OLD MEN AND WOMEN IN GLASSES :
A HEADACHE FOR COMPUTATIONAL LINGUISTS

Abstract
The issue of automatic processing of coordinate constructions (mainly NPs) with pre- and/or post-modifiers is investigated. The paper describes the CASSEX package which is capable of conjunction processing, illustrated with a group of English sentences containing modified coordinate NPs which are parsed and translated into Chinese by the package.

0. Introduction
For a machine translation system to be fully automatic, one task it faces is to resolve the ambiguities caused by coordinate constructions. None of the existing systems can claim a satisfactory resolution of this problem, though. For instance, in the machine translation (MT) system designed by (Nagao et al 82), when dealing with conjunctions, only the nearest two items of the same parts of speech were processed, while the following types of coordinated constructions were not analysed correctly:

(noun + prep + noun) + and + (noun + prep + noun);
(adj + noun) + and + noun..

(*) Present mailing address: Cognitive Studies Centre, University of Essex, Colchester CO4 3SQ, England.
For those MT systems with heavy reliance on human translators (such as WEIDNER or ALPS or LOGOS), the analysis is even less intelligent: anyway, there are intelligent people there to do the necessary modifications.

We can see the problem by examining the following example:

(1) old men and women in glasses

which could be interpreted in four different ways and accordingly result in four translations in the target language (Chinese in our case):

(1a) old men and women in glasses
    lao nanren he dai yanjing de nuren
    OLD MEN AND WEAR GLASSES particle WOMEN

(1b) old men, and old women in glasses
    lao nanren he dai yanjing de lao nuren
    OLD MEN AND WEAR GLASSES OLD WOMEN

(1c) old men in glasses, and women in glasses
    dai yanjing de lao nanren he (dai yanjing de)nuren
    WEAR GLASSES OLD MEN AND WEAR GLASSES WOMEN

(1d) old men in glasses, and old women in glasses
    dai yanjing de lao nanren he (dai yanjing de
    WEAR GLASSES OLD MEN AND WEAR GLASSES
    lao) nuren **
    OLD WOMEN

How to get the most likely reading out of such a bunch of possible ones? That's what this paper will be devoted to.

For conjunction processing in general, the reader is referred to (Woods 73, (Church 80), (Blackwell 81),

(**) (1c) and (1d) could possibly be translated into one and the same expression, which is ambiguous itself in Chinese.
(Nagao et al 82), (Boguraev 83) and (Huand 83). Here I will state a set of rules suggested in (Huang 83).

I. CSDC RULES

The following ordered rules, named CSDS (Conjuncts Scope Determination Constraints), are suggested and embodied in an ATN package so as to meet the need for automatically deciding the scope of the conjuncts:

1. Syntactical constraint.

The syntactical constraint has two parts:

1.1 The conjuncts should be of the same syntactical category;
1.2 The coordinated constituent should be in conformity syntactically with the other constituents of the sentence, e.g., if the coordinated constituent is the subject, it should agree with the finite verb in terms of person and number.

According to this constraint, the sentence

(2) The man with the child and the woman are kicking the ball.

should be analysed as follows (the representation is a tree diagram with 'CLAUSE' as the root and centred around the verb, with various case nodes indicating the dependency relationships between the verb and the other constituents):

```
(CLAUSE
 (TYPE DCL)
 (QUERY NIL)
 (TNS PAST)
 (ASPECT NIL)
 (MODALITY NIL)
 (NEG NIL)
 (V (KICK
 (......)
```


(OBJECT ((BALL 1 ....) )
 (NUMBER SINGLE)
 (QUANTIFIER SG)
 (DETERMINER ((DET 1 ONE)))

(AGENT
 (CONJUNCTION AND)
 ((MAN ...)
 (NUMBER SINGLE)
 (QUANTIFIER SG)
 (DETERMINER ((DET 1 ONE)))
 (ATTRIBUTE ((PREP (PREP WITH))
 (CHILD ....)
 (NUMBER....) ...) 
 (WOMAN....) ... ))))

while

(3) The man with the child and the woman with the um-
brella is kicking the ball.

and more generally,

(4) The man with the child and the woman kicked the
ball.

should be analysed roughly as :

(AGENT
 ((MAN ...)
 (NUMBER SINGLE)
 (QUANTIFIER SG)
 (DETERMINER ((DET 1 ONE)))
 (ATTRIBUTE ((PREP (PREP WITH))
 (CONJUNCTION AND)
 (POBJ
 ((CHILD   )
 (NUMBER   ) ... )
 ((WOMAN    ) ... ))))))
2. Semantic constraint.

NPs whose head noun semantic primitives (Wilks 75) are the same should be preferred when deciding the scope of the two conjuncts coordinated by "and". However, if no such NPs can be found, NPs with different head noun semantic primitives are coordinated anyhow.

According to rule 2,

(5) The man with the telescope and the umbrella kicked the ball.

should be roughly represented as 'The man with (AND (the telescope) (the umbrella))';

(6) The man with the telescope and the umbrella with a handle kicked the ball.

as 'The man with (AND (the telescope) (the umbrella with a handle))';

(7) The man with the telescope and the woman kicked the ball.

as '(AND (the man with the telescope) (the woman))' and

(8) The man with the telescope and the woman with the umbrella kicked the ball.

as '(AND (the man with the telescope) (the woman with the umbrella))'.


When Rules 1 and 2 are not enough for deciding the scope of the conjuncts, as for (4) and

(9) The man with the child and the woman with the umbrella kicked the ball.
this rule of preferring conjuncts with symmetrical pre-modifiers and/or post-modifiers will be in effect:

(4) ... with (AND (the child) (the woman)) ...
(9) (AND (the man with ...) (the woman with ...)) ...


If all the three rules above cannot help, the NP to the left of "and" which is closest to the coordinator should be coordinated with the NP immediately following the coordinator. Thus, for

(10) The man with the child and the umbrella fell.

we get "The man with (AND (the child) (the umbrella) fell."

II. ONE MORE CUE

As (Blackwell 81) observed, the presence of a new determiner (in the conjunct following "and"), even if it is identical with the previous one, marks the following noun as autonomous. Therefore, in her example

(11) the wooden table and chair,

"wooden" does not extend to modify "chair". The same holds for post-modifiers. So, if there is a determiner in front of "women" in our example, making it like

(12) the old men and the women in glasses,

then only "the women" should be thought in glasses, and only "men" are old.

III. PROCESSING MODIFIED NPS

Intuitively, we would think (1d) was the most likely reading of (1). The fact that "men and women
in glasses" is normally understood as "men (in glasses) and women in glasses" and "old men and women" as "old men and (old) women" supports our intuition.

Now if we put "old men and women in glasses" into a sentence (because our system deals with sentences only), say, "The old men and women in glasses kicked the ball", and apply the CSDC rules, what will happen?

In S/ state, the arc PUSH NP will be followed. After the NP "the old men" is built (in the subtree form[(MAN ...) (STATE (OLD1 ...)) (NUMBER PLURAL) (QUANTIFIER PL) (DETERMINER DET1)]), CONJ/ state is entered where the second NP "old women" ([[(WOMAN ...) (NUMBER PLURAL) ... (ATTRIBUTE ((PREP (PREP WITH)) (GLASS ...)))]]) is built. Now before the two NPs are coordinated to construct a new one, the STATE and DETERMINER slots are copied to the right conjunct and the ATTRIBUTE slot to the left one so that when the whole processing finished, we get the final result:

(AND (the old men in glasses) (the old women in glasses) kicked the ball,

which is in accordance with our intuition as well.

Admittedly, the CSDC rules alone would be insufficient to deal with structures like

(13) the old men and women suckling babies,

for whose resolution the semantic specialist should be called to check, then reject the connection between "men" and "suckling babies" and that between "women" and "old", because "suckle" prefers young and female (and animate, of course) subject.

One other problematic case is (Bar-Hillel 1962)'s

(14) slow neutrons and protons

which might be wrongly parsed as "(AND (slow neutrons) (slow protons))" by the package. Here what we need is a
kind of world knowledge (that of physics in our case) which can be incorporated into the package (in fact, the most straightforward way of doing that is collapsing "slow neutron(s)" into "slow/neutron(s)" in the preprocessing stage using certain rules and the data stored in a "fixed expression dictionary". If we haven't equipped the system with this kind of low level expertise, the system will accordingly behave exactly as a layman native speaker would do.

IV. CONCLUSION

CASSEX package is a parser based on (Boguraev 79)'s work. The original system is a combination of ATN grammars (Woods 73) and Preference Semantics (Wilks 75) and its aim is to resolve linguistic ambiguities. Conjunctions, one major source of ambiguities, however, were not taken into consideration. As far as the problems investigated are concerned, the system performs impressively, giving for almost all the ambiguous input sentences the most likely reading; and yet giving all the plausible readings if genuine ambiguity exists.

CASSEX, a developed, version of the system, is featured with its ability of processing automatically coordinate constructions. Equipped with the Chinese sentence generation facility, it makes a promising basis for building a fully automatic high quality (FAHQ) English-Chinese MT system.

The package is written in RUTGERS-UCI LISP and is implemented on the PDP-10 computer at the University of Essex. The above example sentences are correctly parsed by the package into the semantic representations, and then Chinese sentences are generated from those representations. Future work will be devoted to enlarging the vocabulary (at the moment it contains only a few hundred entries) and to adding more facilities (anaphora resolution, more sophisticated preposition phrase assignment, etc.) so that it can
become a working MT system in a couple of year's time.

Acknowledgements

I would like to thank Prof. Wilks and my colleague Fass of the Department of Language and Linguistics of the University of Essex for their help during my work on the package.

REFERENCES


