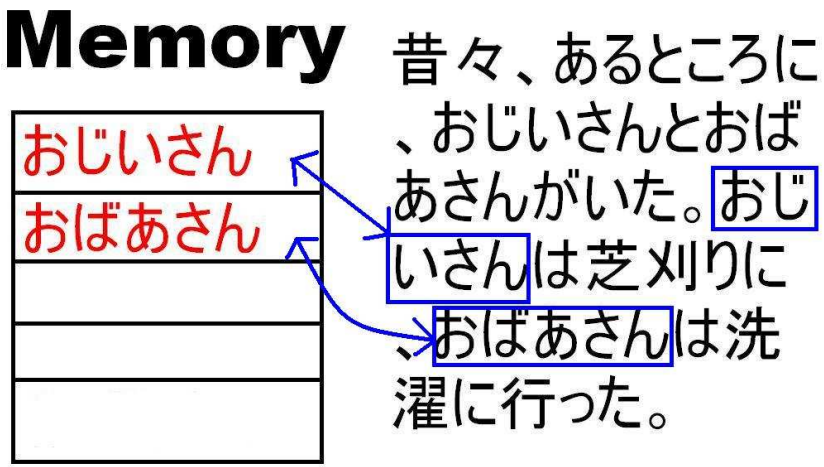
When a noun first appears, there's nothing in the memory.

Why is referentiality necessary?



The noun is added and indefinite article is used.

Why is referentiality necessary?



The second time a noun appears, it will be already in memory, thus, definite article is used.

1. Nouns that appeared before.

Using the rule applied in the previous slide:

Example

像は哺乳類であり、マンモスも哺乳類である。

Elephants are mammals and mammoths are also the mammals

1. Nouns that appeared before.

Using the rule applied in the previous slide:

Example

像は哺乳類であり、マンモスも哺乳類である。

`Elephants are mammals and mammoths are also the mammals`

2. Concordance in number

Example

有名なシャーロックホームズ探偵の物語は大抵ロンドン地域を背景にしたものです。

The tales of famous detective Sherlock Holmes normally have the London area as a background

2. Concordance in number

Example

有名なシャーロックホームズ探偵の物語は大抵ロンドン地域を背景にしたものです。

The tales of famous detective Sherlock Holmes normally have the London area as a background

3. Determining the correct number

Example

この日は社員は来ません。

Employees do not come on this day

3. Determining the correct number

Example

この日は社員は来ません。

Employees do not come on this day

Outline

1. Introduction
   - Understanding the problem
   - NP reference
   - Referentiality in Natural Language Processing
2. Determination of Reference
   - Rules
3. Application to Translation
4. Evaluation
   - ALT-J/E
   - Result
   - Discussion

Definition

Search for patterns within a sentence to determine:

- Structure
- Type

Example

If the subject is marked by one of the focus markers は or も, and the verb is past tense, then the sentence has a specificational reading.

子供はボールで遊んだ　　Search for patterns...

Definition

Search for patterns within a sentence to determine:

- Structure
- Type

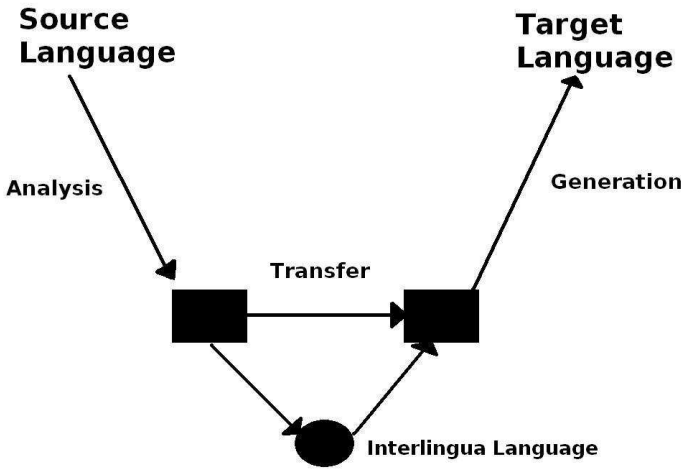
Example

If the subject is marked by one of the focus markers は or も, and the verb is past tense, then the sentence has a specificational reading.

子供はボールで遊んだ **Specificational**

Creation

Rules are created to be applied to the original text in the transfer phase.



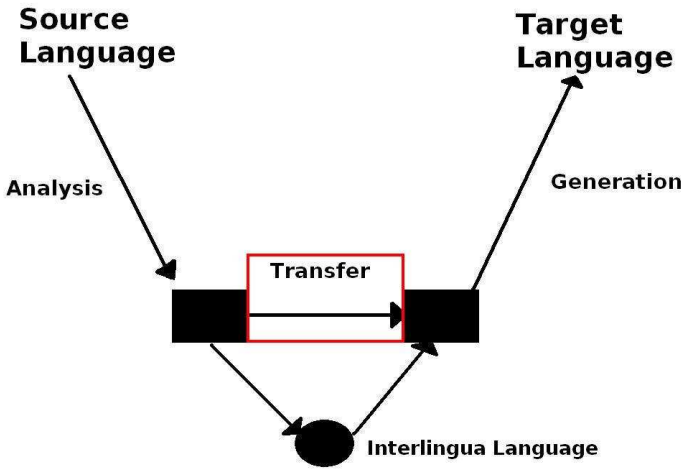
3 levels of rules:

1. Clause Level
2. Case-element Level
3. Modification of and within the NP

All proper nouns and pronouns are assigned the value referential

Creation

Rules are created to be applied to the original text in the transfer phase.



3 levels of rules:

1. Clause Level
2. Case-element Level
3. Modification of and within the NP

All proper nouns and pronouns are assigned the value referential

Example

— The subject of a **specificational** sentence is referential (and definite).
問題は困難であった。 The problem was difficult.
— The complement of the copula in a **predicational** sentence is ascriptive.
マンモスは動物である。 Mammoths are animals.
— The complement of として is ascriptive:
それをマンモスとして扱って。 Treat it as a mammoth.
— Locative NPs are **referential**.
空港に行った。 I went to the airport.
— An NP whose head is modified by a relative clause is **referential**.
店で買ったプレゼント。 The gift I bought at the store.
...

Outline

1. Introduction
   - Understanding the problem
   - NP reference
   - Referentiality in Natural Language Processing
2. Determination of Reference
   - Rules
3. Application to Translation
4. Evaluation
   - ALT-J/E
   - Result
   - Discussion

Countability of Nouns

Based on countability, nouns are divided in:

- Fully countable: **knife**
- Both countable
  - Strongly countable: **cake**
  - Weakly countable: **beer**
- Uncountable: **furniture**
- Pluralia tantum: **scissors**

Generic NPs

GT	FC	BC	U	PT
a	a mammoth	a cake	*a furniture	*a scissors
the	the mammoth	#the cake	#the furniture	#the scissors
ϕ	mammoths	cake/cakes	furniture	scissors

GT: **Generic Type**
FC: **Fully Countable**
BC: **Both Countable**
U: **Uncountable**
PT: **Pluralia Tantum**

- # = Cannot have *generic* reference.
- * = Ungrammatical

Referential NPs

Special rules are used to determine Referential NPs:

Example

— Determine according to verb:
話題がつきた。 I ran out of topics.
— Determine according to adjunct:
全国の学校。 Schools all over the country.
— Specific function words that mark plurality:
達 = plural and countable

Trigger Nouns

Definition

Nouns that have been manually identified and marked in the lexicon

- Basically refers to NPs that in English use a possessive pronoun but in Japanese is often omitted.

Example

財布を落した。 I lost my wallet

Trigger Nouns

Definition

Nouns that have been manually identified and marked in the lexicon

- Basically refers to NPs that in English use a possessive pronoun but in Japanese is often omitted.

Example

財布を落した。 I lost my wallet

Selecting determiners for referential NPs

Example

— `place nouns` heading locative NPs are **definite**

空港に行った。 I went to the airport

— NPs whose head is modified by a `relative clause` is **definite**

昨日来た男。 the man who came yesterday

— An NP whose head is modified by one of certain modifiers is **indefinite**

ある男。 a certain man

Ascriptive NPs

As heuristic, all ascriptive NPs are indefinite = articles will be generated in the same way as for `indefinite referential NPs`:

- Countable singular NPs will take the **indefinite article**.
- Countable plural and uncountable NPs will take ϕ.

Example

- 癌は病気だ。 Cancer is a sickness.
- 子供は天使だ。The children are angels.

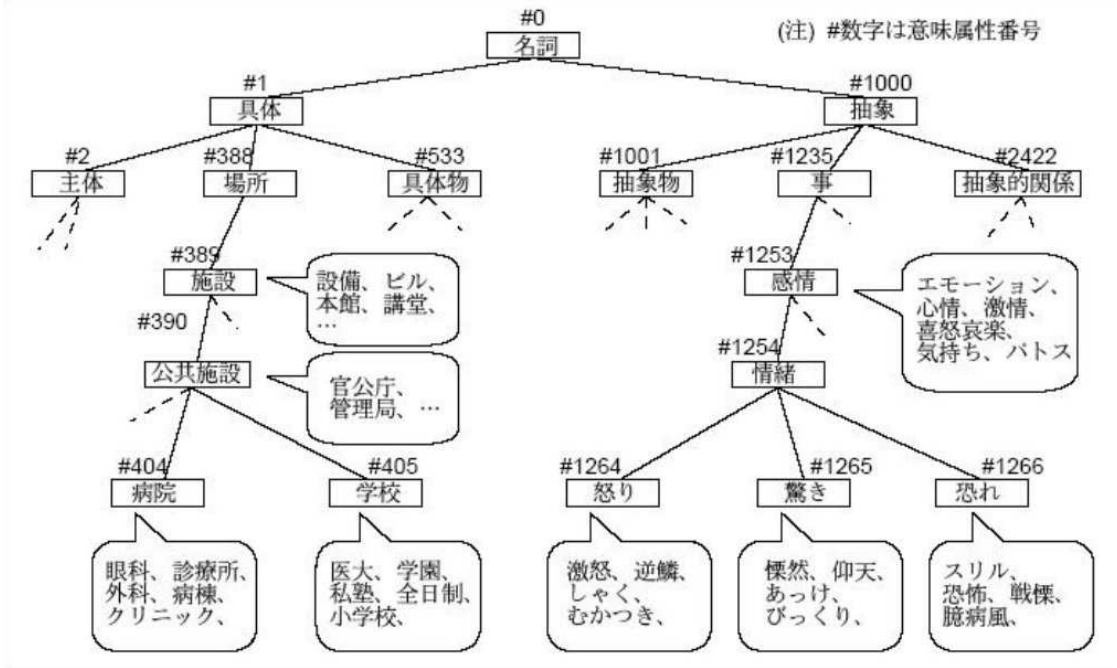
Outline

1. Introduction
   - Understanding the problem
   - NP reference
   - Referentiality in Natural Language Processing
2. Determination of Reference
   - Rules
3. Application to Translation
4. Evaluation
   - ALT-J/E
   - Result
   - Discussion

Description

All the process described has been implemented in ALT-J/E Machine Translation System.

- Developed by NTT Labs between 1982 and 1986
- Very robust
- Uses a 2,715 semantic hierarchy for common nouns



Procedure

- The algorithm was evaluated on a collection of newspaper articles from the `Nikkei-Sangyo` newspaper by an English native speaker not connected with the development of the algorithm.
- The results were compared against and old version of ALT-J/E.

Algorithm	NPs(717)	Sentences(102)
New	85%	36%
Old	65%	5%

Evaluation values

Each NP was given one of the following values:

best:	The most appropriate article and number.
analysis:	Problem with the parse.
NP structure:	Problem with NP structure or choice of translation.
dictionary:	Problem with the dictionary.
article:	Inappropriate article
number:	Inappropriate number.
possesive:	Inappropriate use of possessive determiner.
countability:	problem with countability.
reference:	problem with reference property.

- For a sentence to be judged as correct all the NPs must be correct.

Correct example

Example

会員は 37,000 円、非会員は 47,000 円の参加費が必要。

New: Members need 37,000 yen as the registration fee and non-members need 47,000 yen.

Old: A member needs 37,000 yen as the registration fee and a non-member needs 47,000 yen.

Errors

Problem	Frequency
Analysis	7%
Dictionary	7%
NP structure	18%
Reference	13%
Number	25%
Articles	26%
Possessive	5%

Unrecognized example

Example

NTT は新会社 NTT テレコムエンジニアリング東京を 8 日付で設立する。

- NTT will establish a new company. NTT Telecom Engineering Tokyo, on the 8th.

Example

新会社の資本金は 800 万円。

- The capital of a/the new company is 8 million yen.

Summary

- Referentiality is an important concept necessary in order to get high-quality translations.
- Some rules used to determine NP referentiality in Japanese-English Machine Translation were presented.
- Although the method has been proven to be efficient, several problems remain to be explored.