

Disambiguation of Coordinate Expressions in Japanese by Extracting Mutual Case Relation

Yoshiyuki Kotani

Dept. of Computer, Information & Communication Science,
Tokyo University of Agric. & Tech.
kotani@cc.tuat.ac.jp

Japanese coordinate noun phrases by the particle TO are often ambiguous on whether it means two parallel propositions (*and*) or a mutual case relation (*with* or *against*), as deep case structure. It was a hard problem to determine it, though they are widely used. We propose a method of solving the ambiguity by analyzing mutualness of verbs and adjectives. The mutualness is determined by three features of each verb or adjective. The first feature indicates permission of mutual expression in the subject, and the second in the object. The last shows if a verb is voluntary.

Using this method we design a parsing mechanism, where matching of features is represented as neutralization between predicate arguments.

1. Introduction

The particle TO is often used between noun phrases to represent coordinate relation in Japanese language such as:

JOHN TO MARY GA KENKASURU.
John and Mary SUBJ fight
John and Mary fight.

It is interesting that TO is also used as the case postposition of cooperation, opposition or equality:

JOHN GA MARY TO KENKASURU.
John SUBJ Mary against fight
John fights against Mary.

This first example has in most cases the deep case structure same as the second, which shows the structure directly. But the example

JOHN TO MARY GA WAKAI.
John and Mary SUBJ be young
John and Mary are young.

does not have such deep case structure but has a parallel pair of propositions that John is young and that Mary is young.

Therefore it is necessary to distinguish the first from the third, to produce different semantic structures which correspond to deep case structures.

We investigate the particle TO here, and deep case structure for each classification of it.

2. Analysis of TO

Japanese postpositions, which include conjunctions in European grammars, are classified by usage as follows.

Table 1. classification of postpositions

case postposition, conjunctive postposition, terminal postposition, relational postposition, noun modifier, nominalizer, coordinate postposition
--

The particle TO is used as three of them. As a case postposition, two meanings can be distinguished. The four ways of TO usage are in the Table 2.

Here we discuss the classes 1 and 2 rather than 3 and 4, since the latter can be dealt with

ordinary polysemy disambiguation process. The TO in the class 1 connects a pair of noun phrases, and the combination is also a noun phrase. The grammatical rule is:

noun_phrase ----> noun_phrase TO noun_phrase

where TO agrees with AND of English except that TO cannot be used between other categories. So it cannot be extended to any metarule.

The TO in the class 2 makes a noun phrase a postposition phrase which modifies a verb phrase. This is described as the rules:

verb_modifying_phrase --> noun_phrase case_postposition

and

case_postposition --> TO .

Table 2. classification of the particle TO

class	category	(English)
example sentences / meaning		
1	Coordinate postposition	(and)
JOHN TO MARY GA KITA. John and Mary SUBJ came John and Mary came.		
2	Case postposition	(with, against)
JOHN GA MARY TO KITA. John SUBJ Mary with came John came with Mary.		
3	Case postposition	(that)
JOHN GA KITA TO OMOU. John SUBJ came that think I think that John came.		
4	Conjunctive postposition	(if)
JOHN GA KURU TO MARY GA NIGERU John SUBJ come if Mary SUBJ escape If John comes, Mary escapes.		

3. Deep Case Structures Related to TO

Deep case structures or semantic structures of the sentences in which TO in the class 1 or 2 occurs are investigated here.

The first of what TO means is "a pair of parallel events." It is not understood as a single deep structure but a pair of structures which has the same case contents except one. For example, the sentence

JOHN TO MARY GA WAKAI.
 John and Mary SUBJ be young
 John and Mary are young.

can be divided to

JOHN GA WAKAI.
John SUBJ be young
John is young.

MARY GA WAKAI.
Mary SUBJ be young
Mary is young.

But it cannot be changed into the sentence

***JOHN GA MARY TO WAKAI**
John SUBJ Mary with be young
*John is young with Mary.

The second meaning of TO is "cooperation." It can be adequately adopted as a deep case. It shows that a person or an organization does something cooperatively with another. Here the agent is a person or an organization, and so the act or the predicate has to be a voluntary one. The cooperation case is not an obligatory one, because the cooperation is optional.

For example the class 1 expression

JOHN TO MARY GA BENKYOUSURU.
John and Mary SUBJ study
John and Mary study.

means their cooperative act. It is changed into the class 2 expression

JOHN GA MARY TO BENKYOUSURU.
John SUBJ Mary with study
John studies with Mary.

or into

MARY GA JOHN TO BENKYOUSURU.
Mary SUBJ John with study
Mary studies with John.

On the other hand it cannot be divided to the pair of

JOHN GA BENKYOUSURU.
John SUBJ study
John studies.

and the other. Of course this is not the case if it does not show a cooperative act. For example, it is in the situation that someone is asked which two of three people study well.

The last meanings TO are "mutual cases." One is the mutual subject case (GA case in surface) and the other the mutual object case (O case in surface). Coordinate expression in subject phrases or object phrases indicates mutual relation between the two connected phrases, where one is the subject or the object and the other is the opponent. These are similar to the cooperation case, but different from it in whether related to agent's intention or not.

In the expression of class 1

JOHN TO MARY GA TATAKAU.
John and Mary SUBJ fight
John and Mary fight.

Mary is John's opponent. We classify this into a mutual subject case. It is changed in the same manner as the cooperation case into the class 2 expression

JOHN GA MARY TO TATAKAU.
John SUBJ Mary against fight
John fights against Mary.

or into

MARY GA JOHN TO TATAKAU.
Mary SUBJ John against fight
Mary fights against John.

It cannot be divided to

JOHN GA TATAKAU.
 John SUBJ fight
 John fights

and the other. The example of class 1

DICK GA JOHN TO MARY O KURABERU.
 Dick SUBJ John and Mary OBJ compare
 Dick compares John and Mary.

is a mutual object case. This can be changed into the sentence

DICK GA JOHN O MARY TO KURABERU.
 Dick SUBJ John OBJ Mary with compare
 Dick compares John with Mary.

The discussion above is concluded as the table 3. It also includes the examples of interchangeability between noun modifying phrases, where there are similar phenomena except cooperation case.

Table 3. The Classification of Coordinate Expressions with TO and The Expressions Changeable from them (under the dotted lines)

[verb modifying phrases]

(1-1) parallel:
JOHN TO MARY GA WAKAI. John and Mary SUBJ be young John and Mary are young.
JOHN GA WAKAI. MARY GA WAKAI. John SUBJ be young Mary SUBJ be young John is young. Mary is young.
(1-2) cooperation:
JOHN TO MARY GA BENKYOUSURU. John and Mary SUBJ study John and Mary study.
JOHN GA MARY TO BENKYOUSURU. John SUBJ Mary with study John studies with Mary.
(1-3) mutual subject:
JOHN TO MARY GA TATAKAU. John and Mary SUBJ fight John and Mary fight.
JOHN GA MARY TO TATAKAU. John SUBJ Mary against fight John fights against Mary.
(1-4) mutual object:
DICK GA JOHN TO MARY O KURABERU. Dick SUBJ John and Mary OBJ compare Dick compares John and Mary.
DICK GA JOHN O MARY TO KURABERU. Dick SUBJ John OBJ Mary with compare Dick compares John with Mary.

(Table 3. Continued)

[noun modifying phrases]

(2-1) parallel:
JOHN TO MARY NO WAKASA John and Mary of youth John's and Mary's youth
JOHN NO WAKASA TO MARY NO WAKASA John of youth and Mary of youth John's youth and Mary's youth
(2-2) cooperation:
JOHN TO MARY NO BENKYOU John and Mary of study John's and Mary's study
?JOHN NO MARY TO NO BENKYOU John of Mary with of study John's study with Mary
(2-3) mutual subject:
JOHN TO MARY NO TATAKAI John and Mary of fight fight between John and Mary
JOHN NO MARY TO NO TATAKAI John of Mary against of fight John's fight against Mary
(2-4) mutual object:
DICK NI YORU JOHN TO MARY NO HIKAKU Dick by John and Mary of comparison comparison between John and Mary by Dick
DICK NI YORU JOHN NO MARY TO NO HIKAKU Dick by John of Mary with of comparison comparison of John with Mary by Dick

4. Features and Case Dependency of Verbs and Adjectives

Coordinate noun phrases with TO are possible to occur at any place of a noun phrase in the sentence, independently from the case and from the verb, if they are used as parallel events. For example,

JOHN TO MARY WA INU TO NEKO O YAMA TO KAWA NI TSURETE IKU.

John and Mary TOPIC dog and cat OBJ mountain and river at take

John and Mary take a dog and a cat to a mountain and a river.

This is formulated as:

RULE: coordinate noun phrases with TO whose deep meaning is a pair of parallel matters can occur at any place of a noun phrase arbitrarily.

On the other hand, cooperation case, mutual subject case, and mutual object case can occur in the sentence, dependently upon categories of verbs or adjectives. Such verbs and adjectives have their own case dependency structures that includes TO case which indicates cooperators, opponents etc..

We adopt here the feature "VOL" or "voluntary" of verbs and adjectives, in order to disambiguate cooperation case. It indicates some will or intention by the agent at the subject case in the sentence. If there is no such intention, then the verb or the adjective has the feature "0VOL" or "not voluntary."

For example WAKAI (to be young) has 0VOL and BENKYOUSURU (to study) has VOL, in

JOHN TO MARY GA WAKAI.

JOHN TO MARY GA BENKYOUSURU.

Using this feature, we construct the next:

RULE: verbs and adjectives which have VOL feature has the optional TO case in surface which is the cooperation case in deep structure. Those which have 0VOL does not have it.

There are the following lexical characteristics of Japanese language on VOL and 0VOL.

- (1) Most of Japanese verbs have VOL
(exception: WAKARU (understand), CHIGAU(differ),ICCHISURU(agree with))
- (2) Most of adjectives have 0VOL.
- (3) Most of verbs and adjectives which have O case have VOL(exception: HOSHII(want)).
- (4) The auxiliary verb U and YOU makes verbs attached by it VOL.

Verbs and adjectives are classified into two categories by whether the mutual subject case is permitted or not. We describe them as MSUBJ and 0MSUBJ. This classification is independent from whether it is voluntary or not. So we get four groups by them. The MSUBJ case is obligatory. If there is no MSUBJ case then some omitting process is needed. We can conclude:

RULE: verbs and adjectives which have MSUBJ feature has the obligatory TO case in surface which is the mutual case with the subject in deep structure. Those which have 0MSUBJ do not have it.

The MOBJ case also classify verbs and adjectives by whether they have it or not. The markers for the classes are MOBJ and 0MOBJ. The MOBJ class is included in VOL. It seems that the class cooccurs with MSUBJ. The class MOBJ is also obligatory. We can conclude:

RULE: verbs and adjectives which have MOBJ feature has the obligatory TO case in surface which is the mutual case with the object in deep structure. Those which have 0MOBJ do not have it.

The five classes are determined by the discussion.

Table. 4. Classification of Verbs and Adjectives and Examples

0VOL, 0MSUBJ, (0MOBJ)
WAKAI(be young), SEIKAKU GA II(have good personality), WAKARU(understand), SHIZUKADA(be calm), KANAU(suit, come up to),FURU(rain)
0VOL, MSUBJ, (0MOBJ)
HITOSHII(be equal),SHITASHII(be familiar),NAKA GA II(be intimate), CHIGAU(differ),KANKEI GA ARU(have relation),ICCHISURU(agree)
VOL, 0MSUBJ, 0MOBJ
TABERU(eat), IKU(go), BENKYOUSURU(study),IRU(be)
VOL, MSUBJ, 0MOBJ
KUTTSUKU(be connected), TATAKAU(fight), AU(match), KEKKONSURU(marry) <adjective(0VOL,MSUBJ)>+NARU(come to be)
VOL, MSUBJ, MOBJ
KUTTSUKERU(connect),KURABERU(compare),AWASERU(make match) <adjective(0VOL,MSUBJ)>+SURU(make)

The verbs and adjectives which have VOL permit the cooperative TO case (+COOP). This is expressed by that they have the marker ~COOP to use the case matching. The verbs and adjectives which have the feature MSUBJ need the mutual subject case (+MSUBJ). This obligation is expressed by that they have the marker -MSUBJ which must match +MSUBJ. The verbs with MOBJ have the marker -MOBJ in the same way.

5. Transformation and Preference

We claim that the coordinate noun phrase by TO is considered to be generated by the rules like transformation grammar. They change the TO cases in the three surface cases: cooperation, mutual subject, and mutual object (+COOP, +MSUBJ, +MOBJ) into coordinate phrases.

RULE: the noun phrase with +COOP case can move to the subject noun phrase, attaching the coordinate postposition TO.

RULE: the noun phrase with +MSUBJ case can move to the subject noun phrase, attaching the coordinate postposition TO.

RULE: the noun phrase with +MOBJ case can move to the object noun phrase, attaching the coordinate postposition TO.

There may be the additional rule:

RULE: the subject or object noun phrase in the rules must not be coordinate ones.

The example "JOHN TO MARY GA KEKKONSURU(John and Mary marry)" can be understood usually as that they made a family. Here this coordinate phrase is transformed from a subject case by GA and a MSUBJ case by TO. On the other hand, it is possible to understand that they combine two marriage ceremonies together. It is also possible that they marry apart from each other, when asked who married last year.

We claim here the preference for these possibility, that is, which is liked more than others in understanding. The next rule is extracted for class 1:

**RULE: In coordinate noun phrases, cases are preferred in the order
MSUBJ, MOBJ > COOP > parallel structure**

For class 2,

**RULE: In verb modifying phrases by TO, cases are preferred in the order
MSUBJ, MOBJ > COOP**

There is an additional rule for exclusion:

RULE: the MSUBJ TO case and the MOBJ TO case cannot occur together with each other.

The example

**JOHN TO MARY GA DICK TO TATAKAU
(John and Mary fight against Dick)**

shows the effect of the rule. Here Dick occupies the MSUBJ case, and so Mary cannot do it. By preference Mary occupies COOP case. This means Mary fights with John, rather than against him.

6. Japanese Partial Grammar Dealt with TO

Utilizing our idea above in parsing systems, we adopt definite clause grammar(DCG). An argument of the predicates for categories are used as the parameter to check the correspondence of case dependency. For example, transitive verbs are described as verb(-obj). Objects, which are verb modifying phrases by the case postposition O in Japanese, are described as verb_modifying_phrase(+obj). When one generates a sentence, +obj neutralizes -obj. Generally, the system neutralizes the pair of the markers with plus and minus symbols. This process computes the case correspondence. We use DCG and this notation to make simpler to show the process than other like LFG which can deal with the same.

We can construct a partial grammar in Fig. 1, where parentheses [] for lists of marker sequence are omitted for simplicity. The neutralization mechanism can work by the predicates in Fig 2.

The case postposition TO puts one of the parameters +COOP, +MSUBJ and +MOBJ onto verb modifying phrases. Coordinate noun phrases may have one of them. Otherwise, they mean parallel matters.

The symbol ~ is similar to the minus. It differs from the minus at the point that it is not necessary to be neutralized.

Fig. 1. Japanese Partial DCG which deals TO case

```

sentence --> verb_phrase( )
verb_phrase(X) --> verb(X)
verb_phrase(X) --> adjective(X)
verb_phrase(X) --> verb_modifying_phrase(A), verb_phrase(B),{cancel(A,B,X)}
verb_modifying_phrase(+coop) --> noun_phrase( ),[to]
verb_modifying_phrase(+msubj) --> noun_phrase( ),[to]
verb_modifying_phrase(+mobj) --> noun_phrase( ),[to]
verb_modifying_phrase(+subj,X) --> noun_phrase(X),[ga]
verb_modifying_phrase(+obj,X) --> noun_phrase(X),[o]
noun_phrase( ) --> noun_phrase( ),[to],noun_phrase( )
noun_phrase(+coop) --> noun_phrase( ),[to],noun_phrase( )
noun_phrase(+msubj) --> noun_phrase( ),[to],noun_phrase( )
noun_phrase(+mobj) --> noun_phrase( ),[to],noun_phrase( )
noun_phrase(X) --> noun(X)

```

Fig. 2. Neutralization : Processing for + and - cancellation

```

cancel([ ],Y,Y).
cancel([A|X],Y,U):- cancel1(A,Y,Z),cancel(X,Z,U).
cancel1(A,[ ],[A]).
cancel1(~C,Y,Y).
cancel1(A,[~C|Y],Z):- cancel(A,Y,Z).
cancel1(A,[B|Y],Y):- match(A,B),!.
cancel1(A,[B|Y],[B|AY]):- cancel1(A,Y,AY).
match(+A,-A). match(-A,+A). match(+A,~A). match(~A,+A).

```

The example of lexicon is shown as follows. For example KURABERU(compare) needs +SUBJ, +OBJ and +MOBJ as its cases. The case +COOP is possible.

Fig. 3. Example of Lexicon

```

adjective( subj) --> [wakai]
verb( subj) --> [furu]
adjective( -subj, -msubj) --> [hitoshii]
verb( -subj, -msubj ) --> [icchisuru]
verb( -subj,~coop) --> [iku]
verb( -subj,~coop, -obj) --> [taberu]
verb( -subj, -msubj ,~coop) --> [kutttsuku]
verb( -subj, -msubj ,~coop, -obj) --> [toriau]
verb( -subj,~coop, -obj, -mobj) --> [kuraberu]
noun( ) --> [hito] % HITO: a person
noun( ) --> noun( ),[tachi] % TACHI: and so on
noun(+coop) --> noun( ),[tachi]
noun(+msubj) --> noun( ),[tachi]
noun(+mobj) --> noun( ),[tachi]

```


The expression HITO TACHI(people, HITO: a person, TACHI: pluralizer) has the same characteristics as coordinate noun phrases with TO. The appendix shows the neutralizing process in parsing.

7. Processing of the particle NO

The particle NO makes noun modifying phrases like English OF. Our theory can extend to noun modifying phrases, if the C in

A TO B NO C (C of A and B)

is a noun phrase which means mutual relation. This mechanism is implemented by the following extension.

Fig. 4. The DCG for the noun modifier NO and words

noun_phrase(X)	-->	noun_phrase(A),[no],noun_phrase(B),{cancel(A,B,X)}
noun(msubj)	-->	[naka] % NAKA(personal relation)

For example, the parameter +MSUBJ of JOHN TO MARY is neutralized with the parameter MSUBJ of NAKA.

JOHN TO MARY NO NAKA GA WARUI.
John and Mary of relation SUBJ bad
The relation between John and Mary is bad.

8. Conclusion

Japanese coordinate noun phrases of the form A TO B (A and B) are often ambiguous in the sentences. The A and B may be in the relation of cooperation, or the B may be the partner or the opponent. The form shows not only parallel proposition but also cooperation or mutual relation in deep structure. We proposed the following to clarify this problem and to implement the computational system which resolve this ambiguity:

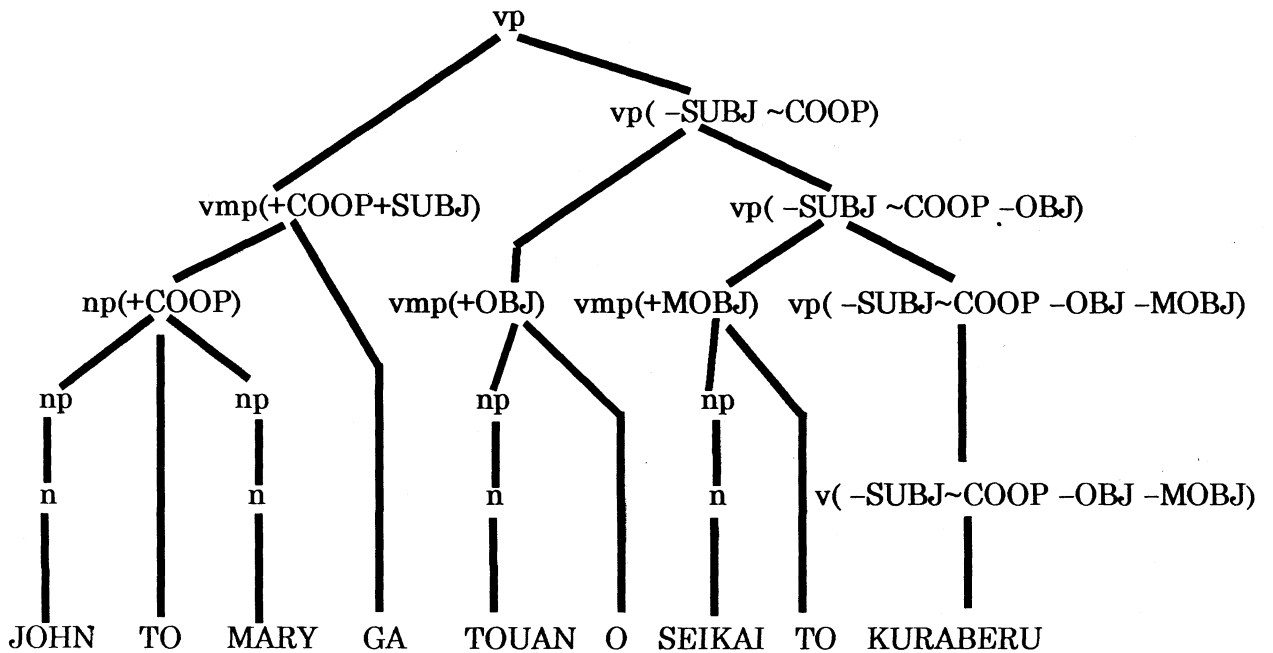
- (1) The three case markers +COOP (cooperation) , +MSUBJ (mutual subject) and +MOBJ(mutual object) are defined.
- (2) Coordinate noun phrases are classified into the four groups, whether they have one of them or not, and which they have.
- (3) Disambiguation of them is done in parsing by neutralizing the plus markers which noun phrases have and the minus ones which the stem verbs or adjectives have.
- (4) Japanese verbs and adjectives are classified by the possibility of neutralization, that means which deep cases are permitted by them.
- (5) The method of implementation is shown as a DCG system, where matching of features is represented as neutralization between predicate arguments.

This method is applicable to today's practical natural-language-processing systems as it is.

References

1. Nomura, Koike, Encyclopaedia of Japanese language, Tokyodo, 1992.
2. Seiichi Makino, Michio Tsutsui, A Dictionary of Basic Japanese Grammar, The Japan Times, 1988.
3. Gazdar, G., Unbounded Dependencies and Coordinate Structure, Linguistic Inquiry, Vol.12, 1981.
4. Sedogbo, C., About Coordination Treatment in Metamorphosis Grammars, Group BULL, 1983.
5. Yoshiyuki Kotani, A System of Practical Linguistic Knowledge Acquisition in Japanese Language, Proc. of Pacific Rim International Conference on Artificial Intelligence '90, pp. 643 648, 1990.
6. Yoshiyuki Kotani, Toshika Sasaki, Nobuo Takiguchi, Syntactic Processing of Parentheses in Japanese Sentences, Proc. of Natural Language Pacific Rim Symposium '91, pp.186 192, 1991.
7. Nobuo Takiguchi, Jian Ming Xie, Yoshiyuki Kotani, Acquisition of Semantic Structure on Case Frame Structure, Proc. of Natural Language Pacific Rim Symposium '91, pp.337 344, 1991.
8. Masaya Yamaguchi, Nobuo Takiguchi, Yoshiyuki Kotani, Hirohiko Nisimura, An Interactive Method of Semantic Ambiguity in Sentences by Selection of Examples, Proc. of Natural Language Processing Pacific Rim Symposium '93, 1993.

Appendix. Example of Syntax Tree



John and Mary SUBJ sheet OBJ correct answer with compare

John and Mary compare their sheet with the correct answer.