TR-IT-0176

Tesnière's structural syntax: notations for tree-banking using BoardEdit

Yves Lepage

July 1996

Abstract

This report is intended as specification for tree-banking, using linguistic representations derived from those of Tesnière in his master work: *Eléments de syntaxe structurale*, and using the text-and-structure editor BoardEdit.

Keywords

Structural syntax, dependency structures, tree-bank.

Contents

3

List of Figures

Introduction

The object of this report is to give a brief description of Tesnière's main ideas concerning syntactic representations. It is somehow a summary of the *Eléments de syntaxe structurale* (Outline of Structural Syntax), but as we just give the conclusions Tesnière arrives at, we ask the reader to refer to the book to get theoretical and experimental justifications.

This brief description should serve as a sufficient basis for a linguist to perform tree-banking, *i.e.* the drawing of linguistic representations in the form of trees, using BoardEdit, the tree-text-correspondence editing tool[1].

Examples are given, which illustrate each of the three basic phenomena according to structural syntax: *connection, junction* and *transference*. Also, linguistic choices to be made for the analysis of Japanese will be mentioned. Methods used in structural syntax as linguistic experiments to make these kinds of decisions will be briefly presented.

---

[1]designed and implemented by Yves Lepage.

1 Introduction to structural syntax

After having given the different types and categories of words, we will introduce the three phenomena to which Tesnière reduces all syntactical phenomena. They constitute the pillars upon which structural syntax is built. For each of these phenomena, we will follow Tesnière's notations whenever possible. In case it is not possible, we shall propose an alternate way of encoding it into forest structures.

1.1 The types and categories of words

Sentences do not exist without words and words do not exist independently of sentences. Our first task will be to categorise words.

According to the classical structuralist view, there are basically two different types of words:

- *content words* (Fr. mots pleins), which bear a proper meaning, and

- *function words* (Fr. mots vides), which bear syntactical information.

1.1.1 Content words

They bear a semantic function, *i.e.* they are directly associated with a concept, and their function is precisely to evoke or represent this concept. Examples are 馬, 飲む, 黒い, あらかじめ.

Categories There are two classes of content words:

- *processes* and

- *substances.*

Because in indo-european languages processes or predicates are often expressed by verbs, and substances are often expressed by nouns, one tends to explain these classes with the help of these corresponding syntactical categories.

These two classes can be in turn divided into two sub-classes:

- *concrete* and

- *abstract.*

This opposes the concrete notion of processes and substances to their abstract attributes. This opposition is clearly viewed in the following examples

黒い馬が早く走っている。

where each concrete content word (馬, 走る) is complemented by an abstract attribute (黒い, 早く).

Hence the classification for content words (see also [Starosta 88]) given in the following array, where Tesnière's notations appear as big letters.

	substances	processes
concrete	substantive O	verbial I
abstract	adjectival A	adverbial E

This notation with capital letters will be used during tree-banking to represent the content words appearing in sentences.

1.1.2   Function words

They are grammatical tools, the role of which is to make explicit, to precise or to transform the category of content words, and to define the relationships between these words.

Typical examples of function words in Japanese are: が, を, に, います, おります, です, します.

Empty words will appear as they are in structural representations.

1.2   Tree representation: stemmas

Historically, as back as in 1934, Tesnière was the first linguist who ever proposed to systematically use trees[2], which he called *stemmas*, to

---

[2] Exactly speaking, Tesnière's representations are more than simple trees, they are graphs.

represent syntactical phenomena [Tesnière 34]. He acknowledged that some Russian linguists, in 1930, did use trees to explain some syntactic phenomenon, but, on the contrary to Tesnière, they never made this representation the central support of their explanations.

Tesnière designed his structural syntax to allow for phenomena from various languages. The representations adopted are dependency structures[3], which basically describe relations between words in a sentence in terms of their subordination relations, or *connection* in Tesnière's terminology. This covers predicate-argument, governor-modifier, *etc* relations.

For a given sentence, for example

黒い馬が早く走っている。

one can draw the *real stemma* on the left in Figure 1.

走っている   I
  /\   /\
が   早く   が E
 | |
馬   O
 | |
黒い   A
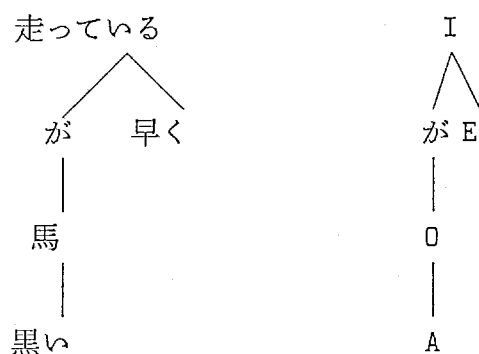
Figure 1: A real and a virtual stemma.

One gets a *virtual (or symbolic) stemma* by substituting content words for their category symbols. Function words are left unchanged. This is shown on the right in Figure 1. The work of the tree-banker will be to draw similar *virtual stemmas* for each sentence of a data-base.

---

[3]Refer to Zemb, Starosta, Gross and others for comments on the fact that constituency structures are English-oriented representations, in which some linguists try desperately to cast by force phenomena from other languages.

Correspondences In order to make explicit which word, or better said, especially in the case of Japanese, which chunk of the text corresponds to which node in the stemma, the tree-banker will use correspondences [Boitet and Zaharin 88].

There basically exist two kinds of correspondences [Lepage 94]:

- words-to-node, and

- sentence parts-to-complete subtrees, or substring-to-subtree.

They are governed by three constraints which are sufficient to capture standard linguistic structural representations (as well dependency as constituency representations).

- **global correspondence**: the entire tree corresponds to the entire sentence;

- **inclusion**: if a subtree is included in another subtree, then the substring in correspondence with the included subtree is included in the substring in correspondence with the including subtree;

- **membership**: if a node is member of a subtree, then the words in correspondence with the node are in the substring in correspondence with the subtree.

In the following example, on each node of the stemma, two intervals stand for the words-node and the substring-subtree correspondences in that order. The BoardEdit tool offers a facility to create correspondences.

```
                    I[4_5/0_5]
                         /\
           が [2_3/0_3]      E[3_4/3_4]
                 |
           O[1_2/0_2]
                 |
           A[0_1/0_1]
```
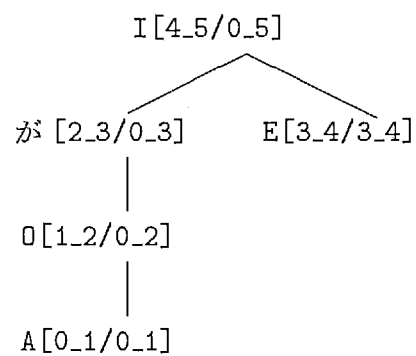
₀ 黒い ₁ 馬 ₂ が ₃ 早く ₄ 走っている ₅

Figure 2: A virtual stemma with correspondences showed with the help of intervals.

1.3 The three syntactical relations

1.3.1 Connection

In a sentence, words are usually in relations of dependency relatively one to another. *The study of sentences, which is the proper object of structural syntax is essentially the study of its structure,* i.e. *the hierarchy of its connections.*

Basically, and in the simpler cases, connections happen between concrete notions and their attributes (Figure 3). They are represented by vertical links (mother-daughter relations in a tree).

```
O                    I
|                    |
A                    E
```

黒い馬   早く走っている
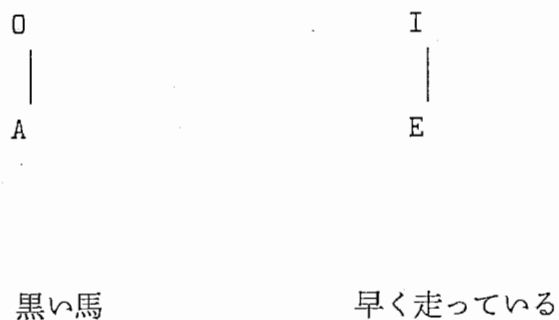
Figure 3: Basic connections.

Figure 1 already showed connections inside a simple sentence.

1.3.2 Junction

Classical grammar puts under the same denomination of *conjunction* two kinds of empty words which allow the expansion of the sentence: *junctive* and *translative* words.

Junction (Fr. jonction) gathers the facts of coordination, and fatorisation.

Some junctive words in Japanese are と, や, し, けど.

The representation we propose is to have special nodes for junctive words:

- the function word is kept in the node as is;

- it is preceded and followed by --, which tries to keep Tesnière's representations by horizontal links.

With this, we easily represent cap junctions as shown in Figure 4.

Cup cases are cases where the same dependent share several governors, *e.g* a same unique subject for several coordinated verbs. These cases need a special representation. We shall represent a complete factored subtree by repeating it in as a node bearing a special label, as many times as required. Correspondences will be responsible for representing the factorisation: they will point to the same portion of text. As for a rule, and so as to reflect the order of appearance in the string, we decide that:

- the factored subtree should always come to the left;

- factoring nodes, with label V, come after, to right. On these nodes, the correspondences are the same as the correspondences borne by the mother node of the factored subtree.

An example of this notation is given in Figure 5.

Because of junction, a structure representing a sentence may be a forest, and not only a tree[4]. Figure 5 already showed such a structure. This is not rare case, and another example is found in the first sentences of the TDMT corpus (see Figure 6).

---

[4]This is the second big difference with constituency representations, which try to enforce sentences into the corset of a tree.

```
                      か
                      |
                      I
                   /     \
                  は       OT
                  |       / \
                  O      I    + い
                  |      |
                  AT     は
                 / \    / | \
                O  の O -- と -- O
```
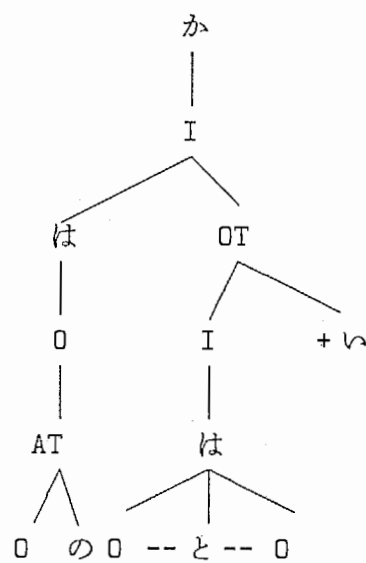
その料金は税金とサービス料は込みですか。

Figure 4: A cap case of junction.

```
A[2_3/0_3]      -- で[3_4/3_4]--      A[4_5/0_2+4_5]
       |                                     |
    は[1_2/0_2]                           V[1_2/0_2].
       |
    O[0_1/0_1]
```

$_0$彼女$_1$は$_2$親切$_3$で$_4$美しい$_5$。

Figure 5: A cup case of junction.

```
E --,--              I
                    / \
                  ET   O
                 / \
                O   で
                |
                AT
               / \
              O   の
```

いいえ、チェックアウトの時で結構です。

Figure 6: Forest case of stemma.

17

1.3.3   Transference

Constituency grammars have failed in perceiving the major discovery made by Tesnière: the difference between the syntactical category of a word and its role in a sentence.

Transference[5], by essence, consists in transferring a content word of a given category into another category. This phenomenon is somehow the dual or the counterpart of the existence of categories. It is precisely transference which allows a speaker of a given language to never be stopped by the fact that a needed concept does not fit, by category, at a given point in the middle of an utterance, into the role required at that point[6].

Transference applies to a content word, the transferee. It is performed by a transferer, which may be:

- a function word の, に, する, *etc.*

- some morphological device + く, + て, *etc.* We note morphological devices with a + on the left (*resp.* on the right) to indicate that it is a suffix (*resp.* a prefix);

- no mark at all (the so-called, and badly called, relatives in Japanese are in fact transferences: a verb is transferred into an adjective, without any marker). In this case, we note the label of the transferer node with a ∅.

The result is that the category of the content word has been transformed into another category, so that it can play the role of the resulting category.

馬:O → 馬の:A

---

[5]Here, we follow the recommendation of Tesnière himself to render the French *translation* by an English term specially coined on this occasion.

[6]The big failure of generative linguistics is here: constituency structures try to capture categories *and* functions at the same time, hence the abherrant privilege of the subject NP in those representations, which do not have any valid explanation in slavic languages or Japanese (see [Gunji 87] for an unsuccessful attempt of justification). The genius of Tesnière is to have recognised that categories do well correspond to some definite function, but also to reveal that any language possesses means to transform any category into any other category, so that any word can potentially fit into the function of any different category.

The two directions in transference Depending on the position of the transferer to the left or to the right of the content word, left and right transferences have to be distinguished. In Japanese, massively, the transferer is placed to the right of the transferee.

We choose to represent transference with the help of a 3-node subtree:

- the mother bears the category into which transference transfers the full word; the order in which the tranferer and the transferee appear is indicated by the big letter T placed:

  - on the right of the resulting category if the transferer is on the right of the transferee in the sentence. This is the usual case in Japanese;

  - on the left of the resulting category in the other case.

- the left (or right) daughter bears the category of the transferee, depending on its actual place in the sentence;

- the right (or left) daughter bears the transferer, *i.e.* the function word *in extenso*.

```
        AT
       / \
      0   の
```
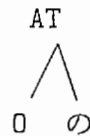
馬の

Figure 7: Representation of transference.

2　Choices

The task of the tree-banker, who should be a linguist, is to make choices in certain cases. We have enumerated some of these necessary choices in the following list.

These choices should be made consequently, after linguistic inquiry, based on linguistic tests. In the next section, we describe two of the most common ways of testing hypotheses in structural linguistics.

As soon as they are made, these choices should be consequently followed in all similar cases, so as to ensure the consistency of the data.

2.1　Some choices made

Morphological transference　Morphological transference will be represented by noting the transferer as it appears in the text, preceded by a + if it is a suffix, as is often the case in Japanese. The + is placed after if it is a prefix.

Hence, the adverbial form of adjectives will be represented as in Figure 8. Suspended forms of verbs are obtained in Japanese either by derivation with the suffix い or by retrogradation (noted by -) for verbs of the type 食べる. This cases of morphological transferences are shown in Figure 8.

```
     ET              OT              OT
    / \             / \             / \
  I  + く         I  + い         I  - る


   早く             付き             食べ
```

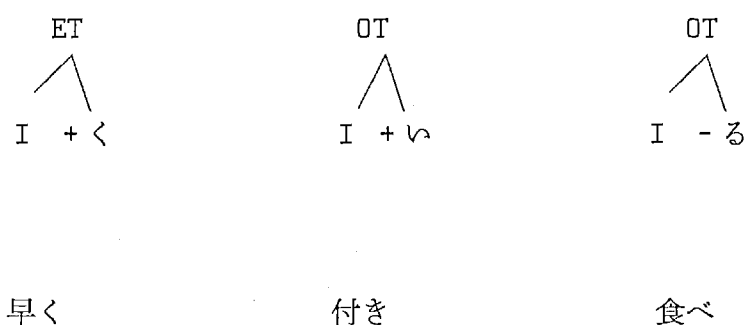Figure 8: Morphological transferences.

Deictics   We have chosen to represent the series of Japanese deictics この, その, あの and どの, as transferences from O to A because the regularity of the series induces this analysis. See examples in Figures 22 or 26.

Polite prefixes   The polite prefixes お +, ご + have been represented as modifiers of the nouns, adjectives, or noun-adjectives they refer to. Examples are shown in Figures 11 or 12.

2.2   Some choices to be made

Verbal morphology   Although in Tesnière's work on French, verbal morphology is not represented in stemmas, because, as morphology, it does not belong to the realm of syntax, the question has to be raised for Japanese.

As a matter of fact, all ます / ました -forms can be analysed as suspended forms of verbs (transference I → O) followed by a suffix. This suffix bears a semantic meaning, *i.e.* the politeness level. For parallelism, if one explains お帰りになりました on the syntactic level, then it would be logic to have also a syntactic analysis for 帰りました.

If this analysis is adopted, all our examples should be consequently modified[7].

Quantifiers   In the examples hereafter, quantifiers and quantity adverbials have been diversely analysed. It will be the responsability of the linguist in charge of tree-banking to propose an analysis of these forms, and consequently correct our trees (Figures 11, 14 and 18)

な -adjectives   The analysis of な -adjectives has also to be studied. may be the right analysis is to consider な as a transferer in a O → A transference. Our examples should thus be modified accordingly.

---

[7]We insist on the fact that *politeness* is the object of the syntactic analysis here, not *tense*. This latter is a morphological phenomenon in Japanese.

3 Tests

3.1 Introspection

The use of instropection as a means to study syntax, and hence to make choices for structural representation, is justified[Tesnière 59], by the fact that syntactical facts do not necessary have surface markers. In other words, there are syntactic facts which do not appear *per se* in given occurrences. The lack of external facts implies the necessity of an internal inspection of the speaking activity.

Now, introspective methods have been criticised for being subjective. They are surely experimental methods, but the facts they work on are not subjective: they are *abstract objective* facts, as syntactical facts and grammaticality are abstract but objective facts.

The second criticism, that of being a method based on intuition, is more relevant, and is totally justified in cases where the experimenter is not a true linguist. Introspection requires skills, practice and exactness.

In trying to determine what stemma is the exact one for a given syntactical phenomenon, one should consider two kinds of tests:

- **Positive** tests. From a given occurrence, one tries to build new occurrences by ([Gross 75]):

    - distribution or

    - transformation.

- **Negative** tests. Incompatibilities and ungrammaticality allow the experimenter to determine, by opposition, what the real meaning of the elements he tried to combine together is[8].

---

[8]Tesnière is famous for having shown in a simple manner, using the *composite word* method, that the French *imparfait* is a marker for "habit, custom", as the verbs "avoir l'habitude", "souloir" imply the use of this tense, and are incompatible with "passé simple".

3.2 Distribution

In a given utterance, one can replace one word by another and try to determine:

- what is common to the series of words commuting;

- what the characteristics of words are, which introduction in the occurrence would make it ungrammatical.

Also, in order to determine if a phenomenon is really syntactical or rather morphological, one has to consider how freely commutations occur, *i.e.* if commutation is reduced to a certain finite number of cases, if it is open, but constrained, or if it is unconstrained. Consequences implied by commutation play also a role: phonological transformations are typical of morphological phenomena.

3.3 Transformation

Given an utterance, one can also create other utterances, with the same or a mechanically derived meaning, by applying transformations, following Harris' invaluable work.

Transformations may be

- **purely syntactic transformations**, like the classical active-passive transformation, or negation, permutation, pronominalisation, extraposition, *etc.*;

- **morphological devices**, like derivations (*i.e.* substantivation of a sentence), *etc.*

3.4 Analogy and consistency

Whereas introspection methods are necessary to establish whether a fact belongs to morphology or syntax, or what the appropriate structural representation should be, the similarity of representations for similar phenomena, and hence the consistency of the tree-bank, should be ensured by the use of analogical reasoning.

BoardEdit, in its form used for tree-banking, will provide a device to perform analysis by analogy, *i.e.* there will be the possiblility of automatically creating a new structure for a new utterance, by analogy with the data already existing in the database.

4 Examples

We present some examples of stemmas drawn for sentences extracted from a corpus used in the TDMT system. The original numbers of sentences in the corpus appear in parentheses for reference.

が
|
I
|
OT
／＼
AT　　　の
／＼
IT　　＋いたい
／｜＼
を　０　する
｜　｜
０　AT
／＼
０　の
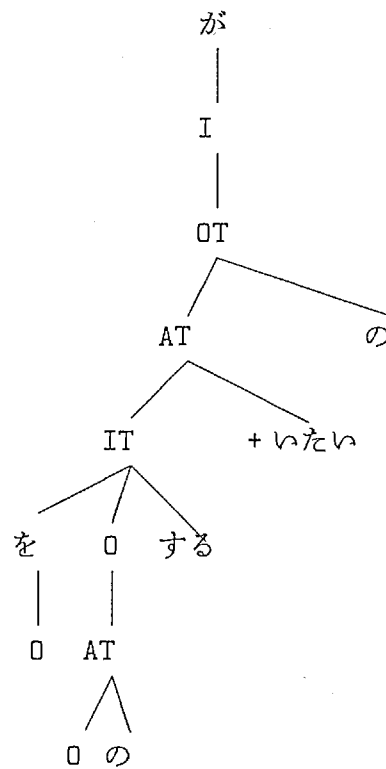
Figure 9: (AU4JP001) 今晩のホテルを予約したいのですが。

```
か
│
I
│
0
```

Figure 10: (AU4JP002) 何名さまですか。

```
I            IT
│           ╱│╲
0     を   0   します
      │    │
      0    お＋
      │
      AT
     ╱ ╲
    0   ∅
```
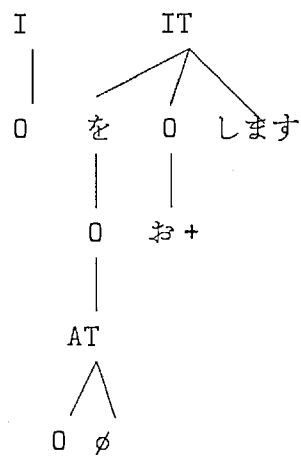
Figure 11: (AU4JP003) 一名です。シングルを一つお願いします。

```
E --,--              IT
                    /\
                 O    できます
                 |
                 ご +
```
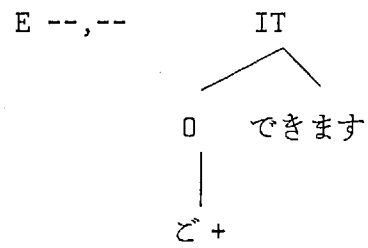
Figure 12: (AU4JP004) はい、ご用意できます。

```
       か
       |
       I
      /\
   は  いくら
   |
   O
```
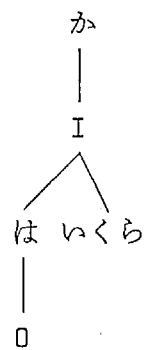
Figure 13: (AU4JP005) 部屋代はいくらですか。
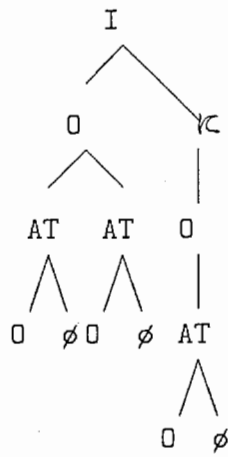
Figure 14: (AU4JP006) シングル一部屋十九万ウォンになります。

```
ET  --,--              が
 /\                    |
0   ∅                  I
                       |
                       AT
                      /   \
                    OT      よう
                   /  \
                  I    ∅
                  |
                  と
                  |
                  0
                  |
                  AT
                 /  \
                I    ∅
                |
                ET
               /\
              0  で
```
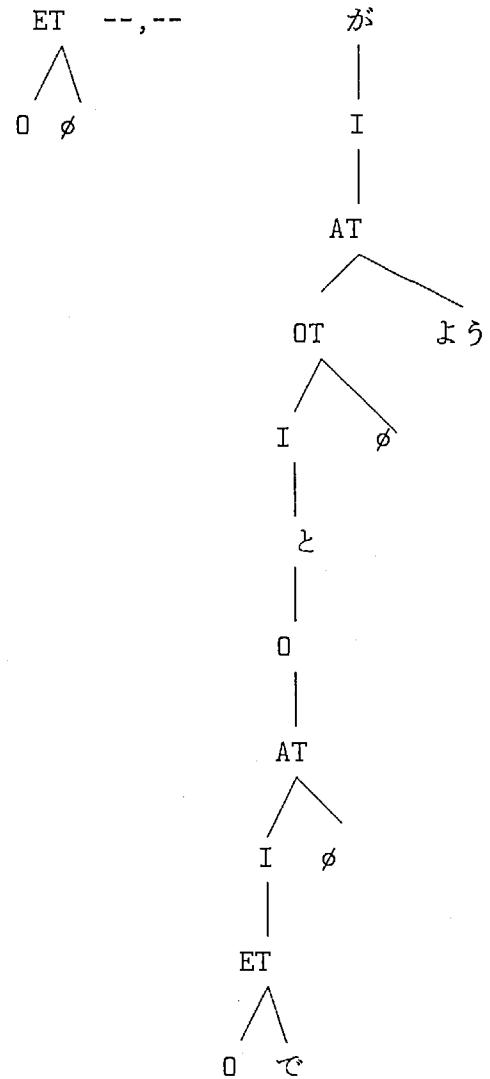
Figure 15: (AU4JP007) あれ、ガイドブックで見た料金と違うようです
が。

```
                          I
                 ┌────────┼────────┐
                ET        は       ET
               ╱  ╲        │      ╱  ╲
             OT    で     O    A   + く
            ╱  ╲
           I    の
           │
           は
           │
           O
```
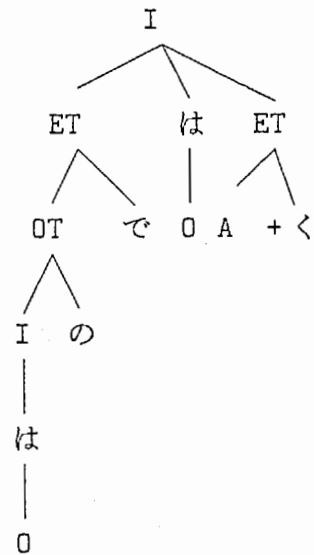
Figure 16: (AU4JP008) 今日は週末なので料金は高くなります。

```
                    IT
                   /  \
                  0    します
                 / \
                を   お+
                 |
                 0
                 |
                 AT
                /  \
              OT    の
             /  \
            I  + い
            |
            0
```
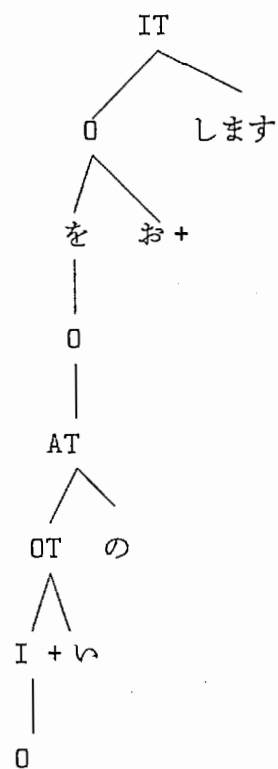
Figure 17: (AU4JP009) エアコン付きの部屋をお願いします。

```
                    I
                   / \
                  は   に
                  |    |
                  0   いくら
                  |
                  AT
                 / \
               0T    ∅
              / \
             I + い
             |
             0
```
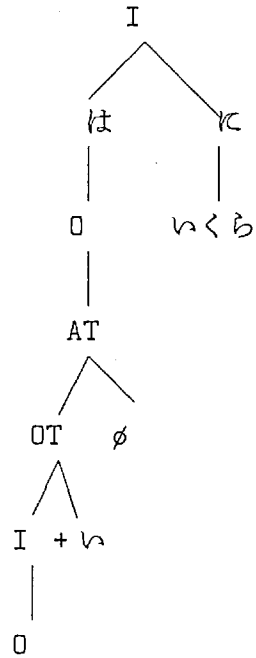
Figure 18: (AU4JP010) バス付きシングルはいくらになりますか。

```
                      が
                      |
                      I
                      |
                     OT
                    /    \
                   A       の
                   |
                   が
                   |
                   O
                   |
                   AT
                  /   \
                OT     の
                / \
               I + い
               |
               O
```
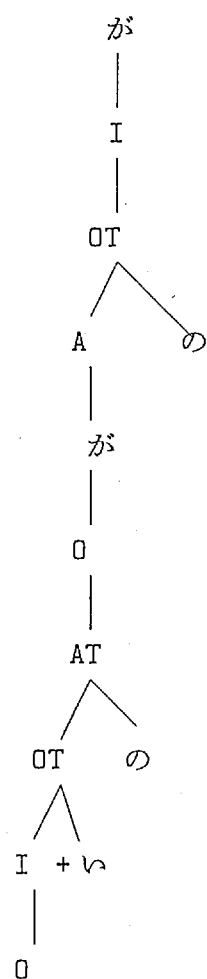
Figure 19: (AU4JP011) オンドル付きの部屋がいいのですが。

```
                    IT
                   /  \
                  O    します
                 / \
               を お +
                 |
                 O
                 |
                 AT
                / \
               O   ∅
```
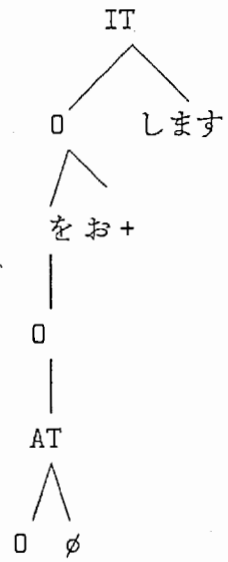
Figure 20: (AU4JP012) 二人部屋をお願いします。

```
か --,--     か
 |            |
 I            I
 |            |
 O.           A
              |
              が
              |
              O
```

Figure 21: (AU4JP013) ツインですか、ダブルがよろしいですか。

```
                    か
                    │
                    IT
                   ╱  ╲
                 ET    います
                ╱╲       │
              I  ＋て    は
              │          │
              に         0
              │
              0
              │
              AT
              ╱╲
             0  の
```
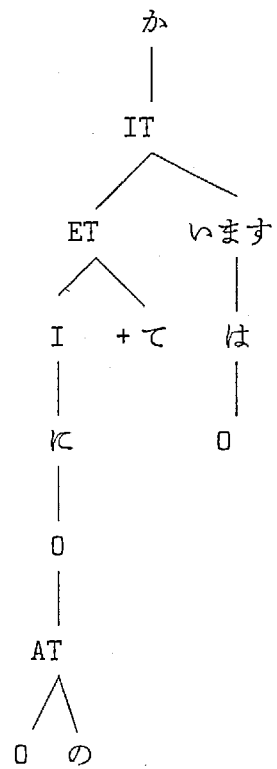
Figure 22: (AU4JP014) その部屋にシャワーはついていますか。

```
      が    --,--            IT
      |                     /\
      IT                  ET おります
     /\                  /\    |
  ETありません          I +て  は
  /\      |                     |
 I +て   は                    0
         |
         0
```
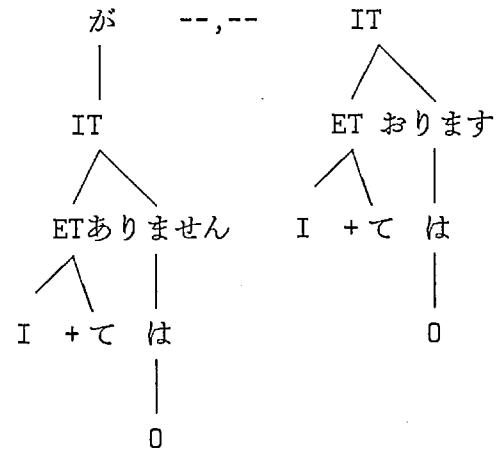
Figure 23: (AU4JP015) バスはついておりませんが、シャワーはついて
おります。

```
E --,--              IT
                    /|\
              ET    0  します
             /\      |
            0 で    お+
```
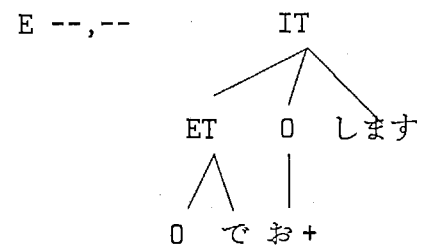
Figure 24: (AU4JP016) じゃあ、それでお願いします。

```
                       が
                       │
                       I
                       │
                       OT
                      ╱ ╲
                    AT     の
                   ╱╲
               I + いたい
                │
                0
```
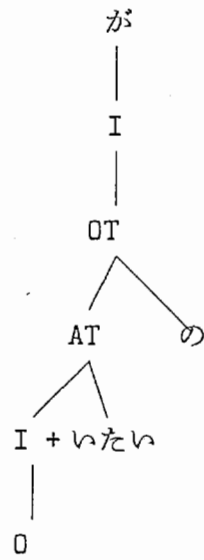
Figure 25: (AU4JP017) 二泊したいのですが。

```
                           か
                           │
                           IT
                    ┌──────┼──────┐
                    を      0   いたしましょう
                    │      │
                    0      ご +
                 ┌──┴──┐
                AT     お +
             ┌──┴──┐
            OT     な
           ╱  ╲
         AT   よう
        ╱ ╲
       0   の
```
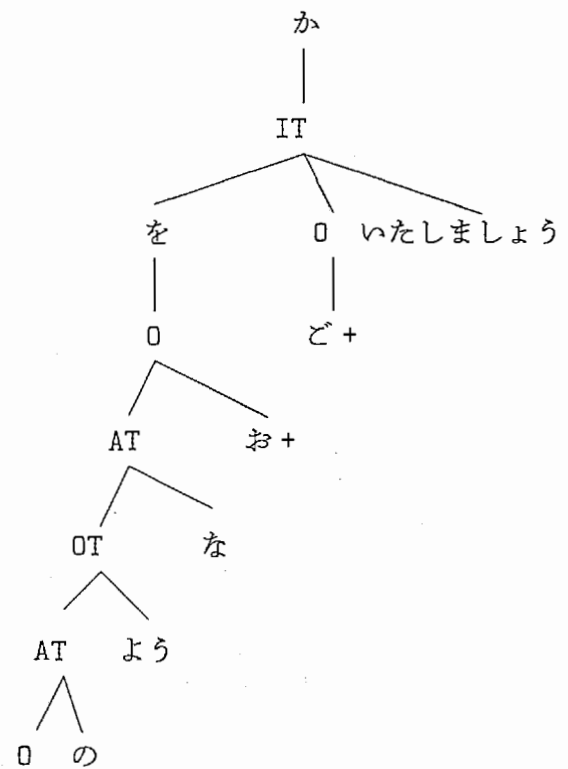
Figure 26: (AU4JP018) どのようなお部屋をご用意いたしましょうか。

```
                    か
                    │
                    I
                  ╱   ╲
                は        OT
                │        ╱  ╲
                0       I      ＋い
                │       │
                AT      は
                ╱╲    ╱  │  ╲
              0  の0 -- と -- 0
```
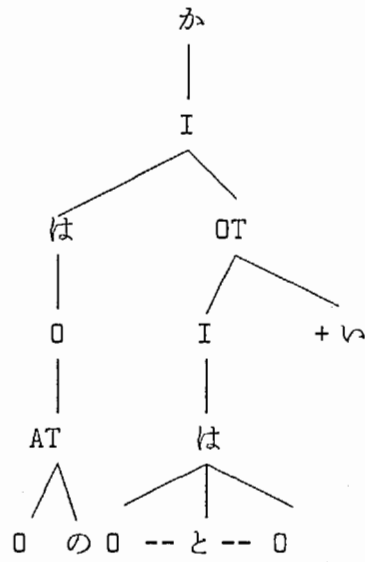
Figure 27: (AU4JP019) その料金は税金とサービス料は込みですか。

```
E--,--                    IT
                        ／      ＼
                     ET        おりません
                    ／＼
                 IT    ＋て
                ／＼
              I  ＋あれる
```
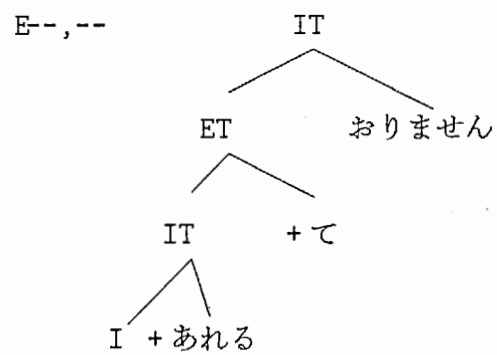
Figure 28: (AU4JP020) いいえ、含まれておりません。

```
                か
                │
                IT
               ╱  ╲
            ET    います
           ╱ ╲      │
         I   +て    は
                    │
                    0
```
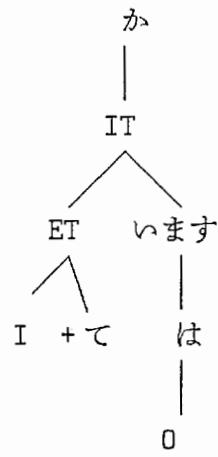
Figure 29: (AU4JP021) 朝食はついていますか。

```
        が
        │
        I
       ╱╲
      は  に
      │   │
      0   0
```
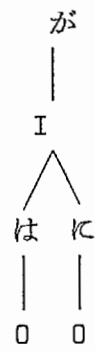
Figure 30: (AU4JP022) 朝食は別料金になりますが。

```
か
│
Ｉ
│
は
│
０
```

Figure 31: (AU4JP023) 前金はいりますか。

```
E --,--            I
                  / \
               ET    O
              / \
             O    で
             │
             AT
            / \
           O   の
```

Figure 32: (AU4JP024) いいえ、チェックアウトの時で結構です。

```
                        か
            ┌───────────┼───────────┐
           IT         -- と --       I
          ╱╲                         │
         O   する                     ET
                                     ╱╲
                                    A  + く
```
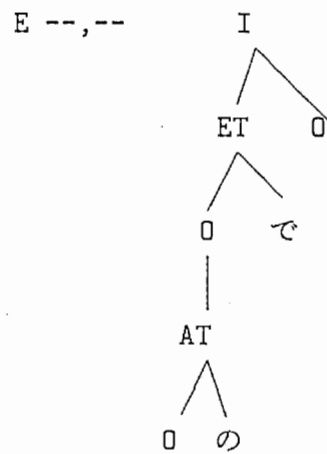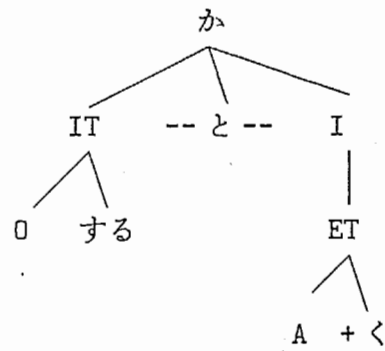
Figure 33: (AU4JP025) 連泊すると安くなりますか。
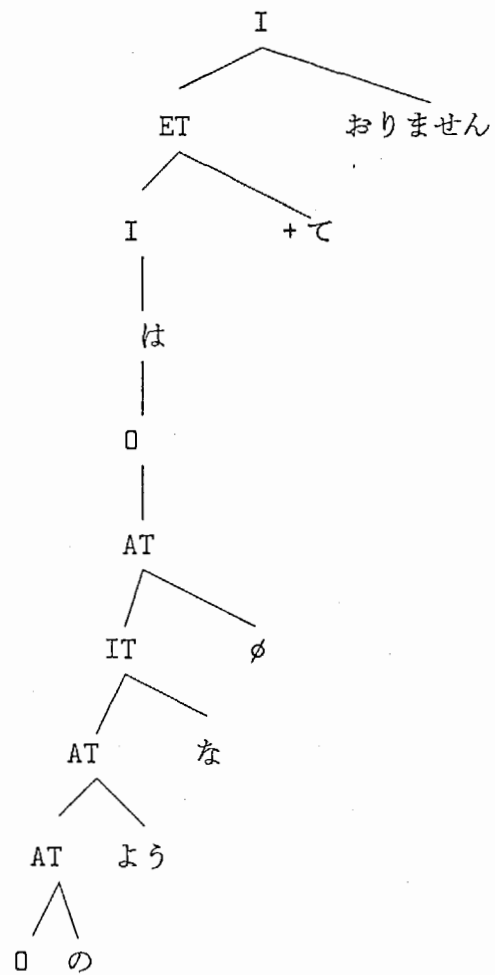
Figure 34: (AU4JP026) そのようなサービスはしておりません。

```
          か
          │
          I
         ╱ ╲
       ET    は
      ╱╲     │
     O  は    O
             │
             AT
            ╱ ╲
          OT    ∅
         ╱╲
        I ＋ い
```

Figure 35: (AU4JP027) こんばんは、空き室はありますか。

47

```
E --,-- I か     か
         |      |
         I      I
         |      |
         A      A
         |      |
         が     が
         |      |
         O      O
                |
                AT
               /\
              O   ∅
```
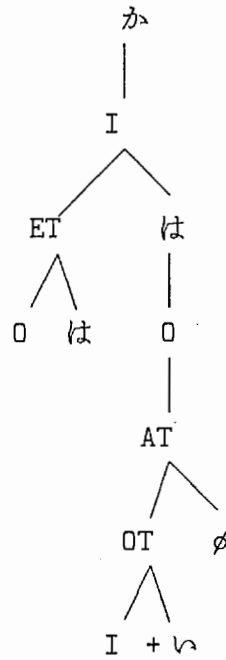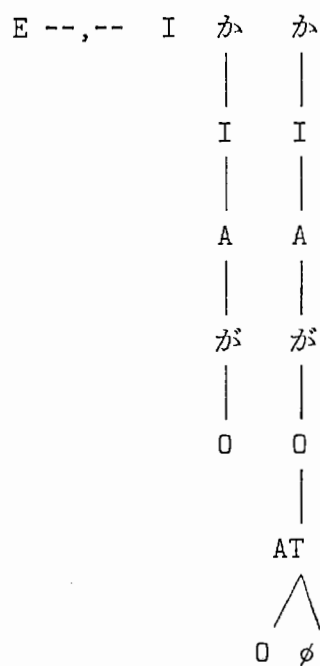
Figure 36: (AU4JP028) はい、あります。 — ベッドがよろしいですか。
— オンドル部屋がよろしいですか。

```
                              か
                              │
                              I
                              │
                             OT
                            ╱    ╲
                           I       の
                         ╱  ╲
                        は   くらい
                        │    │
                        O    AT
                             ╱ ╲
                            O  の.
```
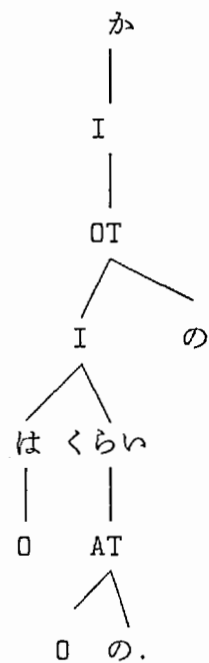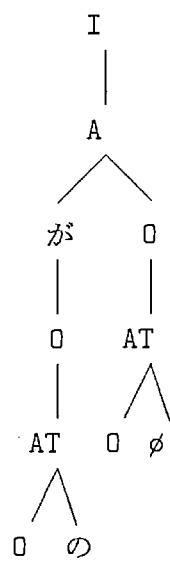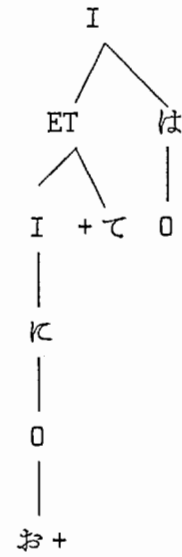
Figure 37: (AU4JP029) 料金はどのくらい違うのですか。

```
                I
                |
                A
               / \
             が    0
             |     |
             0    AT
             |    / \
            AT   0  ∅
            / \
           0   の
```

Figure 38: (AU4JP030) オンドルの方が一万五百ウォン高いです。

```
                    I
                   / \
                 ET    は
                / \     |
               I  +て   0
               |
               に
               |
               0
               |
               お +
```

Figure 39: (AU4JP031) お部屋によって料金はちがいます。

```
E --,--                    か
                           |
                           IT
                          ╱ ╲
                   ET        I  + あれる
                  ╱ ╲
              I     + て
             ╱ ╲
           を   に
           |    |
           0    0
           |    |
          お + ご +
```
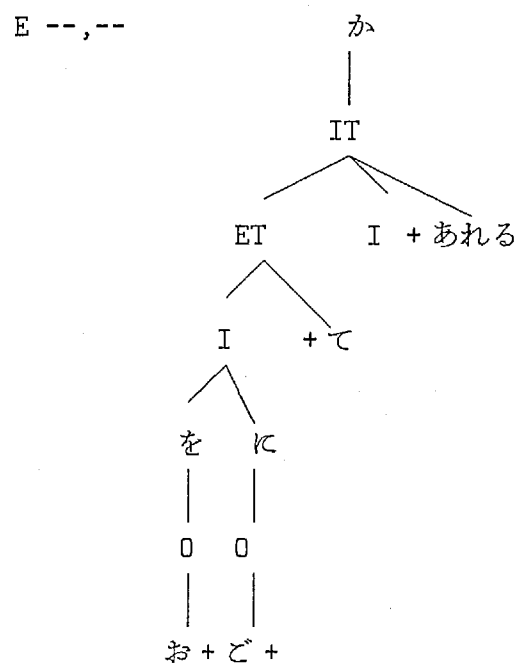
Figure 40: (AU4JP032) 直接、お部屋をご覧になって決められますか。

Conclusion

This report is intended as specifications for tree-banking using Tesnière's structural syntax representations. We presented the major features of structural syntax, and showed how to express them using forests and correspondences under BoardEdit.

A number of examples have be drawn for the purpose of illustratation.

Tree-banking is a linguistic task. The results of linguistic enquiries, such as distribution or transformation tests, will help the linguist in charge of tree-banking in making justified representational choices.

References

[Boitet and Zaharin 88] Christian Boitet and Zaharin Yusoff
Representation trees and string-tree correspondences
Proceedings of COLING-88, Budapest, 1988, pp 59-64.

[Gunji 87] Gunji Takao
Japanese Phrase Structure Grammar
D. Reidel Publishing Company, 1987.

[Gross 75] Maurice Gross
Méthodes en syntaxe. Régimes des constructions complétives.
Hermann, Paris, 1975.

[Lepage 94] Yves Lepage
Texts and Structures – Pattern-matching and Distances
ATR report TR-IT-0049, Kyoto, March 1994.

[Mizutani et al. 83] 水谷しずお、石棉としお、荻野たかの、賀来なお
こ、草薙ゆたか、青山ふみひろ
文法と意味 *I*
朝倉日本語新講座, 1983.

[Starosta 88] Stanley Starosta
The Case for Lexicase
Pinter Publishers, London and New York, 1988.

[Tesnière 34] Lucien Tesnière
Comment construire une syntaxe
Bulletin de la Faculté des Lettres de Strasbourg, 12e année,
n° 7, mai-juin 1934, pp. 219-229.

[Tesnière 59] Lucien Tesnière
Eléments de syntaxe structurale
Klincksieck, Paris, 1959.

Index