the nouns denoting "inalienable property" (parts of the body, relatives etc.) a special feature (iprop) in the lexicon and inserting the appropriate reflexive pronoun on the transfer stage.

Another translation problem is caused by the fact that Russian possessive reflexive pronouns often have to be translated as *of (one's) own* or Swedish *egen/eget* (own) when stressed, as in *u menja svoja mašina* (at me own car – I have a car of my own). Difficulties like these can to a certain extent be handled on the transfer stage, e.g. by rules like: if a Russian possessive reflexive pronoun is found in a possessive construction *u X Y*, like *u menja svoja mašina*, it should be treated as emphasized and translated into Swedish/English as *egen/eget* respectively phrases with *own*. As can be seen, reflexivization poses many interesting problems to linguistic theory and to machine translation.

REFERENCES

Diderichsen, P. 1939. 'Om Pronominerne *sig* og *sin*'. Acta Philologica Scandinavica 13.

Gawrońska-Werngren, B. 1990. 'Translation great problem – on the problem of inserting articles when translating from Russian into Swedish'. Proc Coling 1990. Helsinki (to appear)

Shopen, T. (ed). 1985. *Language typology and syntactic description*, vol. 1: Clause structure. Cambridge University Press.

Sigurd, B. 1987. Referent Grammar. A generalized phrase structure grammar with built-in referents. *Studia Linguistica* 41, 115-35.

Sigurd, B. & B. Gawrońska-Werngren, 1988. 'The Potential of SWETRA - A multilanguage MT system'. *Computers and Translation* 3, 237-50.

Wellander, E. 1946 (3rd ed). *Riktig svenska*. Stockholm: Svenska Bokförlaget.

Zribi-Herz, A. 1989. 'Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse'. *Language* 65:4, 695-727.

Lund University, Dept. of Linguistics
Working Papers 36 (1990), 151-162

Understanding Coordination by Means of Prolog

Bengt Sigurd and Per Warter

Abstract
Coordinated structures are very frequent in texts, but generally grammatical theories have little to say on the subject. This paper describes the different types of coordination and shows how they can be analyzed - and understood - using Prolog. The work has been carried out within the automatic translation project SWETRA (Sigurd & Gawrońska-Werngren 1988).

INTRODUCTION

Almost all sentence constituents can be coordinated, and coordination is very common in texts, a fact which is not reflected duly in the treatment of coordination in traditional or modern grammar. The following sentence illustrates coordination of phrases and words:

The rich, nice and beautiful (adjectives)
boys and girls (nouns)
who live in Florida and in New York (prep phrases)
can and do (auxiliaries)
give or throw (main verbs)
parties
when and where they want (subjunctions)
over and over. (lexicalized coordinated phrase)

Coordination of clauses may be illustrated by the following sentence: *The boy, whom the parents loved and (whom) the neighbours hated, took the car and drove to New York in order to work and (to) have fun.* This sentence illustrates coordinated relative clauses, coordinated main clauses, and coordinated infinitives, the latter with optional *to*. The sequence: *took the car and drove to New York* can alternatively be considered as a case of coordinated verb phrases.

There are at least some words which can hardly be coordinated, however, e.g. articles: *?A or the boy may come.* Similarly, some pronouns can hardly be coordinated with anything, e.g. *who*: *?The boy, who or that* *came*, and *which* : *?The house which or that I like*. The reason seems to be that *who* and *that* and *which* and *that* have the same referent and so they do not indicate any semantic difference. Normally each nominal constituent of a coordinated NP has a referent of its own. This seems to hold true above for nominal phrases with "nominal referents" (Gawrońska-Werngren 1990). For obvious reasons, it is also hard to coordinate coordinators without moving into a kind of metalanguage, as in: *?The boy sang and or or laughed*. (In more technical texts, such sentences may be rendered by: *The boy sang and/or laughed.*

The standard rule for coordination is that constituents of the same category may be coordinated. This is certainly true, but there are more or less clear cases of coordination of phrases of different categories as illustrated by the following examples: *Kim is a republican and proud of it, Lee is on his way up and looking to take over, They decided to leave and without losing a minute* (Sheila Dooley Collberg, personal communication).

Coordination exists in all natural languages and it is thus a characteristic of human language (and human thinking). The markers (coordinators) are typically small words, such as: *and*, German *und*, Swedish *och*, Russian *i*, Latin *et* and *-que*. Statistically, coordinators are short, as one would expect given these words, which are among the top ten on the word frequency lists of most languages studied so far.

We note that the Swedish coordinator *och* is etymologically related to Modern Swedish *öka* and Old Norse *aukian*, both meaning 'increase, add'. English *and* is related to German *und* and an Old Indian word meaning 'then'. Comprehensive studies are needed in order to find out whether words for 'increase' or 'then' are common origins of coordinators in the languages of the world. One problem with discovering the etymologies of coordinators is that they often are short and reduced forms. Often there are several etymological possibilities, as illustrated by the discussion of the origin of Swedish *eller* (Hellqvist 1980).

English *and* illustrates the coordinator as an infix operator between the conjuncts. In Latin there is both the marker *et* and the enclitic postfix marker *-que*, as illustrated by: *SENATUS POPULUSQUE ROMAE* (the senate and people of Rome).

Coordinators may also be discontinuous, as illustrated by: *both...and*, *either...or, neither...nor*. In that case the prefix part is a word which specifies the number of constituents to be coordinated (and a scope indicator). Most languages seem to have a way of indicating that the number

of coordinated items is 2 (dualis), which may be of special importance to the speakers. We also note that there may be special negative discontinuous coordinators (*neither...nor*) in languages. Swedish has both *varken... eller* and *vare sig... eller* to convey the same meaning.

In Hungarian *both A and B* is expressed as *A is B is* (i.e. literally *A and B and*), where the coordinator *is* is found after each term.

The first (prefix) part of a discontinuous coordination marker adds emphasis. Often, this first part is also seen to indicate whether the use of *or* is exclusive or inclusive.

As mentioned already, traditional or modern standard reference grammars and grammatical theories do not generally devote much space to coordination, but there are some comprehensive books where examples as those mentioned and many more are treated in depth (see references). These also study the semantics of coordination and relate the linguistic expressions to logic.

This paper will demonstrate how coordinated constructions can be interpreted by a computer program written in Definite Clause Grammar, a formalism available in most Prolog implementations. We will mainly use English adjective phrases for illustration. The generalization to NP, VP, PP and even S is obvious. The work is part of the research in the project Swetra (Swedish Computer Translation Research: Sigurd & Gawrońska-Werngren 1988) supported by the Swedish Council for Research in the Humanities and Social Sciences, and we hereby acknowledge the help of Mats Eeg-Olofsson and Barbara Gawrońska-Werngren in the Swetra group at Lund.

TYPES OF COORDINATION, SEMANTIC REPRESENTATIONS AND PROLOG RULES

In general there are two ways to deal with coordination (Goodall 1987, pp. 17-19). The first is often called "derived conjunction", and grammars which apply derived conjunction contain one general rule which conjoins sentences along with rules which delete identical elements in the sentences. The sentence:

The girls and the boys went into the garden.

can be seen as the conjunction of the two sentences:

The girls went into the garden.

The boys went into the garden.

making up the sentence (the elements to be deleted in brackets):

The girls (went into the garden) and the boys went into the garden.

The other way to handle coordination is the so-called "phrasal" or constituent coordination, which requires rules such as the following, where *co(C)* is a coordinator:

```
np(co(C,A,B)) --> np(A),co(C),np(B).
vp(co(C,A,B)) --> vp(A),co(C),vp(B).
```

We will not adopt the first method here, because we do not want to use any deletion rules. Implementing phrasal coordination, however, seems to be quite easy in a DCG-grammar, and we will investigate in detail more how phrasal coordination can be handled in Prolog below.

Prolog rules and coordination

A first approach to coordination is to give each coordinator a distinct meaning representation and to create a rule which regulates the linearization of coordinator and conjuncts, such as:

```
ap(A) --> a(A).
ap(co(C,A,B)) --> a(A),co(C),ap(B).
```

Corresponding lexical rules may then be added:

```
co(and) --> [and].
co(or) --> [or].
co(',') --> [','].
a(big) -->[big].
a(nice) --> [nice].
a(strong) --> [strong].
a(great) --> [great].
```

In Prolog the variables *A*, *B*, and *C* are replaced by the meaning of the analysed components. These rules will therefore analyse an adjective phrase such as:
[big,',',nice,and,strong]
as
co(',',big,co(and,nice,strong));
and

[big,and,nice,and,strong,and great] will be analysed as
co(and,big,co(and,nice,co(and,strong,great))).
The structure demonstrated above has two disadvantages: on the one hand, the structure is unnecessarily complicated, and on the other hand, the representation of the comma is not quite appropriate because the comma can represent *or* or *and*. This can pose problems when translating to a language in which such a replacement of a coordinator by a comma is not possible.

This prompted us to suggest a semantic representation which also includes some of the stylistic properties of coordination. "Understanding" means representing in some other form, and the form chosen here is a predicate logic prefix form. The "meaning" of the phrase *big or strong* is now represented as:

co(or,m,[big,strong]),

and the "meaning" of the phrase: *nice,big, strong* as:

co(and,a,[nice,big,strong]).

For possible use in a translation project we have to find some means of representing the stylistic properties of the coordination as asyndetic and polysyndetic, and so the letter *m* denotes 'monosyndetic' and *a* 'asyndetic' coordination (see below). Such representations can easily be changed into another form, e.g. or(big,strong), and(nice,big,strong), if required; but these representations carry less information about the surface structure. (Special precautions are needed if the prefixes "and" and "or" are to be used as predicates, as they are sometimes built-in predicates in Prolog).

Types of coordination

The following are the main types of coordination (illustrated by adjectives):

1. Monosyndetic coordination

The most common and convenient way to linearize (compress) a coordinated structure such as: A co B co C co ... co N (where *A*, *B*, *C* and *N* are the coordinated categories, and *co* is a coordinator such as *and* or *or*) is to replace all the coordinators, except the last, by a comma as in *nice, big and strong*. This last coordinator then indicates the nature of the coordination, which can be conjunction or disjunction. We call this type of coordination 'monosyndetic', as there is only one "overt" coordinator, and our way to represent it is by the formula co(Co,m,F), where *Co* is the coordinator, *m* the coordination type monosyndetic, and *F* a list of the categories to be coordinated.

The Prolog rules for monosyndetic coordination written in DCG-grammar could be written as follows (illustrated for adjectives):

(1)

(a) ap(co(C,m,F)) --> apm(C,F).
(b) apm(C,[F1,F2]) --> a(F1), co(C), a(F2).
(c) apm(C,[F1|F2]) --> a(F1), [','], apm(C,F2).

We assume further lexical rules such as:
a(big).
a(nice).
a(strong).
co(and).
co(or).

2. Polysyndetic coordination

A structure A co B co C ... co N like *nice and big and strong*, with no coordinator left out, is a stylistic means of bringing out each of the coordinated constituents as well as the type of coordination. A coordination of this type is given the representation: co(Co,p,F), where *p* denotes 'polysyndetic'.

A DCG-rule for a polysyndetic construction is very similar to the monosyndetic one, except for the last rule:

(2)

(a) ap(co(C,p,F)) --> app(C,F).
(b) app(C,[F1,F2]) --> a(F1), co(C), a(F2).
(c) app(C,[F1|F2]) --> a(F1), co(C), app(C,F2).

3. Asyndetic coordination

A coordination in which all coordinators are replaced with commas, as in *nice,big,strong*, is called asyndetic. A well-known example is Latin: *Veni, vidi, vici*. The coordinator which is left out must be conjunctive (*and*). Such a structure receives the representation: co(and,a,[nice,big,strong]).

The following DCG rules cover this case:

(3)

(a) ap(co(and,a,F)) --> apa(F).
(b) apa([F1,F2]) --> a(F1), [','], a(F2).
(c) apa([F1|F2]) --> a(F1), [','], apa(F2).

4. A special case of asyndetic coordination?

In adjective phrases there is a special case where the last adjective and the noun make up a separate unit and have a composite meaning as in *the young American girl*. We call it the zero case, because there is no coordinator and no comma and, if a representation is required, we suggest co(and,0,F), even if this is not quite appropriate since the adjectives are not completely on the same level. A characteristic feature of this type is that a coordinator cannot be inserted. A more appropriate way of treating this case would perhaps be to say that there are head nouns consisting of an adjective and a noun. But as we want a unified treatment, we shall treat this case as a special syntactic type of coordination.

The DCG-rules could be the following:

(4)

(a) ap(and,0,F) --> ap0(F).
(b) ap0([F1,F2]) --> a(F1), a(F2).
(c) ap0([F1|F2]) --> a(F1), [','], ap0(F2).

5. Complex coordination

Normally a coordination is only of one semantic type (conjunctive or disjunctive) and only includes one type of coordinator. A coordination may, however, be more complex and include several types of coordination and coordinators, e.g. *nice and big or strong*. Such complex strings are generally ambiguous. In that case the string has two possible representations (disregarding the type of coordination):

co(or,co(and,nice,big),strong) or
co(and,nice,co(or,big,strong)).

The following rules cover such complex cases:

(5)

```
(a)    ap(F) --> apc(F,2).
(b)    apc(F,_) --> a(F).
(c)    apc(co(C,F1,F2),N) -->
           {N>0, N1 is N-1},
           apc(F1,N1), co(C), apc(F2,N1).
```

Rule (5b) says that a complex coordinated structure can be an adjective; (5c) that it can consist of two ordinary structures with a coordinator in between. It is necessary to restrict the recursion depth in this case, as the rule is left-recursive, so within (5c) there is a condition which says that N must be greater than zero. N initializes to 2 in (5b), and every time rule (5c) is called (recursively) N counts down (N is N-1), so that the greatest recursion depth is 2. The figure 2 is arbitrary, but it serves to indicate the restricted capacity of the human brain in processing complex coordinations. The rule above cannot process structures more complex than: *nice or big and strong or wise* - which seems to be sufficient.

Up to now we have only treated the coordinators *and* and *or*, but it is easy to add codings of discontinuous coordinators such as *either...or, neither...nor,* and *both...and*: The following DCG rules cover such cases:

```
(d)    apc(co(either-or,F1,F2),N) -->
           {N>0, N1 is N-1},
           [either], apc(F1,N1), [or], apc(F2,N1).
(e)    apc(co(both-and,F1,F2),N) -->
           {N>0, N1 is N-1},
           [both], apc(F1,N1), [and], apc(F2,N1).
(f)    apc(co(neither-or,F1,F2),N) -->
           [either], apc(F1,N1), [or], apc(F2,N1).
```

Besides fulfilling a certain intensifying function, the first component serves as a scope indicator. The phrase *nice and big or strong* is ambiguous in that one does not know whether it is *nice* and *big* which form a disjunction with *strong* or whether it is *big* or *strong* which form a conjunction with *nice*. The phrase *either nice and big or strong* is not ambiguous, however, as *either* here indicates the scope of *or*. The representation of this becomes co(either-or, co(and,nice,big),strong).

The most interesting feature of coordination is its recursive character. The phrases: *strong, big, and nice* can be prolonged infinitely by adding new adjectives after *big*. This is taken care of by our Prolog rules.

Tentative investigations and intuition indicate that the most common types are: *big and strong* (two members, a pair), and *big, strong and nice* (three members).

COORDINATION AND ELLIPSIS

A constituent in a coordination is often reduced in that certain elements are deleted or elided (cf. derived conjunction above). This holds true above all for sentences (clauses) and natural languages thus utilize the possibilities of economic compression. There are interesting differences between languages in the possibilities of deleting constituents: 'forward or backward gapping' (Ross 1970). The following are the main types.

Clause ellipsis:
Bill ran and (Bill) laughed (subject forward ellipsis)
Bill hit the ball and Sue (hit) the window (predicate forward ellipsis)
Bill hit (the window) and Sue crashed the window (backward object ellipsis)

Prepositional phrase ellipsis:
The boy from Australia and the girl (from Australia) (forward postnominal attribute ellipsis)

We will not study ellipsis any further here nor show how such cases are taken care of in Referent Grammar, however.

CONCLUSION

In conclusion we give a test run (demo) which shows how various coordinated adjective structures are analysed and understood according to our rules. As can be seen the program is capable of analyzing a wide range of coordination phenomena. The representations derived also indicate the stylistic properties of the expressions.

We give the whole program in an appendix. When the user writes e.g. *analysis([nice,',',big,and,strong]).* the program responds *co(and,p,[nice,big, strong]).* If instead *regen* is used, the semantic representation is printed and then the same phrase is regenerated from this semantic representation.

TEST RUNS (DEMO)

```
analysis([nice,´,´,big,´,´,strong]).
co(and,a,[nice,big,strong])
Nº1 yes

analysis([nice,´,´,big,strong]).
co(and,0,[nice,big,strong])
Nº1 yes

analysis([nice,´,´,big,and,strong]).
co(and,m,[nice,big,strong])
Nº1 yes

analysis([nice,and,big,and,strong]).
co(and,p,[nice,big,strong])
Nº1 yes


analysis([nice,and,big,or,strong]).
co(or,co(and,nice,big),strong)
Nº1 yes;
co(and,nice,co(or,big,strong))
Nº2 yes;
no more solutions

analysis([either,nice,and,big,or,strong]).
co(either-or,co(and,nice,big),strong)
Nº1 yes;
no more solutions

analysis([nice,and,either,big,or,strong]).
co(and,nice,co(either-or,big,strong))
Nº1 yes;
no more solutions
```


APPENDIX. PROGRAM COORD

```
ap(co(C,m,F)) --> apm(C,F).
ap(co(and,0,F)) --> ap0(F).
ap(co(C,p,F)) --> app(C,F).
ap(co(and,a,F)) --> apa(F).
ap(F) --> apc(F,2).
ap(F) --> a(F).

apm(C,[F1|F2]) --> a(F1),[´,´],apm(C,F2).
apm(C,[F1,F2]) --> a(F1),co(C), a(F2).
app(C,[F1|F2]) --> a(F1),co(C),app(C,F2).
app(C,[F1,F2]) --> a(F1),co(C),a(F2).
apa([F1|F2]) --> a(F1),[´,´],apa(F2).
apa([F1,F2]) --> a(F1),[´,´],a(F2).
ap0([F1|F2]) --> a(F1),[´,´],ap0(F2).
ap0([F1,F2]) --> a(F1),a(F2).

apc(co(C,F1,F2),N) -->
   {N>0,N1 is N - 1}, apc(F1,N1),co(C),apc(F2,N1).
apc(co(either-or,F1,F2),N) -->
   {N>0,N1 is N-1},[either],apc(F1,N1),[or],apc(F2,N1).
apc(co(both-and,F1,F2),N) -->
   {N>0,N1 is N-1},[both],apc(F1,N1),[and],apc(F2,N1).
apc(co(neither-nor,F1,F2),N) -->
   {N>0,N1 is N-1},[neither],apc(F1,N1),[nor],apc(F2,N1).
apc(F,_) --> a(F).

co(and) --> [and].
co(or) --> [or].

a(big) --> [big].
a(strong) --> [strong].
a(nice) --> [nice].
a(great) --> [great].


/* Interactions for analysis and regeneration */
analysis(X) :- ap(F,X,[]),write(F),nl.
regen(X) :- ap(F,X,[]),write(F),nl,ap(F,Y,[]),write(Y),nl.
```

REFERENCES

Cresswell, M. 1973. *Logics and Languages*. London: Methuen.

Gawrońska-Werngren, B. 1990. 'Translation great problem - on the problem of inserting articles when translating from Russion into Swedish'. *Proc. Coling* 1990. Helsinki (to appear).

Gazdar, G. & Pullum, G. 1976. 'Truth-functional connectives in natural language'. *Chicago Linguistic Society* (CLS) 12, 220-234.

Gazdar, G. 1980. 'A cross-categorial semantics for coordination'. *Linguistics and Philosophy* 3 , 307-409.

Gazdar, G. 1981. 'Unbounded dependencies and coordinated structure'. *Linguistic Inquiry* 12, 155-184.

Goodall, G. 1987. *Parallel Structures in Syntax. Coordination, Causatives, and Restructuring*. Cambridge: Cambridge Univ. Press.

Hellqvist, E. 1980. *Svensk etymologisk ordbok*. (4:th ed). Lund: Liber.

Hesse, H. & Küstner, K. 1985. *Syntax der koordinativen Verknüpfung*. Berlin: Akademie-Verlag.

Lang, E. 1984. *The Semantics of Coordination*. Studies in Language Companion Series (SLCS) 9. Amsterdam: John Benjamin.

Mallinson, G. & Blake, B. J. 1981. *Language typology. Crosslinguistic Studies in Syntax*. Amsterdam.

Ross, J. 1970. 'Gapping and the order of constituents'. *Papers in linguistics*., eds. M. Bierwisch & K. Heidolph. The Hague: Mouton.

Sigurd, B. 1987. 'Referent grammar (RG). A generalized phrase structure grammar with built-in referents'. *Studia Linguistica* 41:2, 115-135.

Sigurd, B. & Gawrońska-Werngren, B. 1988. 'The potentials of SWETRA, a multilanguage MT-system'. *Computers and Translation* 3, 238-250.

Lund University, Dept. of Linguistics
Working Papers 36 (1990), 163-190

Semiotic Play : A Child Translates Text into Pictures

Ragnhild Söderbergh

INTRODUCTION

Much time and energy has been devoted to devising clever experiments in order to find out about children's linguistic abilities. There always remains a feeling of uncertainty, however, as to whether the test really captured what it was was intended to - that , after all, the fish did not slip through the meshes of the net. Superior to the specimens elicited by researchers in laboratories or on casual visits in children's homes are the spontaneous examples found, for instance, in parents' diaries, where the parent is also a trained scholar with a sensitive ear and eye and the rich experience of the child that only a shared everyday life can give.

This paper has been based on a diary documenting a girl's learning to read and her reading, from two to seven years of age, and on some of her spontaneous drawings that have been dated at the time of their production and saved. The corpus of pictures presented and analyzed here were spontaneously made during a period of two months - November and December 1968 - as illustrations of episodes from texts that she had earlier read and which she reread before she produced the drawings. At the time of the production of these drawings the girl was 5 1/2 years old. In comparison with her other drawings from the same period, the book-illustrating ones are much more stereotyped and conventionalized. A close analysis indicates that they are transformations of the written text into a pictorial mode. As such they not only give interesting information about the child's comprehension of the texts and her linguistic abilities, but also invite more speculative comparisons with pictorial precursors of written language in the history of writing.