

TR-IT-0036

Syntactic trees for the sentences of ten dialogues

Yves Lepage and Laurel Fais

January 1994

Abstract

This report lists sentences of ten of the ATR dialogues with their associated syntactic trees. The choices made for syntactic representation are explained. These choices have been consistently adopted throughout the corpus.

Keywords

ATR dialogues, syntactic tree, VP node.

©ATR Interpreting Telecommunications Research Laboratories

Contents

Introduction	9
1 Representational choices	11
1.1 Morpho-syntactic categories	11
1.2 Phrase types	12
1.3 Noun phrases	13
1.3.1 Addresses	13
1.3.2 Names	13
1.4 Verb phrases	13
1.4.1 Verb types	13
1.4.2 Past	14
2 The data	15
2.1 Data	15
2.2 Programs	15
3 The syntactic trees	19
3.1 Interjections	19
3.2 Agreements, Confirmation, Denial	20
3.3 Copulas	22
3.4 Declarative sentences	25
3.5 Negative declarative sentences	31
3.6 Interrogative sentences	33
3.7 Negative interrogative sentences	40
3.8 Incitative sentences	41
3.9 Confirmatory sentences	44
3.10 Infinitive clauses	45
3.11 Relative clauses	48
3.12 Subordinate clauses	50
3.13 Coordinate clauses	56
A Appendix: Against the VP node	61
Bibliography	65

List of Figures

1	Dialogue 1: Sentence 1: "Hello."	19
2	Dialogue 1: Sentence 3: "Yes."	19
3	Dialogue 1: Sentence 10: "No."	19
4	Dialogue 1: Sentence 11: "Not yet."	20
5	Dialogue 1: Sentence 12: "I see."	20
6	Dialogue 1: Sentence 19: "Thank you very much."	20
7	Dialogue 1: Sentence 20: "Well then, good-bye."	21
8	Dialogue 4: Sentence 16: "All right."	21
9	Dialogue 1: Sentence 4: "That 's right."	22
10	Dialogue 1: Sentence 16: "My name is Mayumi Suzuki."	22
11	Dialogue 2: Sentence 2: "This is the Conference Office."	22
12	Dialogue 2: Sentence 19: "You 're welcome."	23
13	Dialogue 3: Sentence 9: "It 's English and Japanese."	23
14	Dialogue 4: Sentence 8: "The attendance fee is 40,000 yen."	23
15	Dialogue 4: Sentence 12: "My name is Mayumi Suzuki."	23
16	Dialogue 4: Sentence 13: "My address is 2-27-7 Tamatsukuri, Higashi-ku, Osaka."	24
17	Dialogue 4: Sentence 17: "It 's 372-8018."	24
18	Dialogue 9: Sentence 6: "I 'm at Kyoto Station now."	24
19	Dialogue 1: Sentence 13: "Well then, I 'll send you a registration form."	25
20	Dialogue 1: Sentence 18: "I 'll send you a registration form immediately."	25
21	Dialogue 2: Sentence 6: "At present the attendance fee is 35,000 yen per person."	26
22	Dialogue 2: Sentence 8: "The proceedings and the reception fee are included in the attendance fee."	26
23	Dialogue 2: Sentence 9: "I 'm a member of the Information Processing Society."	27
24	Dialogue 2: Sentence 14: "You should pay the attendance fee with a bank-transfer."	27
25	Dialogue 3: Sentence 3: "I 'd like to present a paper at the Conference."	28
26	Dialogue 3: Sentence 5: "This Conference includes an extensive sphere of research related to Interpreting Telephony."	28
27	Dialogue 3: Sentence 6: "We also expect people who specialize in linguistics, psychology, and so on to attend."	29

28	Dialogue 3: Sentence 13: "We have made arrangements for a simultaneous translation into English."	29
29	Dialogue 4: Sentence 7: "The Conference will take place from August 22nd to 25th at the Kyoto International Conference Center."	30
30	Dialogue 5: Sentence 3: "I have a request."	30
31	Dialogue 2: Sentence 11: "There's no discount this time."	31
32	Dialogue 3: Sentence 10: "I do n't understand Japanese at all."	31
33	Dialogue 4: Sentence 5: "I do n't have it."	31
34	Dialogue 5: Sentence 15: "For cancellations after September 27th we can't make refunds."	32
35	Dialogue 5: Sentence 18: "As far as that goes, there is n't a special problem."	32
36	Dialogue 7: Sentence 19: "There is n't."	32
37	Dialogue 1: Sentence 2: "Is this the Conference Office?"	33
38	Dialogue 1: Sentence 5: "May I help you?"	33
39	Dialogue 1: Sentence 7: "What sort of procedure should I follow?"	33
40	Dialogue 1: Sentence 9: "Do you already have a registration form?"	34
41	Dialogue 1: Sentence 14: "May I have your name and your address, please?"	34
42	Dialogue 2: Sentence 3: "Could you tell me about the attendance fee for the Conference?"	35
43	Dialogue 2: Sentence 13: "How should I pay the attendance fee?"	35
44	Dialogue 3: Sentence 8: "By the way, what's the official language at the Conference?"	35
45	Dialogue 4: Sentence 3: "Do you have the Announcement of the Conference?"	36
46	Dialogue 5: Sentence 12: "Could you possibly refund the registration fee?"	36
47	Dialogue 5: Sentence 17: "Well then, can someone attend the Conference instead of me?"	36
48	Dialogue 6: Sentence 9: "How much is the attendance fee?"	37
49	Dialogue 6: Sentence 16: "Well then, may I have your name and the number of people in your party, please?"	37
50	Dialogue 7: Sentence 7: "Specifically, what sort of topic is this?"	37
51	Dialogue 7: Sentence 10: "Would you please take a look at it?"	38
52	Dialogue 7: Sentence 17: "Is there anything else?"	38
53	Dialogue 8: Sentence 11: "With what kind of form should I write the summary?"	38
54	Dialogue 9: Sentence 10: "About how much does it cost to go from Kyoto Station to the Conference Center by taxi?"	39
55	Dialogue 2: Sentence 10: "Is n't there an attendance fee discount?"	40

56	Dialogue 1: Sentence 8: <i>"Please use a registration form."</i>	41
57	Dialogue 2: Sentence 15: <i>"Please deposit it into the bank account which is indicated in the announcement."</i>	41
58	Dialogue 3: Sentence 4: <i>"Please tell me about the topic of the Conference."</i>	42
59	Dialogue 4: Sentence 2: <i>"Please tell me some details about the Conference."</i>	42
60	Dialogue 6: Sentence 20: <i>"Please pay the attendance fee that day at the meeting place."</i>	43
61	Dialogue 7: Sentence 12: <i>"Well then, please mail me the Announcement as soon as possible."</i>	43
62	Dialogue 4: Sentence 18: <i>"It 's 372-8018, right?"</i>	44
63	Dialogue 8: Sentence 17: <i>"It 's Mr. Mayumi Suzuki from ATR, right?"</i> . .	44
64	Dialogue 8: Sentence 18: <i>"Your address is 3-2-5 Higashi-Ikebukuro, Toshima-ku, Tokyo, right?"</i>	44
65	Dialogue 1: Sentence 6: <i>"I 'd like to apply for the Conference."</i>	45
66	Dialogue 4: Sentence 15: <i>"I 'd like to ask your phone number, too."</i>	45
67	Dialogue 5: Sentence 5: <i>"I 'd like to cancel my attendance."</i>	46
68	Dialogue 8: Sentence 4: <i>"I 'd like to make a presentation at this Conference."</i>	46
69	Dialogue 9: Sentence 5: <i>"I 'd like you to tell me how I should go to the Conference Center."</i>	47
70	Dialogue 10: Sentence 4: <i>"I 'd like to ask about hotel accommodations for the Conference."</i>	47
71	Dialogue 7: Sentence 3: <i>"I 'd like to ask some questions about the topics which will be treated at the Conference."</i>	48
72	Dialogue 7: Sentence 9: <i>"The titles of the papers which will be presented at the Conference are listed in the second version of the Announcement."</i> . . .	48
73	Dialogue 8: Sentence 3: <i>"There 's something I 'd like to ask."</i>	49
74	Dialogue 10: Sentence 7: <i>"The hotels we can suggest are the Kyoto Hotel and Kyoto Prince Hotel."</i>	49
75	Dialogue 2: Sentence 4: <i>"If I apply for the Conference now, how much is the attendance fee?"</i>	50
76	Dialogue 2: Sentence 7: <i>"If you apply next month, it will be 40,000 yen."</i> .	50
77	Dialogue 2: Sentence 20: <i>"If there 's something which you do n't understand, please ask any time."</i>	51
78	Dialogue 3: Sentence 11: <i>"In case a presentation is made in Japanese, is there a simultaneous translation into English?"</i>	51
79	Dialogue 4: Sentence 9: <i>"If you would like to make a presentation, please submit a summary by March 20th."</i>	51

80	Dialogue 5: Sentence 19: "In case a substitute will attend, please let me know in advance."	52
81	Dialogue 5: Sentence 21: "When a substitute is selected, I 'll let you know."	52
82	Dialogue 6: Sentence 3: "I 've heard that there is a city tour during the conference."	53
83	Dialogue 8: Sentence 8: "If your contribution was accepted, we 'll enclose manuscript forms."	53
84	Dialogue 9: Sentence 11: "If it 's from Kyoto Station, it costs about 6,000 yen."	54
85	Dialogue 9: Sentence 12: "Well then, if it 's from Kitaooji Station, about how much does it cost?"	54
86	Dialogue 9: Sentence 13: "If it 's from Kitaooji Station, it 's about 900 yen."	54
87	Dialogue 10: Sentence 28: "I 'll check if I can take a room."	55
88	Dialogue 4: Sentence 10: "I 'll send the Conference Announcement to you, so please have a look at it."	56
89	Dialogue 5: Sentence 13: "I 'm terribly sorry, we can 't."	56
90	Dialogue 7: Sentence 8: "I 'm terribly sorry, but I can 't answer any technical questions here."	57
91	Dialogue 8: Sentence 7: "We 'll review it here and we 'll send you the result by May 20th."	57
92	Dialogue 8: Sentence 12: "There 's a special application form, so please fill it in."	57
93	Dialogue 8: Sentence 13: "Well then, we 'll send you an application form, so may I have your address, please?"	58

Introduction

For their Interpreting Telephony project, the ATR team chose an illustrative scenario: someone calls a secretariat to get some information about an international conference.

A collection of simulated conversations were recorded and transcribed. With annotations of segmentation and syntactic categories, they constitute the ATR dialogue database.

For the design of the experimental prototype demonstrated on the 28th of January 1993, 12 dialogues were selected. They were numbered A-B, 1-10. There exists a version of these dialogues in Japanese, German and English.

To our surprise, there did not exist any complete syntactic description of each of the sentences of the dialogues. In a previous work [Lepage 93a], we automatically extracted the syntactic part from the feature structures produced by the analysis in the ASURA system. Unfortunately, they happened to contain inconsistencies and even errors of representations. Our feeling is that, any linguistic work on a corpus requires a predefined complete syntactic description, something like a tree bank. The goal of the work reported here is to fill this demanding lack. We intended this to be a reference for syntactic structures of the English ATR dialogue database.

This report lists a good part of the syntactic trees associated with English sentences of dialogues 1 to 10. These syntactic trees were designed by hand. The trees shown here are projective and do not contain any trace node. We insisted on consistency, and as a consequence, the same phenomenon is described by exactly the same representation throughout the corpus.

1 Representational choices

The conventions adopted for the trees are described in this section.

1.1 Morpho-syntactic categories

We used 14 morpho-syntactic categories. Here is an exhaustive list. The signification of the abbreviation is followed by examples.

- adj adjective *closer, your*
- adv adverb *yet, not, already*
- conj conjunction *if, and*
- copula copula *is, was, are*
- det determiner *a, the*
- interj interjection *yes, no, hello*
- modal modal verb *please, will, can, do*
- noun noun *Hotel, address, registration*
- post adjunct particle of a verb *in in fill it in*
- prep preposition *at, in, of*
- pron pronoun *it, you, what, when*
- prtp participle *listed, included*
- quant quantifier *any, some, all*
- verb verb *help, apply, follow*

Of course, the same word may be ambiguous with respect to the morpho-syntactic classes. Hence, for example, *what* may be adj or pron, as in the following sentences.

What sort of procedure should I follow?
What is it?

The previous categories apply to single words. As one of our main concerns was to keep exact projectivity, we were forced to decompose sequences such as *take place* or *at all*, although other analyses would consider them as single units bearing only one category (for example, the analysis of *well then* in [Fais 93, p. 25]). Reciprocally, for the same reason, we decomposed clusters like *don't* or *we'll* into two words:

1.2 Phrase types

We used 7 types of phrases. Here is a list of them with examples.

- AP adjectival phrase *closer to the Conference Center*
- AVP adverbial phrase *Not yet*
- NP noun phrase *the Kyoto Prince Hotel*
- PP prepositional phrase *at the Kyoto Prince Hotel*
- PRPP prepositional complex *instead of*
- S sentence *Please., Is this the Conference Office.*
- VP verbal phrase *help you, use a registration form*

The analysis of *please* was problematic. It is more a discourse element than a syntactical element [Fais and Kikui 91]. For this reason, preverbal *please* was analysed as a modal, and, post-sentential *please* was analysed as a sentence.

*Please tell me about the topic of the conference.
May I have your name and your address, please?*

This is different from the analysis described in [Fais 93] where *please* gets the category greeting-word, and special rules are added to authorise both *vp please* and *please vp* structures.

1.3 Noun phrases

1.3.1 Addresses

The representation of addresses is one of the most difficult problems as lack of order and optionality play a big role in these phrases. To make it simple we chose to flatten the tree and normalise it in the following pattern.

```
      NP
      |
-----|-----
adj noun noun noun noun
```

1.3.2 Names

In a similar way, names are represented by a flattened tree.

```
      NP
      |
---|---
noun noun
```

1.4 Verb phrases

Although we adopted a rather unusual representation for the past, we did not follow completely our conviction, and refrained from eliminating the VP nodes. Anyhow, good arguments against VP may ultimately put an end to the use of this label. Some are given in the appendix.

1.4.1 Verb types

We divided the set of all verbs into three subsets:

- copula this labels the forms of *to be* when it works as a copula. The main reason for this distinct node is that, considering it as a verb would have consequently led us to build a NP node and to introduce discontinuous structures for inversions, as in:

Is this the conference office?

- modal this stands for modals properly, as *can* for auxiliaries, as *will, do* in negatives, and verbal clause modifiers, as *please*. The reason for this node is similar to that for the node copula.
- verb this class collects all other verbs.

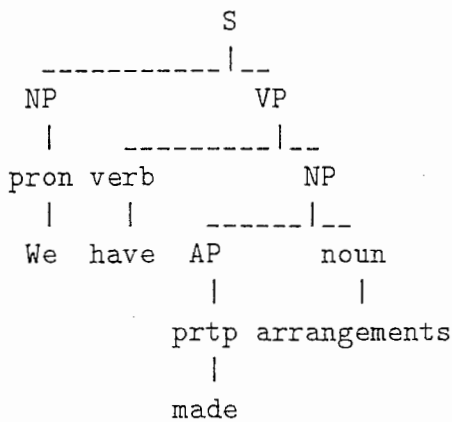
As was seen, the introduction of different nodes was mainly motivated by problems originating in the presence of a VP node. If it would have been only for us, we would have removed any VP node. The appendix gives a short linguistic discussion and references against the VP node. Until this view prevail in mainstream linguistics, we submitted to the prevailing representation and kept the VP nodes in the syntactic trees.

1.4.2 Past

We represented the past tense according to Zemb's representation. In the sentence

We have made arrangements...

the verb is *have* and *made arrangements* is a noun phrase which has to be analysed as an adjective phrase followed by a noun.



2 The data

2.1 Data

The trees are available on the machine as11. They are located under the directory

```
/export/as11/share/data/dialogue
```

as the files 1 through 10. Each line of these files contains two fields separated by a tabulation. The first one is the pure syntactic tree, *i.e.* without the words, in a parenthesised form. The second one contains the sentence enclosed in double-quotes. Words are separated by spaces. The escape character \ appears before commas and apostrophes.

Here is an excerpt of the file dialogue/3.

```
S(interj) "Hello."
S(NP(pron),copula,NP(det,NP(noun),noun)) "This is the Conference
Office."
...
S(NP(pron),VP(verb)) "I see."
S(VP(verb,NP(pron)),AVP(AVP(adv),adv)) "Thank you very much."
S(interj) "Good-bye."
```

2.2 Programs

The separation of the fields by a tabulation character allows the use of the standard UNIX commands cut and paste. For example, the following command lists the sentences of dialogue 3.

```
prompt> cd /export/as11/share/data
prompt> cut -f2 dialogue/3
"Hello."
"This is the Conference Office."
...
"I see."
"Thank you very much."
"Good-bye."
prompt>
```

Other programs for the manipulation of these data may be found under the /home/-lepage/bin/tree directory. These programs use features described in [Lepage 92a] and objects of the C toolbox documented by [Lepage 92b]. We give some examples in the following.

projForest produces the trees with words as leaves.

```
prompt> projForest < dialogue/3
S(interj(Hello.))
S(NP(pron(This)),copula(is),NP(det(the),NP(noun(Conference)),
noun(Office.)))
...
S(NP(pron(I)),VP(verb(see.)))
S(VP(verb(Thank),NP(pron(you))),AVP(AVP(adv(very)),adv(much.)))
S(interj(Good-bye.))
prompt>
```

2Dforest draws a forest given under parenthesised form. By using the -l command, which applies the same command on each line of a file, one can draw all the trees of a dialogue.

```
prompt> cut -f1 dialogue/3 | -l 2Dforest
```

```
  S
  |
interj

          S
-----|-----
NP  copula      NP
 |           -----|-----
pron         det  NP  noun
 |
noun

...

```

yieldForest gives the yield of a forest, that is the sequence of its leaves. Applied on the trees of the dialogue files, this gives the sequence of morpho-syntactic categories for each sentence.


```

prompt> cut -f1 dialogue/3 | -l 2Dforest
interj
pron,copula,det,noun,noun
...
verb,pron,adv,adv
interj
prompt>

```

enumsubTree enumerates all subtrees of a forest. Here is how to enumerate without repetition all subtrees in the structure of sentence 2 of dialogue 3.

```

prompt> head -2 < dialogue/3 | tail -1
S(NP(pron),copula,NP(det,NP(noun),noun)) "This is the Conference Office."
prompt> head -2 < dialogue/3 | tail -1 | cut -f1 | enumsubTree | sort -u
NP(det,NP(noun),noun)
NP(noun)
NP(pron)
S(NP(pron),copula,NP(det,NP(noun),noun))
copula
det
noun
pron
prompt>

```

These data have been used for experiments on a bi-directional system with a genetic algorithm engine. The system has been described in [Lepage 93a]. Results obtained with the data have been reported in [Lepage 93b] and [Lepage 93c].

3 The syntactic trees

Here is the listing of the syntactic trees for the ten dialogues. For each sentence, the number of the dialogue and the number in the dialogue are given first. Then, the corresponding tree is drawn, with the terminals as leaves. The sentence is also explicitly given.

3.1 Interjections

```
  S
  |
interj
  |
Hello.
```

Figure 1: Dialogue 1: Sentence 1: "Hello."

```
  S
  |
interj
  |
Yes.
```

Figure 2: Dialogue 1: Sentence 3: "Yes."

```
  S
  |
interj
  |
No.
```

Figure 3: Dialogue 1: Sentence 10: "No."

3.2 Agreements, Confirmation, Denial

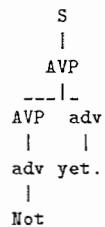


Figure 4: Dialogue 1: Sentence 11: "Not yet."

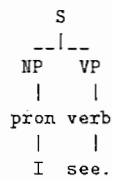


Figure 5: Dialogue 1: Sentence 12: "I see."

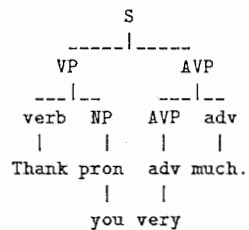


Figure 6: Dialogue 1: Sentence 19: "Thank you very much."

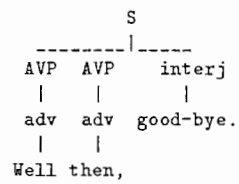


Figure 7: Dialogue 1: Sentence 20: "Well then, good-bye."

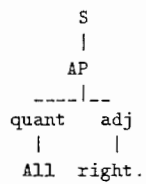


Figure 8: Dialogue 4: Sentence 16: "All right."

3.3 Copulas

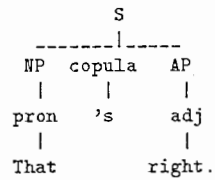


Figure 9: Dialogue 1: Sentence 4: *"That 's right."*

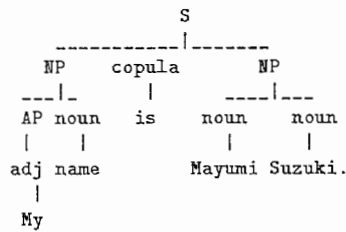


Figure 10: Dialogue 1: Sentence 16: *"My name is Mayumi Suzuki."*

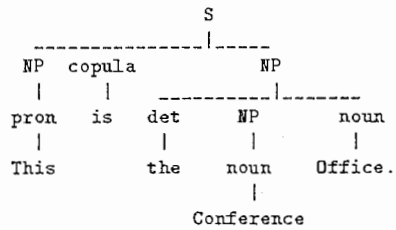


Figure 11: Dialogue 2: Sentence 2: *"This is the Conference Office."*

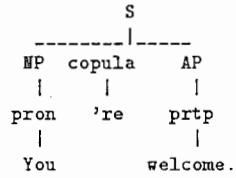


Figure 12: Dialogue 2: Sentence 19: "You 're welcome."

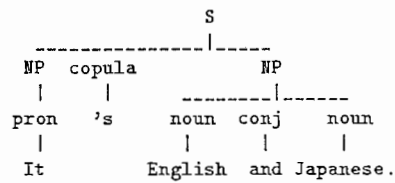


Figure 13: Dialogue 3: Sentence 9: "It 's English and Japanese."

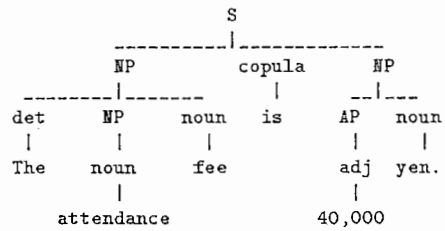


Figure 14: Dialogue 4: Sentence 8: "The attendance fee is 40,000 yen."

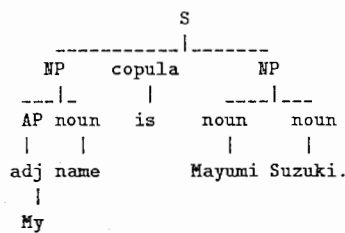


Figure 15: Dialogue 4: Sentence 12: "My name is Mayumi Suzuki."

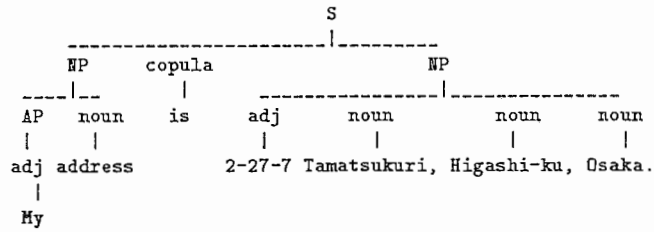


Figure 16: Dialogue 4: Sentence 13: "My address is 2-27-7 Tamatsukuri, Higashi-ku, Osaka."

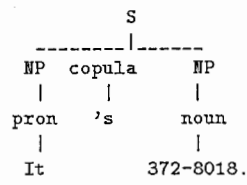


Figure 17: Dialogue 4: Sentence 17: "It 's 372-8018."

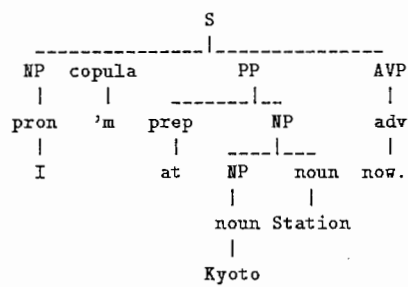


Figure 18: Dialogue 9: Sentence 6: "I 'm at Kyoto Station now."

3.4 Declarative sentences

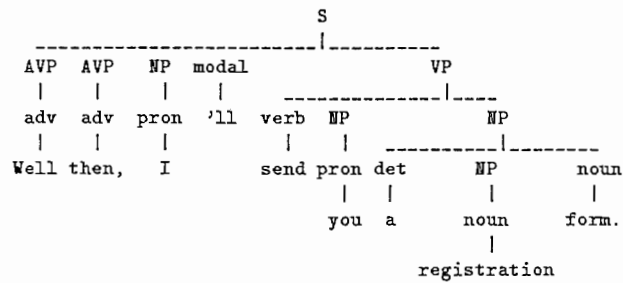


Figure 19: Dialogue 1: Sentence 13: "Well then, I 'll send you a registration form."

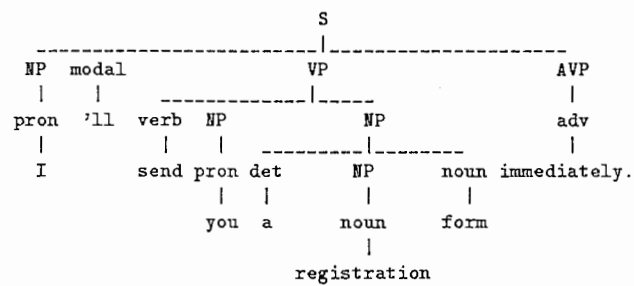


Figure 20: Dialogue 1: Sentence 18: "I 'll send you a registration form immediately."

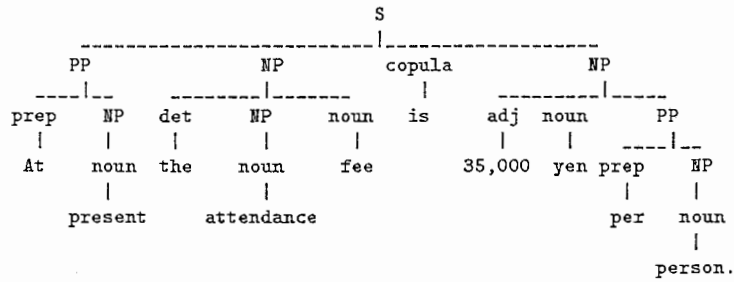


Figure 21: Dialogue 2: Sentence 6: "At present the attendance fee is 35,000 yen per person."

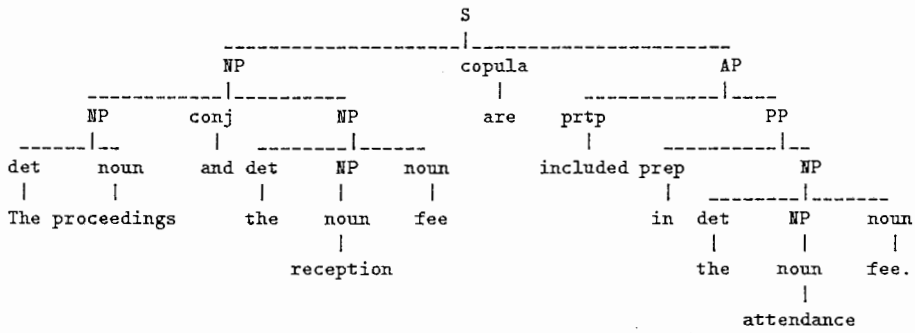


Figure 22: Dialogue 2: Sentence 8: "The proceedings and the reception fee are included in the attendance fee."

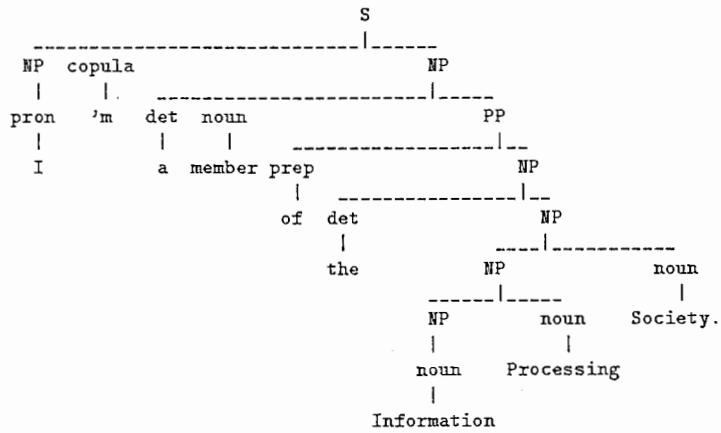


Figure 23: Dialogue 2: Sentence 9: "I 'm a member of the Information Processing Society."

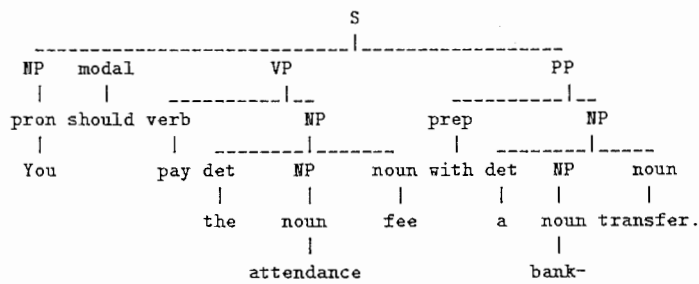


Figure 24: Dialogue 2: Sentence 14: "You should pay the attendance fee with a bank-transfer."

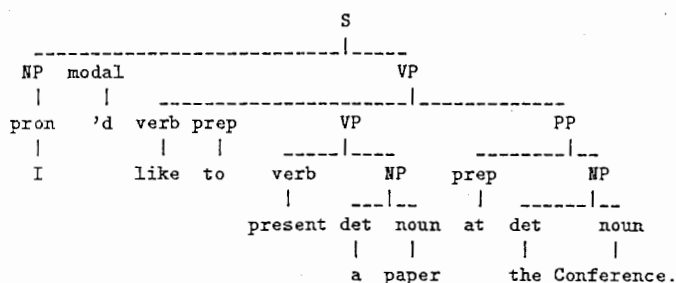


Figure 25: Dialogue 3: Sentence 3: "I'd like to present a paper at the Conference."

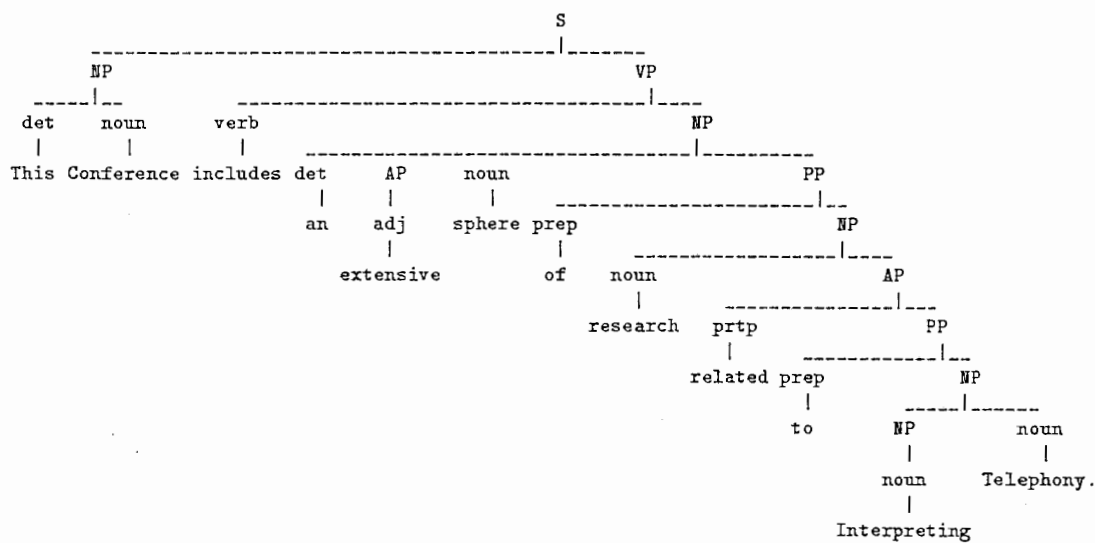


Figure 26: Dialogue 3: Sentence 5: "This Conference includes an extensive sphere of research related to Interpreting Telephony."

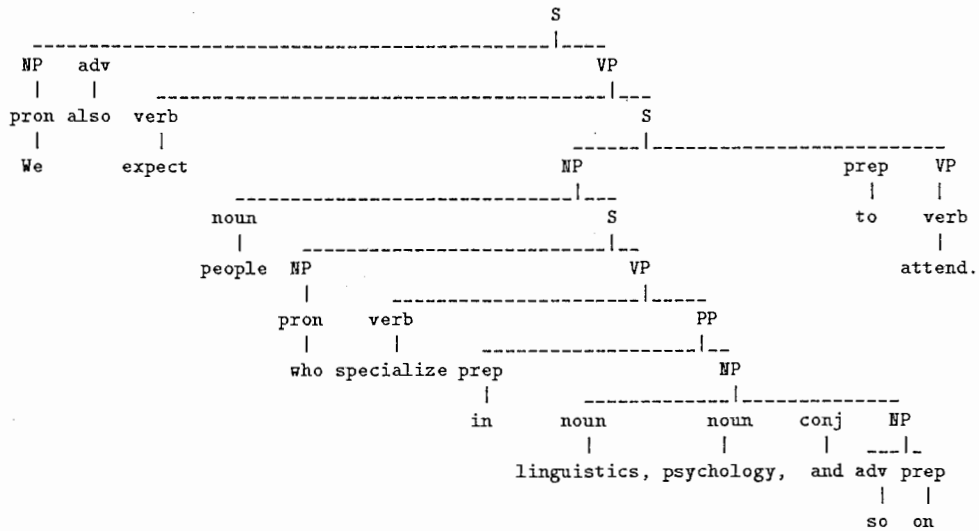


Figure 27: Dialogue 3: Sentence 6: "We also expect people who specialize in linguistics, psychology, and so on to attend."

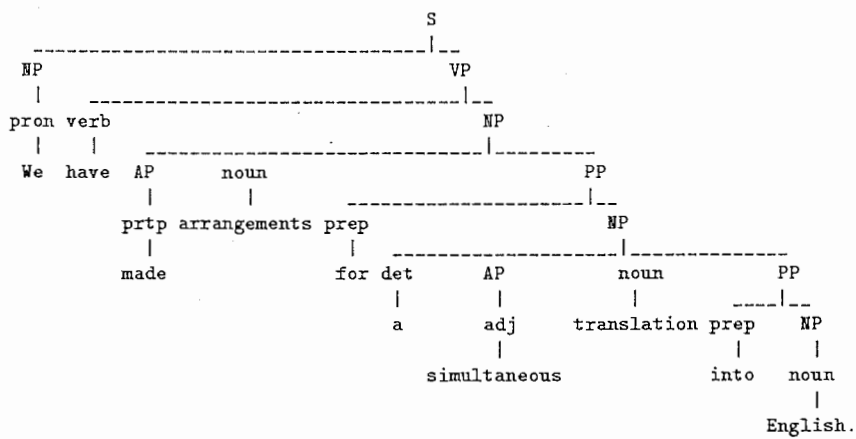


Figure 28: Dialogue 3: Sentence 13: "We have made arrangements for a simultaneous translation into English."

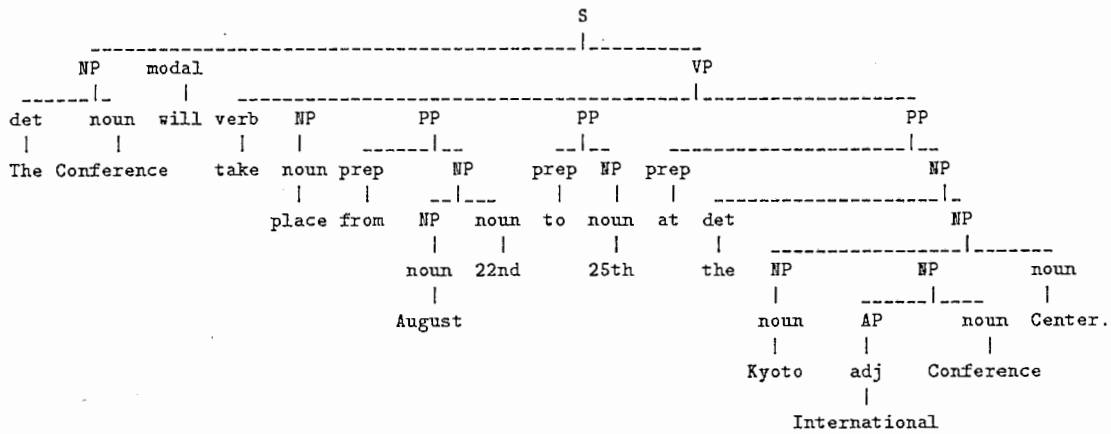


Figure 29: Dialogue 4: Sentence 7: "The Conference will take place from August 22nd to 25th at the Kyoto International Conference Center."

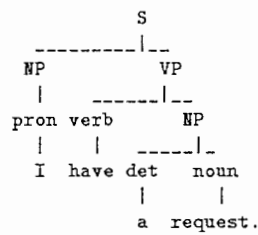


Figure 30: Dialogue 5: Sentence 3: "I have a request."

3.5 Negative declarative sentences

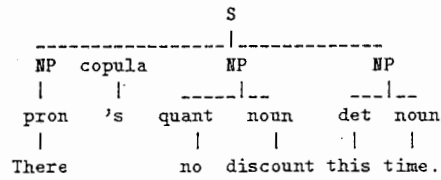


Figure 31: Dialogue 2: Sentence 11: *"There 's no discount this time."*

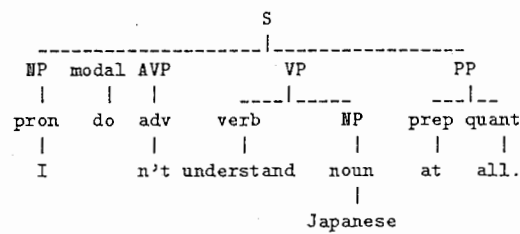


Figure 32: Dialogue 3: Sentence 10: *"I do n't understand Japanese at all."*

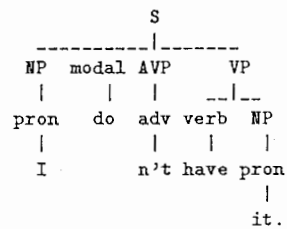


Figure 33: Dialogue 4: Sentence 5: *"I do n't have it."*

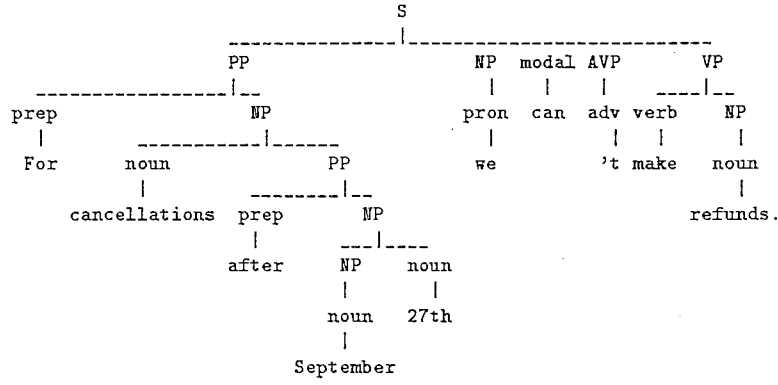


Figure 34: Dialogue 5: Sentence 15: "For cancellations after September 27th we can't make refunds."

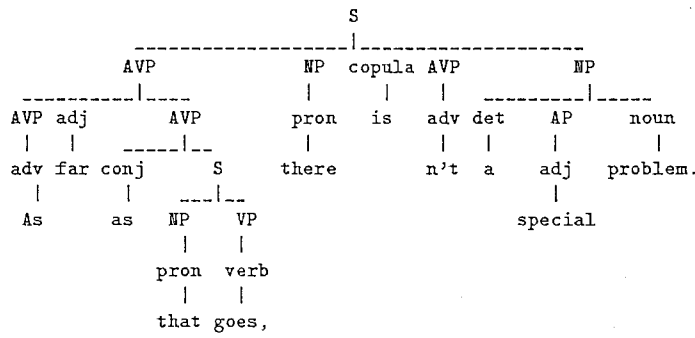


Figure 35: Dialogue 5: Sentence 18: "As far as that goes, there is n't a special problem."

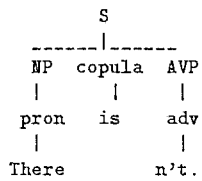


Figure 36: Dialogue 7: Sentence 19: "There is n't."

3.6 Interrogative sentences

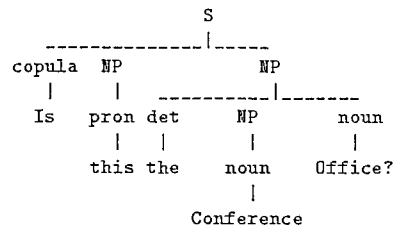


Figure 37: Dialogue 1: Sentence 2: "Is this the Conference Office?"

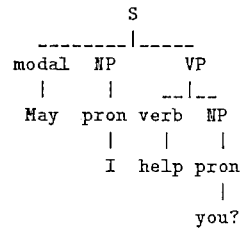


Figure 38: Dialogue 1: Sentence 5: "May I help you?"

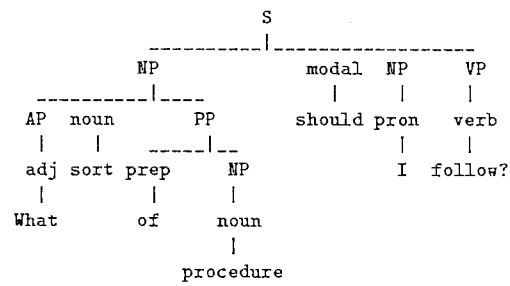


Figure 39: Dialogue 1: Sentence 7: "What sort of procedure should I follow?"

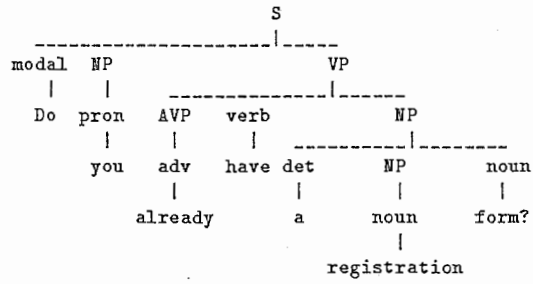


Figure 40: Dialogue 1: Sentence 9: "Do you already have a registration form?"

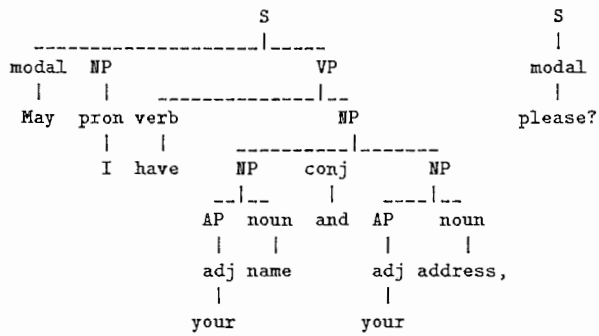


Figure 41: Dialogue 1: Sentence 14: "May I have your name and your address, please?"

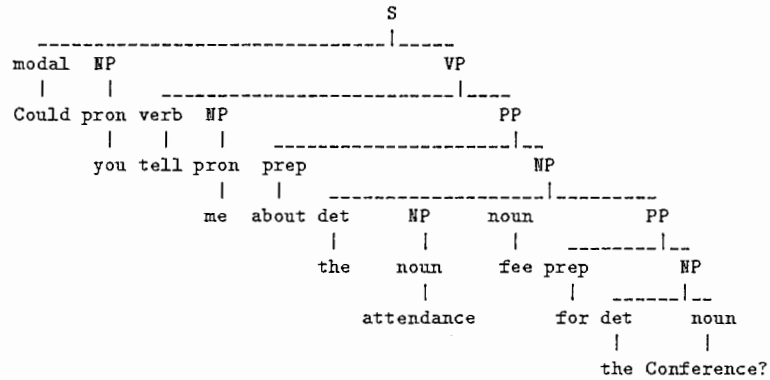


Figure 42: Dialogue 2: Sentence 3: "Could you tell me about the attendance fee for the Conference?"

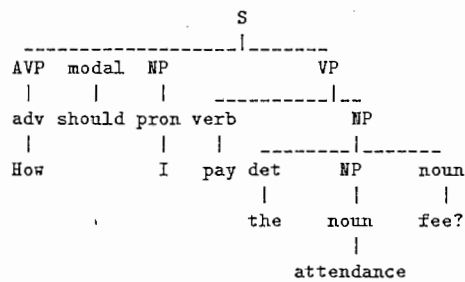


Figure 43: Dialogue 2: Sentence 13: "How should I pay the attendance fee?"

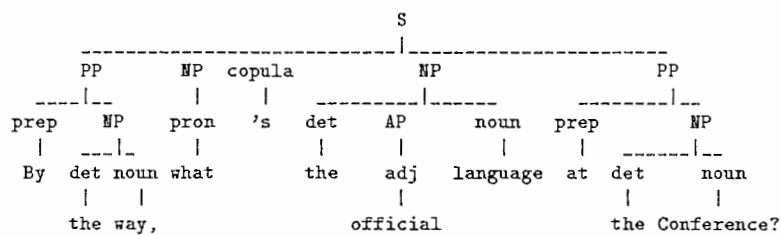


Figure 44: Dialogue 3: Sentence 8: "By the way, what 's the official language at the Conference?"

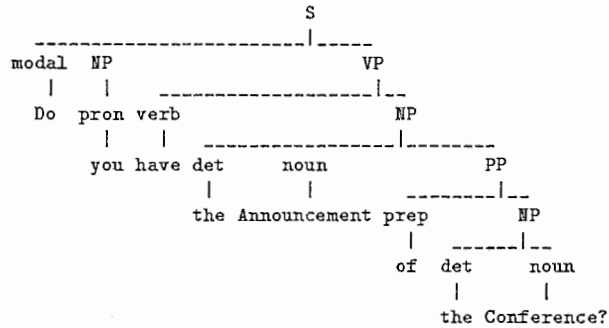


Figure 45: Dialogue 4: Sentence 3: "Do you have the Announcement of the Conference?"

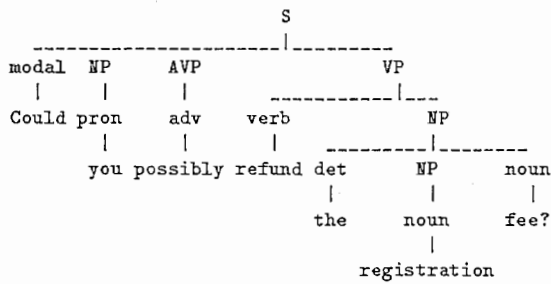


Figure 46: Dialogue 5: Sentence 12: "Could you possibly refund the registration fee?"

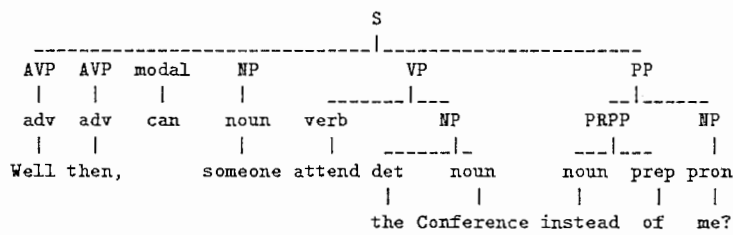


Figure 47: Dialogue 5: Sentence 17: "Well then, can someone attend the Conference instead of me?"

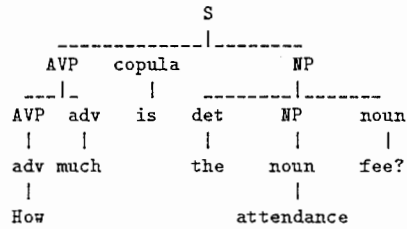


Figure 48: Dialogue 6: Sentence 9: "How much is the attendance fee?"

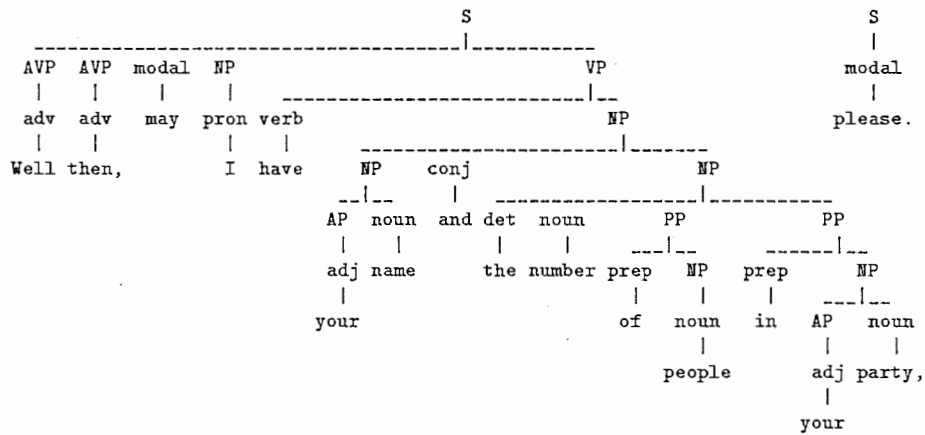


Figure 49: Dialogue 6: Sentence 16: "Well then, may I have your name and the number of people in your party, please."

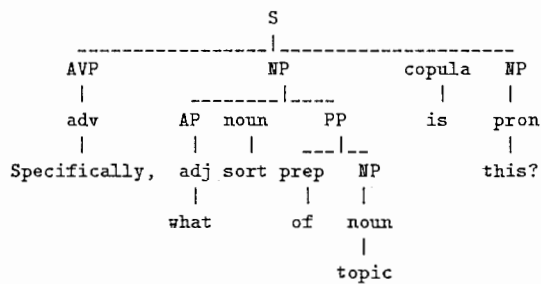


Figure 50: Dialogue 7: Sentence 7: "Specifically, what sort of topic is this?"

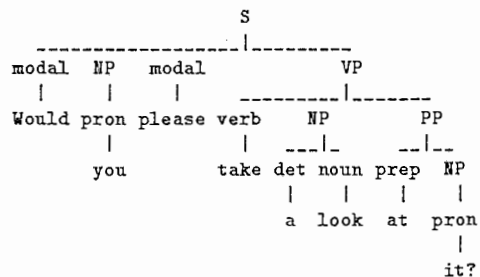


Figure 51: Dialogue 7: Sentence 10: "Would you please take a look at it?"

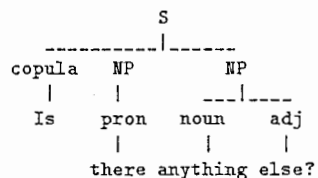


Figure 52: Dialogue 7: Sentence 17: "Is there anything else?"

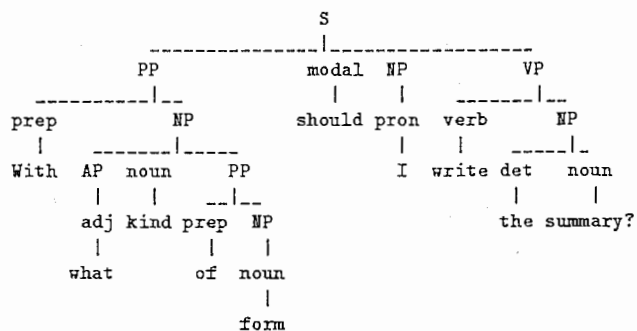


Figure 53: Dialogue 8: Sentence 11: "With what kind of form should I write the summary?"

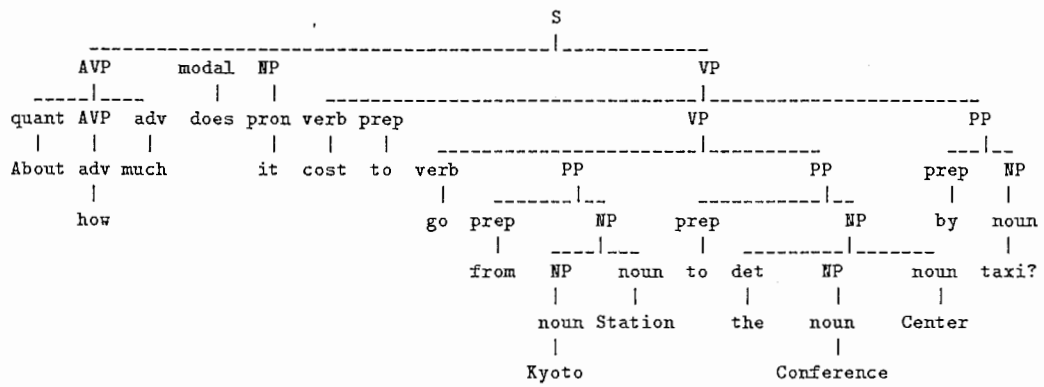


Figure 54: Dialogue 9: Sentence 10: "About how much does it cost to go from Kyoto Station to the Conference Center by taxi?"

3.7 Negative interrogative sentences

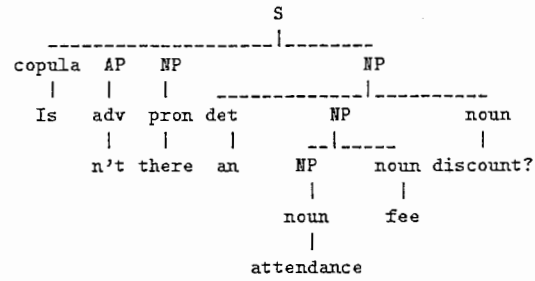


Figure 55: Dialogue 2: Sentence 10: "Is n't there an attendance fee discount?"

3.8 Incitative sentences

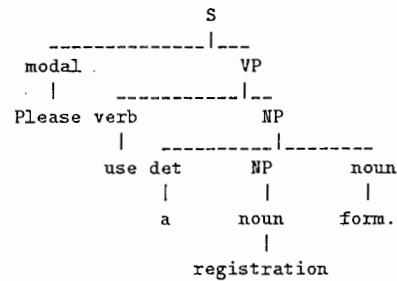


Figure 56: Dialogue 1: Sentence 8: "Please use a registration form."

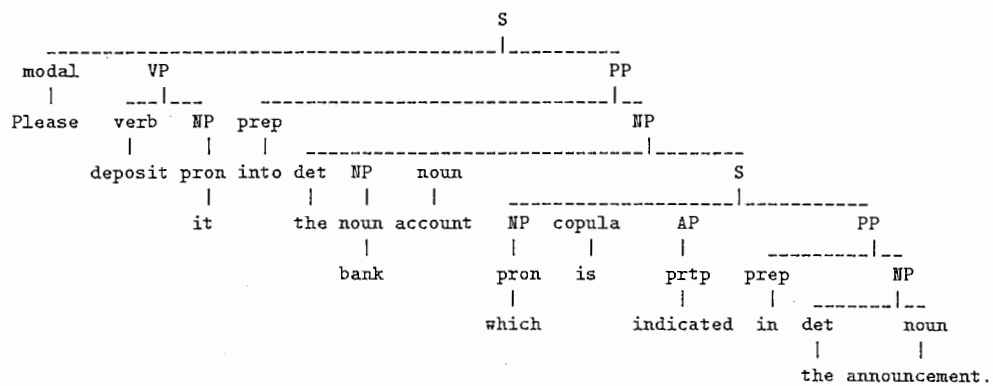


Figure 57: Dialogue 2: Sentence 15: "Please deposit it into the bank account which is indicated in the announcement."

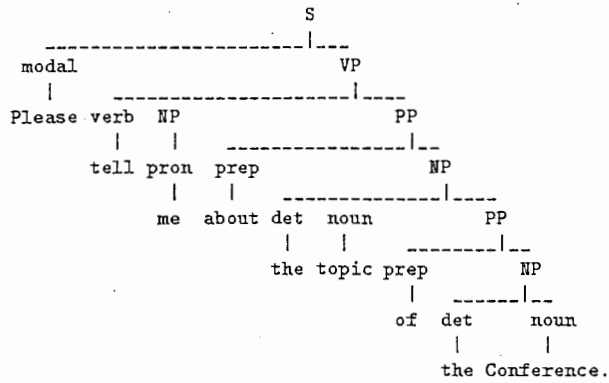


Figure 58: Dialogue 3: Sentence 4: *"Please tell me about the topic of the Conference."*

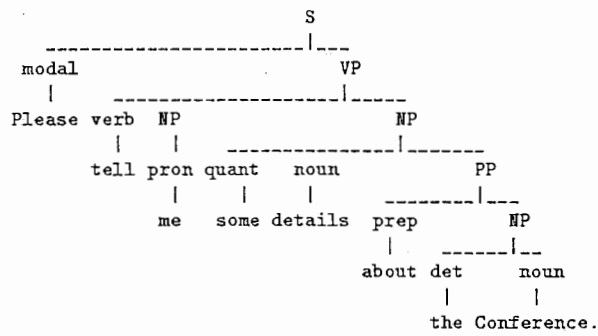


Figure 59: Dialogue 4: Sentence 2: *"Please tell me some details about the Conference."*

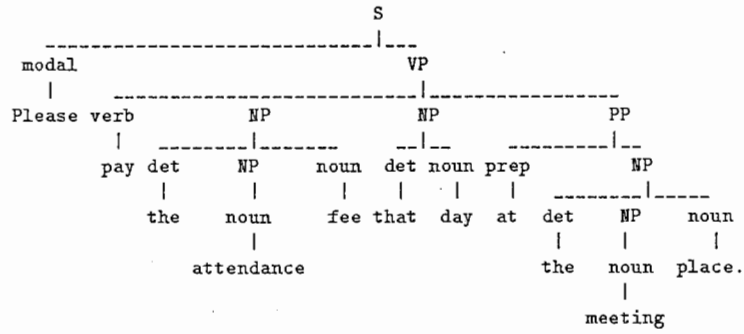


Figure 60: Dialogue 6: Sentence 20: "Please pay the attendance fee that day at the meeting place."

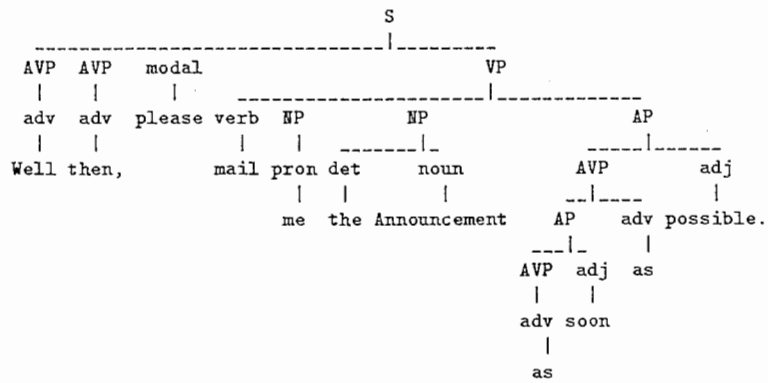


Figure 61: Dialogue 7: Sentence 12: "Well then, please mail me the Announcement as soon as possible."

3.9 Confirmatory sentences

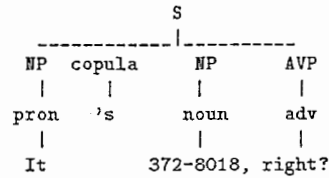


Figure 62: Dialogue 4: Sentence 18: "It 's 372-8018, right?"

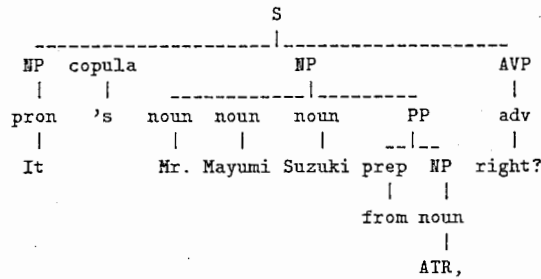


Figure 63: Dialogue 8: Sentence 17: "It 's Mr. Mayumi Suzuki from ATR, right?"

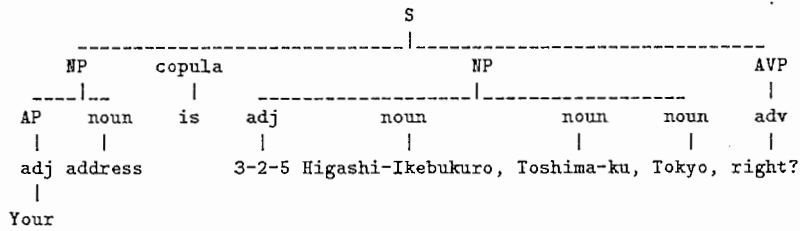


Figure 64: Dialogue 8: Sentence 18: "Your address is 3-2-5 Higashi-Ikebukuro, Toshima-ku, Tokyo, right?"

3.10 Infinitive clauses

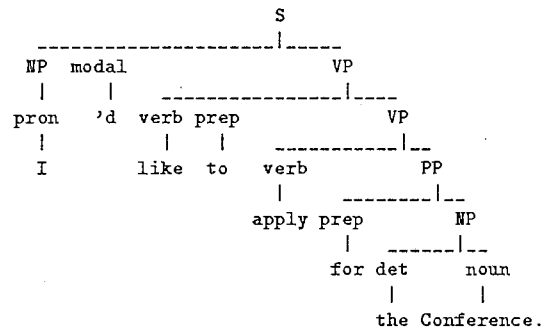


Figure 65: Dialogue 1: Sentence 6: *"I 'd like to apply for the Conference."*

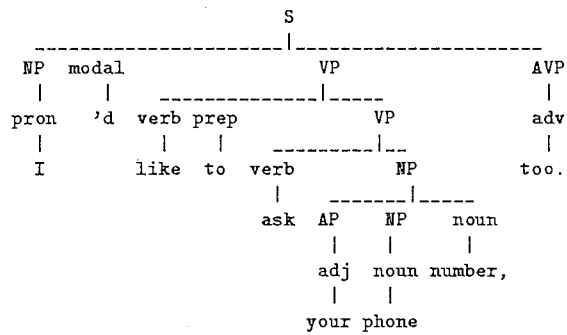


Figure 66: Dialogue 4: Sentence 15: *"I 'd like to ask your phone number, too."*

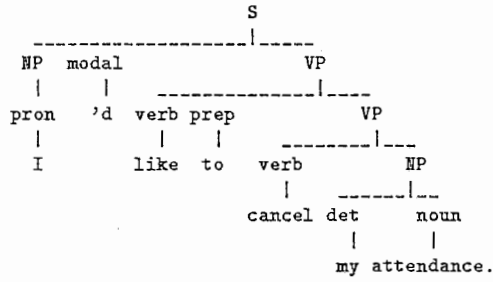


Figure 67: Dialogue 5: Sentence 5: "I 'd like to cancel my attendance."

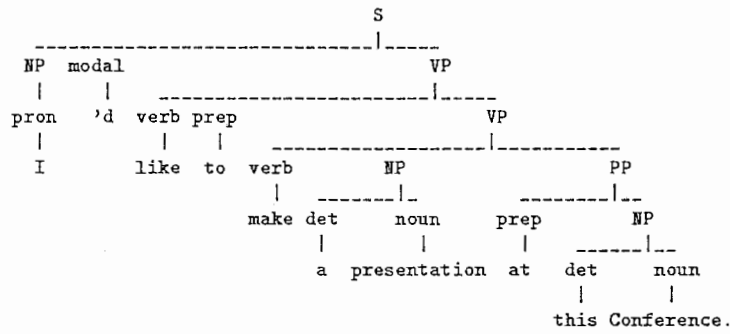


Figure 68: Dialogue 8: Sentence 4: "I 'd like to make a presentation at this Conference."

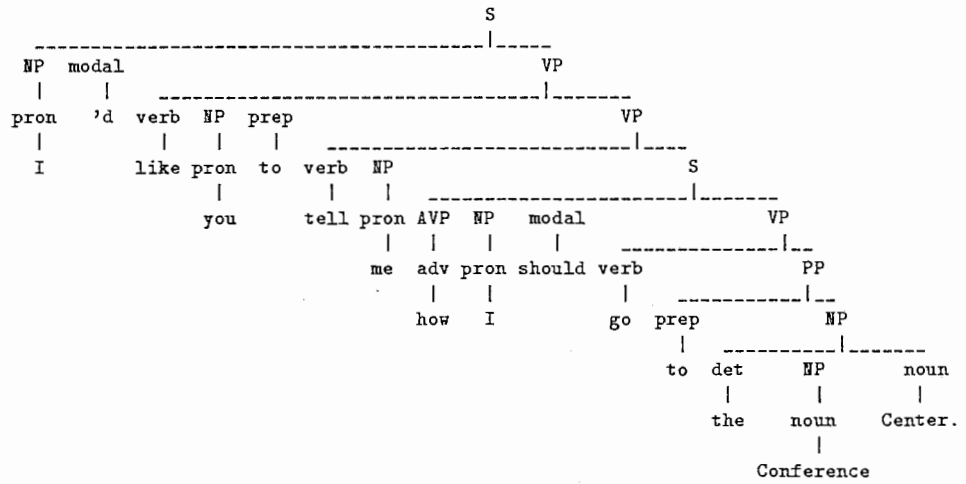


Figure 69: Dialogue 9: Sentence 5: "I 'd like you to tell me how I should go to the Conference Center."

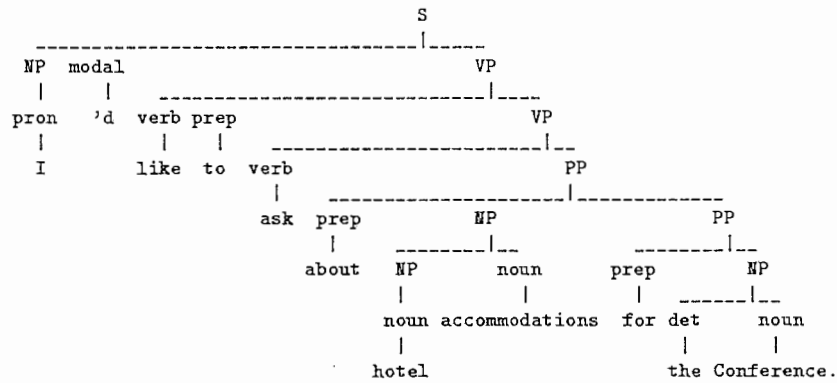


Figure 70: Dialogue 10: Sentence 4: "I 'd like to ask about hotel accommodations for the Conference."

3.11 Relative clauses

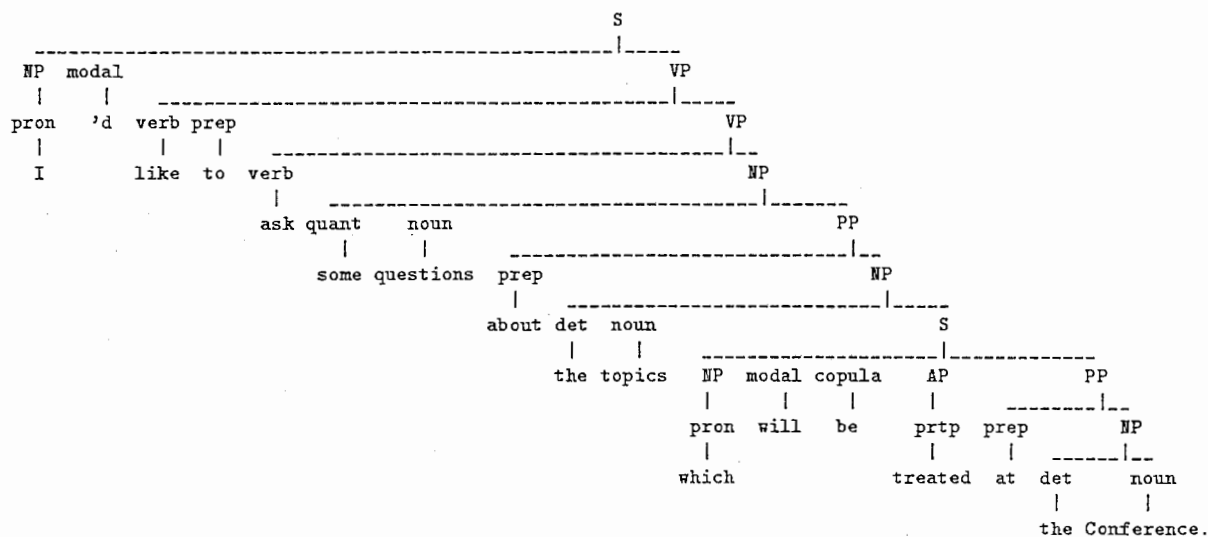


Figure 71: Dialogue 7: Sentence 3: "I'd like to ask some questions about the topics which will be treated at the Conference."

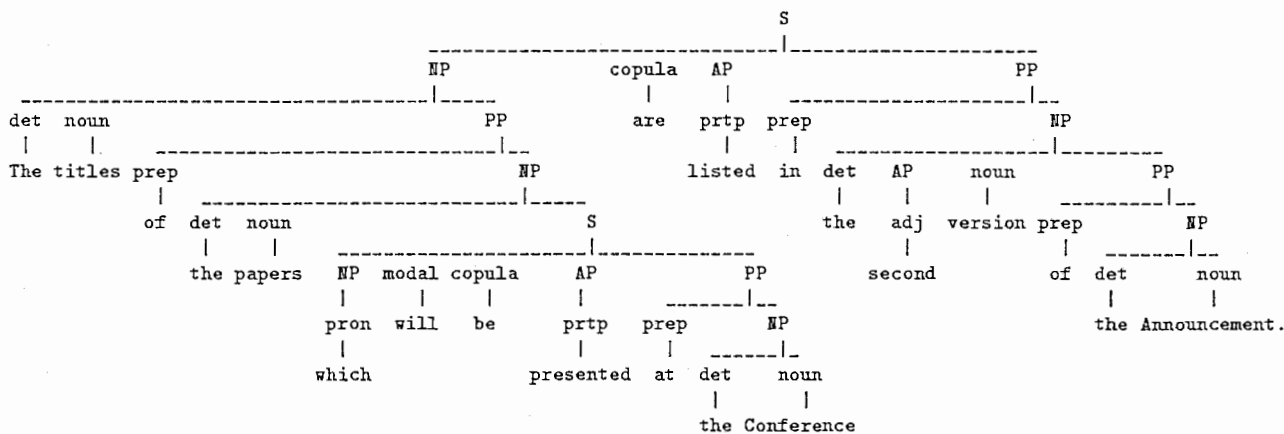


Figure 72: Dialogue 7: Sentence 9: "The titles of the papers which will be presented at the Conference are listed in the second version of the Announcement."

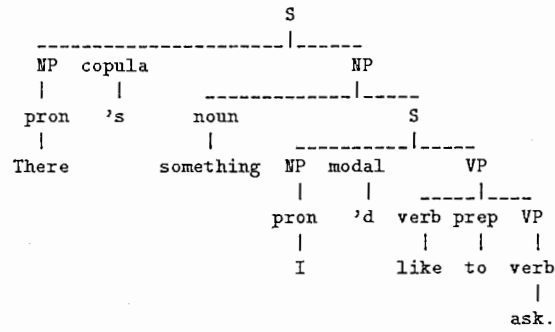


Figure 73: Dialogue 8: Sentence 3: "There 's something I 'd like to ask."

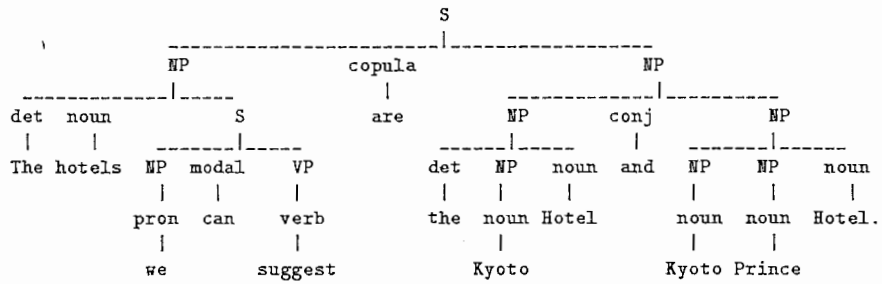


Figure 74: Dialogue 10: Sentence 7: "The hotels we can suggest are the Kyoto Hotel and Kyoto Prince Hotel."

3.12 Subordinate clauses

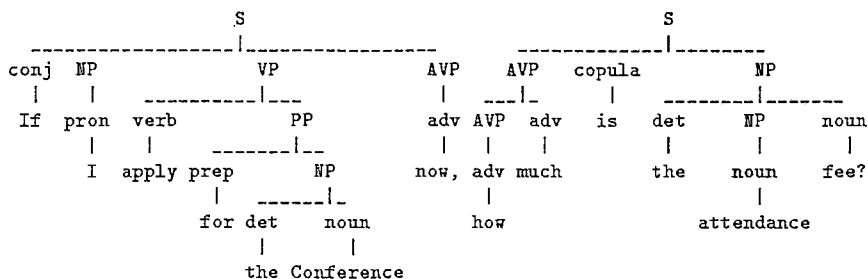


Figure 75: Dialogue 2: Sentence 4: "If I apply for the Conference now, how much is the attendance fee?"

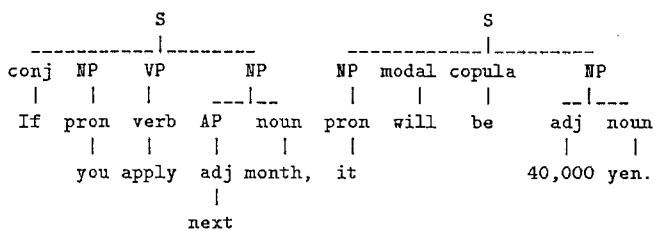


Figure 76: Dialogue 2: Sentence 7: "If you apply next month, it will be 40,000 yen."

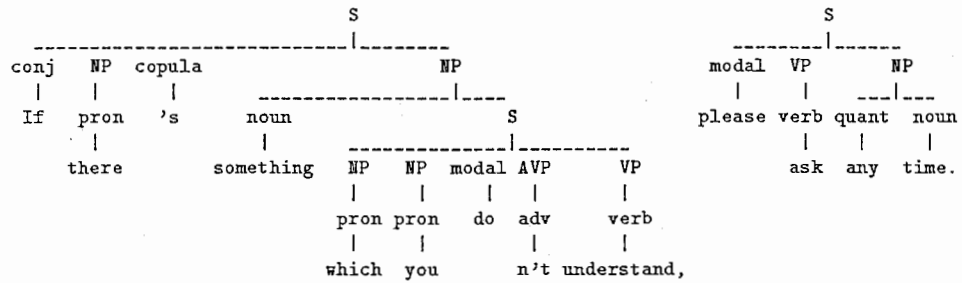


Figure 77: Dialogue 2: Sentence 20: "If there 's something which you do n't understand, please ask any time."

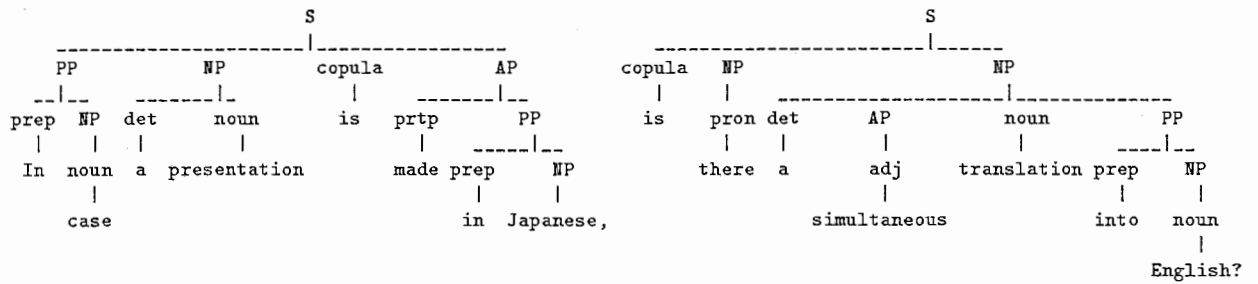


Figure 78: Dialogue 3: Sentence 11: "In case a presentation is made in Japanese, is there a simultaneous translation into English?"

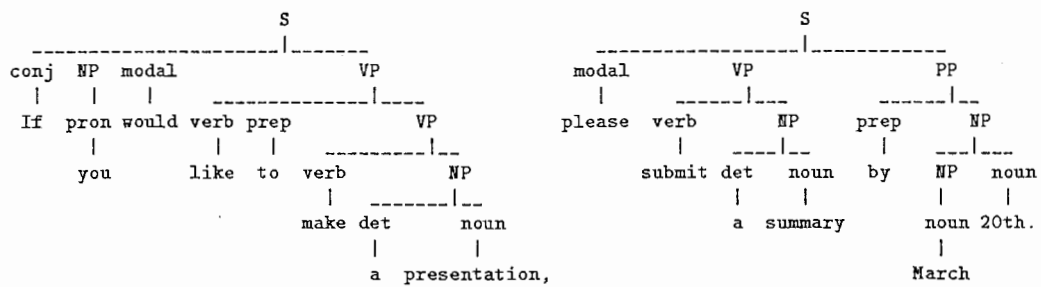


Figure 79: Dialogue 4: Sentence 9: "If you would like to make a presentation, please submit a summary by March 20th."

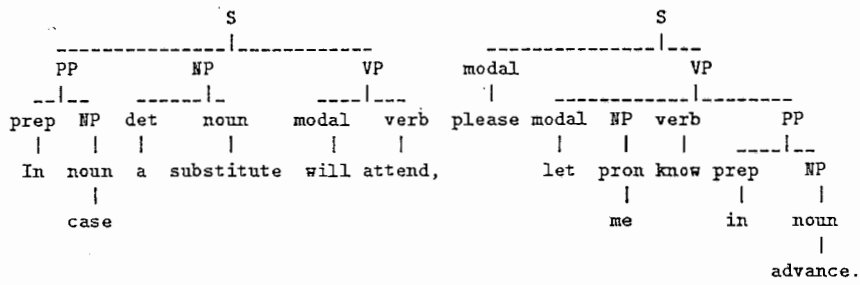


Figure 80: Dialogue 5: Sentence 19: "In case a substitute will attend, please let me know in advance."

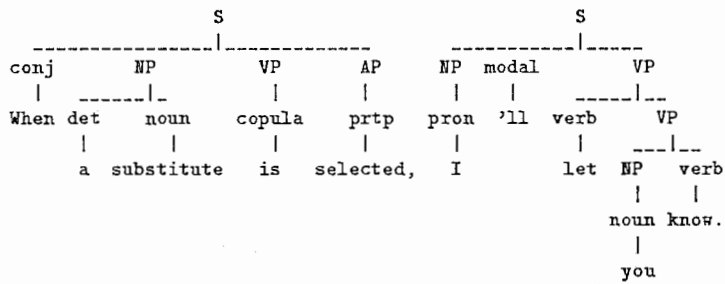


Figure 81: Dialogue 5: Sentence 21: "When a substitute is selected, I 'll let you know."

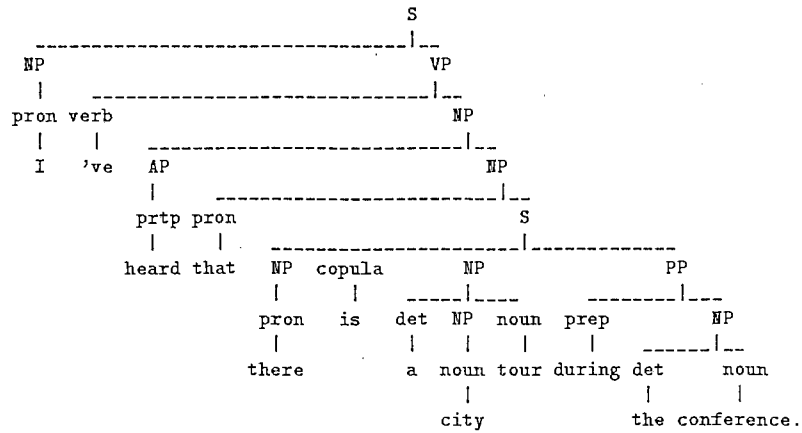


Figure 82: Dialogue 6: Sentence 3: "I've heard that there is a city tour during the conference."

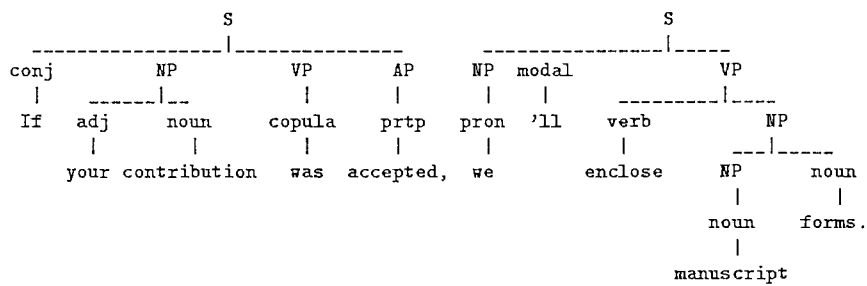


Figure 83: Dialogue 8: Sentence 8: "If your contribution was accepted, we'll enclose manuscript forms."

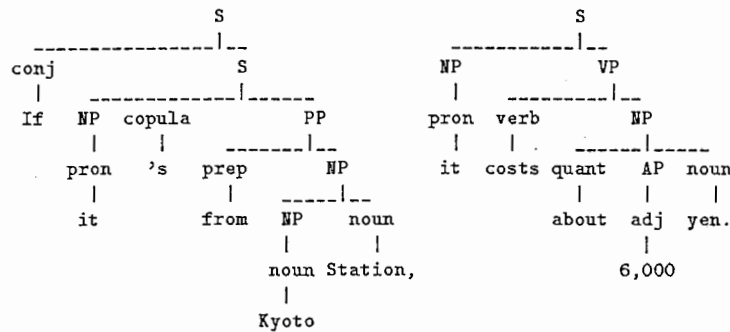


Figure 84: Dialogue 9: Sentence 11: "If it 's from Kyoto Station, it costs about 6,000 yen."

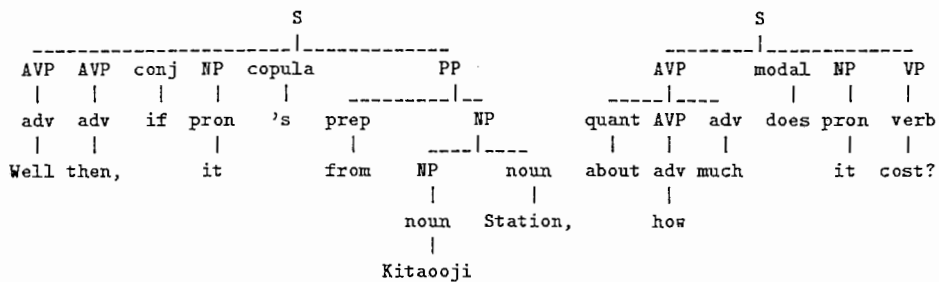


Figure 85: Dialogue 9: Sentence 12: "Well then, if it 's from Kitaooji Station, about how much does it cost?"

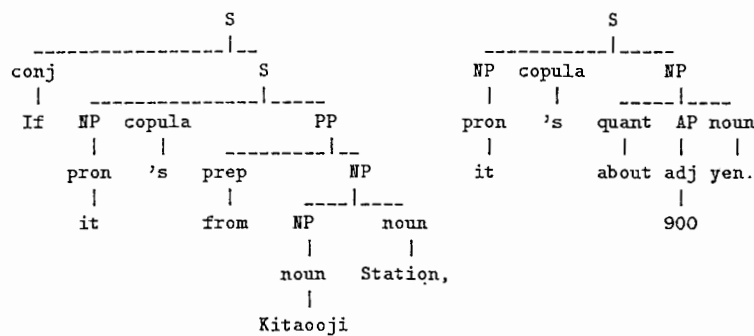


Figure 86: Dialogue 9: Sentence 13: "If it 's from Kitaooji Station, it 's about 900 yen."

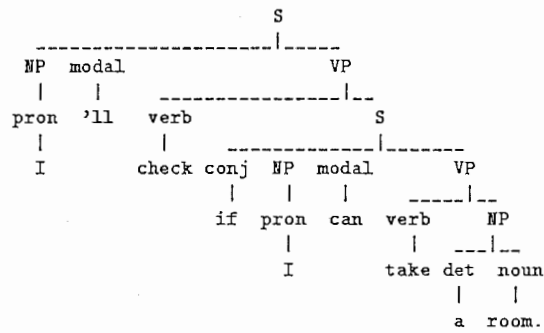


Figure 87: Dialogue 10: Sentence 28: "I 'll check if I can take a room."

3.13 Coordinate clauses

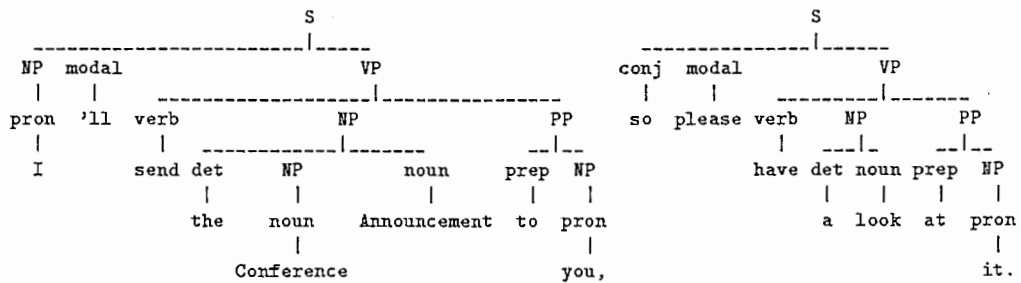


Figure 88: Dialogue 4: Sentence 10: "I 'll send the Conference Announcement to you, so please have a look at it."

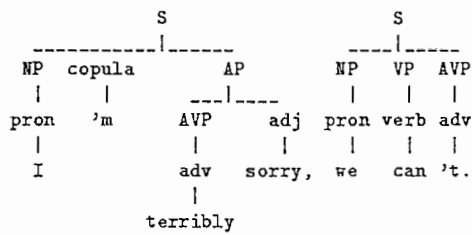


Figure 89: Dialogue 5: Sentence 13: "I 'm terribly sorry, we can 't."

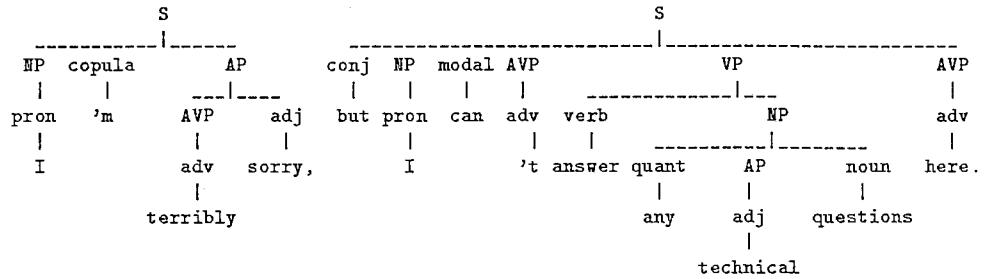


Figure 90: Dialogue 7: Sentence 8: "I 'm terribly sorry, but I can 't answer any technical questions here."

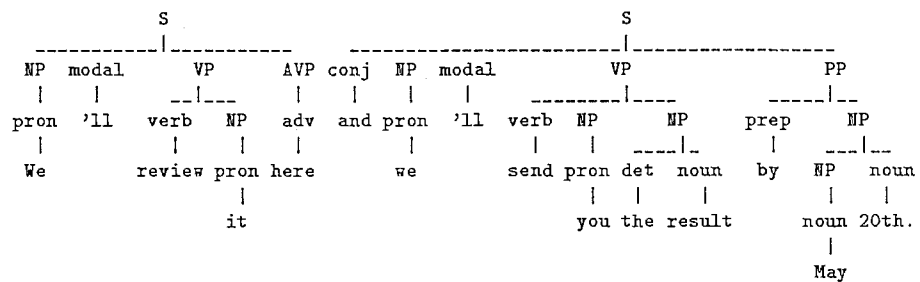


Figure 91: Dialogue 8: Sentence 7: "We 'll review it here and we 'll send you the result by May 20th."

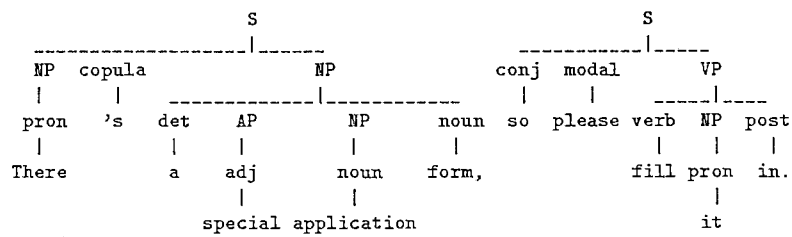


Figure 92: Dialogue 8: Sentence 12: "There 's a special application form, so please fill it in."

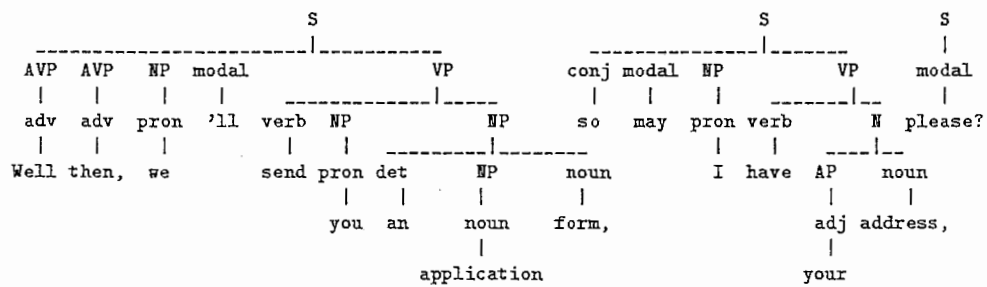


Figure 93: Dialogue 8: Sentence 13: "Well then, we'll send you an application form, so may I have your address, please?"

Conclusion

This report has presented a list of some sentences of the ten dialogues of the ATR corpus with their associated syntactic trees.

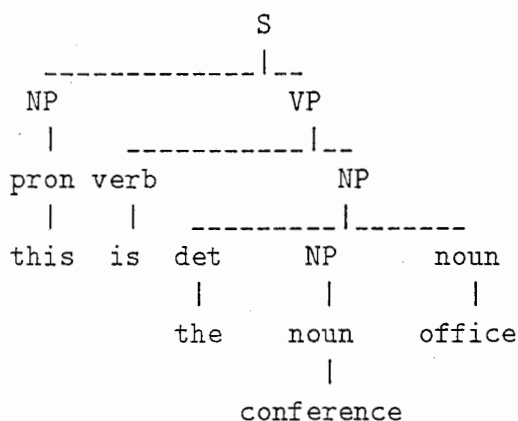
These trees are totally projective and were written with a preoccupation of consistency.

A Appendix: Against the VP node

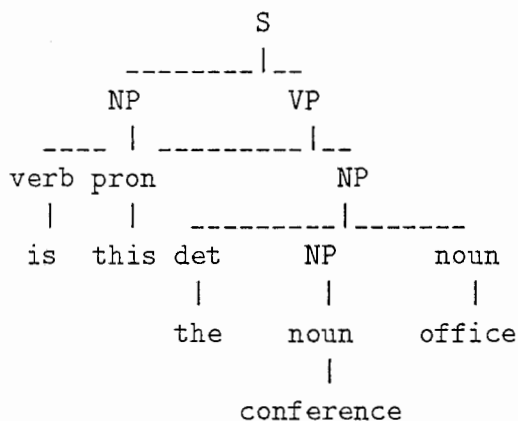
In the process of building the previous syntactic trees, it appeared clear that the VP node is more a difficulty than a convenience. Hence, for example, if the forms of *to be* had been considered those of a verb subcategorising a noun phrase, the representation for

This is the conference office.

would have been



But the interrogative form is problematic as some discontinuous representation has to be introduced.



As a matter of fact, the presence of a VP node is responsible for discontinuous representations. Starosta goes even farther [Starosta 88, p. 17]:

Positing a VP in the GB framework forces the loss of generalizations at every turn (Starosta 1977; Starosta 1985f). Anyone who cares to can verify this claim by (a) going carefully through Chomsky's 'Lectures on Government and Binding' (Chomsky 1981), including the footnotes, and making a list of the conventions, constraints, principles, and constituents, including the INFL node, which owe their existence solely to the NP-VP split and could be eliminated without it; and then (b) counting the number of generalizations for which an exception must be stated solely because the subject is not the sister of the verb. Almost all of these onerous difficulties vanish when the S-VP distinction is eliminated and the subject becomes the sister of the verb.

and page 21:

It is hard to avoid the dark suspicion that having a VP has become an end in itself, and that the awkward Spec analysis of subjects and INFL analysis of clauses is an artifact of the determination to maintain a VP category at all costs. It may not be an exaggeration to state that the main obstacle that is now standing between Government and Binding framework and a genuinely generative monostratal theory is VP.

This criticism can be addressed to Gunji's argumentation [Gunji 87] in favour of a VP node for a language, Japanese, where the order of subject, object and indirect object is tremendously well-known to be free.

Historically, the existence of VP is a consequence of the Chomskyan infamous misinterpretation of Port-Royal Grammar analysis – a sentence is composed of a subject and a predicate – and its transcription into a constituent context-free rule $S \rightarrow NPVP$. Here is the place to quote Zemb [Zemb 84, pp. 3]:

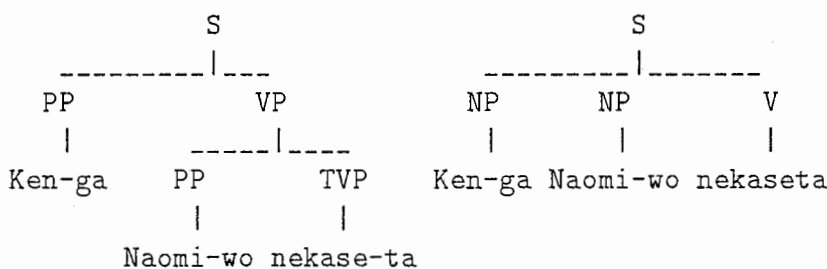
... à me disputer si souvent avec des contemporains tout en invoquant un patronage des Anciens dont ils eussent pu se réclamer plus directement que moi-même s'ils ne s'étaient pas interdit l'aveu de leur traditionalisme, p.ex. VP(NP) par rapport à f(x) ou prédicat(sujet)¹.

This misinterpretation originates in Chomsky's refusal to differentiate between constituents and functions. One of the most persuasive advocates of this distinction is to be found

¹... (I have to) argue so often with contemporary linguists by invoking the heir of the Ancients which they could claim theirs more than I can, would not they forbid themselves to confess their traditionalism, e.g. VP(NP) compared with f(x) or predicate(subject). (my translation)

in French linguistics itself, namely Tesnière. This distinction is at the very core of his theory established in the fifties [Tesnière 59]. The need for functions has been rediscovered only recently in the Government and Binding framework under the inflated name of θ -theory. But sadly, the failure of recognising the constituent/function distinction created an artificial separation in the linguistic spheres between constituency and dependency representation proponents. Arguments in favour of the latter representation may be found in [Hajičová 88]. Again, Starosta elegantly bridged the gap by removing precisely ...VP. He shows that structures respecting his X-bar-only representation can be viewed as constituent *and* dependency structures at the same time [Starosta 88, pp. 104 and 110]. By the way, these structures are very similar to Vauquois' intermediate structures [Vauquois 88].

For its *Japanese Phrase Structure Grammar*, Gunji insisted on having a VP node in the structures drawn, *i.e.* to prefer the following left structure over the right one.



(3.3)

(3.4)

He honestly states:

I am currently not aware of any strong argument for preferring (3.4) over (3.3). Even though it may also be difficult to give a strong argument for (3.3) over (3.4), there seem to be several pieces of evidence that suggest the constituency of the object and the transitive verb in Japanese.

His argumentation is based on two facts.

- the anaphorical value of the form *soo-su*. *Ken-ga neta node Marie-mo soo-sita.*
- the commutation of an intransitive verb with its factitive correspondent and its associated object: *Ken-ga Naomi-wo nekaseta / Ken-ga neta.*

Against the first argument, one can say the following. *soo-su*² is not a basic unit, as it is by no means felt as a verb by native speakers of Japanese. The verb *su* exists and the similar forms *soo-iw*, *soo-omow*, etc. exist also. Consequently, the only deduction from the examples given in [Gunji 87] is that *soo* could get a similar representation to that of an object noun-phrase.

Against the second argument, we must examine the underlying assumption. It is that commutating pieces of text have to be represented by similar *subtrees*. Unfortunately, this assumption gives rise to many problems of representation, not the least of which being representation of coordination. For instance, following this never explicitly settled assumption, [Gazdar 81] admits that a simple rule for coordination such as

$$a \rightarrow a_1 a_2 \dots \text{and } a_n$$

cannot apply directly to the sentence

John gave the book to Mary and the records to Sue.

because the direct and indirect objects do not form a constituent. In terms of data structures, the representation of the direct and indirect objects is not one tree, but two trees making a forest.

This problem leads to a fundamental question: the well-foundedness of linguistic operations dealing only with subtrees. We have argued [Lepage 91] elsewhere, that an enormous amount of perspicuity is gained by allowing operations on forests, and not only on subtrees. This question loops back to the problem of discontinuous structures, a number of which have been exhibited as arguments against the existence of a VP node. For example, in German the three following sentences ([Uszkoreit 86]).

Geben sollte er seiner Frau das Buch.

Das Buch geben sollte er seiner Frau.

Seiner Frau geben sollte er das Buch.

Other such phenomena have been shown in Finnish and treated without any VP in a lexicalist approach in [Karttunen 86].

In conclusion, we have argued here against the existence of the VP node, which is one of the touchstones of one of the dominant (if not imperialistic) streams in linguistics. We have shown that the advantages of its absence have been advocated by renowned linguists. We hope that this discussion will help people not to consider as untouchable axioms which appear to hinder simple descriptions.

²We keep Gunji's notation.

References

- [Fais and Kikui 91] Laurel Fais and Kikui Gen-Ichiro
Determining Surface Form for Indirect Speech Acts in English
ATR report TR-I-0235, Kyoto, December 1991.
- [Fais 93] Laurel Fais
An English Analysis Grammar in a Unification-based Framework
ATR report TR-I-0351, Kyoto, March 1993.
- [Gazdar 81] Gerald Gazdar
Unbounded Dependencies and Coordinate Structure
Linguistic Inquiry, Volume 12, Number 2, Spring 1981, pp. 155-184.
- [Gunji 87] Gunji Takao
Japanese Phrase Structure Grammar
D. Reidel Publishing Company, 1987.
- [Hajičová 88] Eva Hajičová
Reasons Why We Use Dependency Grammar
Proceedings of COLING-88, p 451, Budapest, 1988.
- [Karttunen 86] Lauri Karttunen
Radical Lexicalism
Report No. CSLI-86-68, Stanford, December 1986.
- [Lepage 91] Yves Lepage
Parsing and Generating Context-Sensitive Languages with Correspondence Identification Grammars
Proceedings of the Pacific Rim Symposium on Natural Language Processing, Singapore, November 1991, pp. 256-263.
- [Lepage 92a] Yves Lepage
Easier C programming
Input/output facilities
ATR report TR-I-0293, Kyoto, November 1992.
- [Lepage 92b] Yves Lepage
Easier C programming

Some useful objects

ATR report TR-I-0294, Kyoto, November 1992.

- [Lepage 93a] Yves Lepage
Analysis, generation and more by means of genetic algorithms
ATR report TR-IT-0002, Kyoto, June 1993.
- [Lepage 93b] Yves Lepage
Analysis, generation and more in natural language processing by means of genetic algorithms
Proceedings of the First Meeting of Special Interest Group on Parallel Processing for Artificial Intelligence, pp 13-18, Tokyo, September 1993.
- [Lepage 93c] Yves Lepage
Analysis, generation and more using genetic algorithms
Technical Report of IEICE, NLC 93-42-47, October 1993, pp. 1-8.
- [Tesnière 59] Lucien Tesnière
Éléments de syntaxe structurale
Klincksieck, Paris, 1959.
- [Starosta 88] Stanley Starosta
The Case for Lexicase
Pinter Publishers, London and New York, 1988.
- [Uszkoreit 86] Hans Uszkoreit
Categorial Unification Grammars
Report No. CSLI-86-66, Stanford, December 1986.
- [Vauquois 88] Bernard Vauquois
Bernard Vauquois et la TAO, vingt-cinq ans de Traduction Automatique
Bernard Vauquois and MT, twenty-five years of MT C.Boitet ed., association Champollion & GETA, Grenoble, 1988.
- [Zaharin and Lepage 92] Zaharin Yusoff and Yves Lepage
On the specification of abstract linguistic structures in formalisms for Machine Translation
Proceedings of the International Symposium on Natural Language Understanding and AI, pp 145-153, Iizuka, July 1992.

[Zemb 78] Jean Marie Zemb
Vergleichende Grammatik Französisch-Deutsch
Comparaison de deux systèmes - Teil 1
Dudenverlag, 1978.

[Zemb 84] Jean Marie Zemb
Vergleichende Grammatik Französisch-Deutsch
L'économie de la langue et le jeu de la parole - Teil 2
Dudenverlag, 1984.

