

リレーショナル文法に基づく機械翻訳	
David E. Johnson	
IBM 東京基礎研究所	
千代田区三番町5-19	

関係文法に基づく機械翻訳システムについて述べる。関係文法は異言語間の文法的な関係も一様な方法で扱うことができ、これによって機械翻訳において、原言語の解析と、目標言語の生成部分を独立にすることが可能になる。同時に、トランスフュー部分の設計が理論的にすっきりしたものになる。これは日本IBMの和英翻訳システムで用いられている。

A RELATIONAL-GRAMMAR APPROACH TO MACHINE TRANSLATION

David E. Johnson

IBM Tokyo Research Laboratory

5-19 Sanbancho, Chiyoda-ku, Tokyo 102, Japan

This paper introduces an approach to machine translation that is based on the universally oriented, linguistic theory of relational grammar. Relational grammar provides a framework for analyzing and representing the syntactic structure of all natural languages in a cross-linguistically uniform manner. This universal orientation enables the independent development of analysis and generation grammars, simultaneously increasing modularity and reducing the arbitrariness of the transfer component. The relational-grammar approach to MT is illustrated with respect to several examples from Japanese and English, involving so-called adversity passives and desideratives (-tai/ -te hoshii).

1.0 Introduction

This paper discusses an approach to machine translation that is based on the universally oriented linguistic theory of relational grammar (RG) (cf. Perlmutter and Postal 1974, Johnson 1974, and Johnson and Postal 1980). It is argued that RG provides a methodologically and theoretically sound framework within which to develop machine translation systems. RG provides a framework for analyzing and representing the syntactic structure of all natural languages in a cross-linguistically uniform manner. The universal orientation of RG permits the independent development of analysis and generation grammars, simultaneously increasing modularity and reducing the arbitrariness of the transfer component (i.e., transfer can be developed on the basis of theoretically coherent, comparative linguistics). Further, since RG has been applied to dozens of languages, it is empirically well-founded and so is a plausible candidate for the MT application (cf. the bibliography in Dubinsky and Rosen 1983).

In this paper, an RG approach to MT is illustrated with respect to several examples from Japanese and English. While these examples will be discussed informally, it should be noted that the required English relational generator, *Genie*, has been implemented (Johnson 1988b, Schindler in prep.) and is being used as the backend of the JETS Japanese/English machine translation project being developed by the Japanese Processing Group at IBM Japan's Tokyo Research Laboratory. Moreover, the requirements for implementing in JETS the analysis and transfer parts of the linguistic analyses discussed here are in place.

1.1 Relational Grammar

Two of the central tenets of RG are:

- that the syntax of every natural language is properly characterized in terms of a universal set of primitive grammatical relations/functions such as subject, direct object, indirect object, genitive, and
- that clauses are, in general, **multi-stratal**, meaning that they have more than one level of grammatical structure, where the deepest stratum is language independent (the latter is called here **canonical (relational) structure (CRS)**).

Since canonical structures abstract away from language-particular properties such as word order and morphology, they provide a linguistically natural interface between source and target language. That is, in contrast to the widely varying, superficial structures found in natural languages, canonical structures are relatively few in number and simple in shape. It should be noted that RG grammatical relations correspond neither to superficial cases nor to so-called deep cases in the Fillmorean sense; they are purely *syntactic* relations without any inherent/invariant meaning.

For example, "regular" passive clauses are universally taken to have the same **canonical (relational) structure** as the corresponding actives. As first proposed by Perlmutter and Postal (1974), the active/passive relation is universally characterizable in terms of a relation-changing rule that in the *generation* direction, demotes a subject to chomeur and advances a direct object to subject. Relation-changing rules in individual languages differ in the so-called *side effects*, associated with relational changes. So, in the case of English Passive, the side effects involve marking the verb for the past participle form, flagging the subject chomeur with the preposition *by* and, in most cases, introducing the auxiliary verb *be* (see Figure 1 on page 2).

While the use of primitive grammatical relations/functions was also later adopted by lexical-functional grammar, the multi-stratal assumption -- and the notion of clause-level, relation-changing rules this entails -- is maintained only by RG, clearly separating the two frameworks. It is precisely the distinctive characteristic of multiple relational strata which is argued to be of theoretical importance and practical use in the machine translation domain.

A machine translation system based on RG would have, in a qualified sense, some of the characteristics of a pivot system, in that languages can share canonical structures. Furthermore, canonical structures are semantically oriented in the sense that basic predicate/argument structure is, wherever possible, recovered and many superficially different grammatical forms are reduced to the same representation. In spite of this, an RG-based system would be a transfer system. The theoretical

Relational Changes

Direct Object --> Subject

Subject --> Chomeur

Side Effects

Insert (Prep = by) on chomeur

Insert (VFORM = Pastpart) on verb

(Splice "be" into the structure, with verb as its complement)

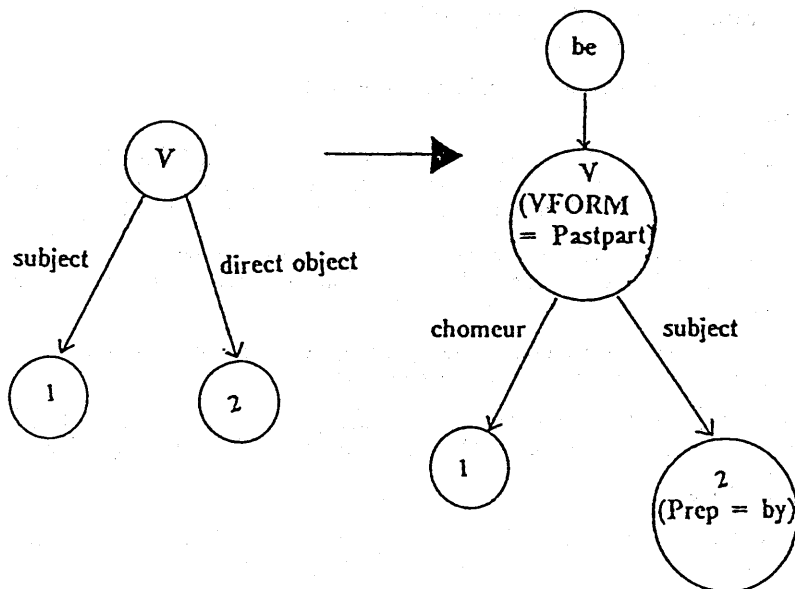


Figure 1. English Passive in Relational Grammar

reason for this is that there is no guarantee that the syntactically justified canonical structure for a given construction of the source language will be the canonical structure of some construction of the target language. As a matter of fact, there are frequently occurring construction types, e.g., passive, clause-union, subject (to subject) raising, subject-to-object raising, etc., which are amenable to uniform relational analysis. But, in general, transfer must map, in a truth-preserving manner, a canonical structure from Japanese into another canonical structure underlying a grammatical English sentence. A case in point is the Japanese so-called suffering/affected passive, which has no single analog in English, e.g.:

(1a) 私は雨に降られた。

watakushi wa ame ni furareta (I topic rain by rain-passive-past)

(1b) It rained on me. (cf. *I had it rain (on me))

(2a) 私は先生に絵を誉められた。

watakushi wa sensei ni e o homerareta (I topic teacher by picture particle praise-passive-past)

(2b) I had (a) picture praised by the teacher

(cf. *The teacher praised a picture on me)

1.2 Example Relational Grammar Analyses

In this section, RG analyses for sentences (1) and (2) above, as well as those below, will be sketched.

(3a) 私は日本へ行きたい。

watakushi wa Nihon e ikitai (I topic Japan to go-want)

(3b) I want to go to Japan.

(4a) 私は彼女に日本へ行ってほしい。

watakushi wa kanojo ni Nihon e itte hoshii (I topic her particle Japan to go want)

(4b) I want her to go to Japan.

(5a) 私は先生に絵を誉められたい。

watakushi wa sensei ni e o homeraretai

(I topic teacher particle picture particle praise-passive-want)

(5b) I want to have (a) picture praised by the teacher.

The relational rules needed to account for these sentences are:

Japanese Relational Rules

- **Equi Clause Union (ECU)** (collapses two clauses into one, deleting the lower subject coreferential to the higher subject and reassigns all other arguments to the new (unioned) clause with the same relations held originally), for sentences with -tai, e.g., (3 and (5),
- **Intransitive Clause Union (ICU)** (collapses two clauses into one, where the higher clause is intransitive, reassigning all arguments to the new (unioned) clause with the same relations held to the old clauses), for adversity passives, e.g., (1), (2), (5).
- **Subject-to-Indirect-Object Demotion (S-IO)** (demotes a subject to indirect object), for sentences like (2), involving hoshii and a complement clause,
- **Indirect-Object Passive (IO-PASS)** (advances an indirect object to subject and demotes the former subject to chomeur), for clauses like (1), (2) and (5) with adversity passives.

Transfer Relational Rules

- **Inversion (INV)** (advances a direct object to subject and demotes the subject to indirect object), for adversity passives like (2) which involve transitive predicates.
- **Intransitive Clause Union (ICU)** (as above)

English Relational Rules

- **Passive** (as above), for ...praised by the teacher in (5),
- **Subject-Deletion (EQUI)**
(deletes a complement subject co-referential to higher subject), triggered by want in (3) and (5), (3) and (5), and
- **Subject-to-Object Raising (B-RAISE)** (raises a complement subject up as higher direct object), for want in (4).

(For the parsing, the inverses of the above rules must, of course, be used.) Except for the analyses of adversity passives, which is admittedly controversial, the rules used are fairly standard RG ones, occurring in a variety of languages.

If we now take the above translation equivalents and work backwards from both sides, using the RG rules assumed above, the RG analyses shown in Figure 2 on page 5, Figure 3 on page 6, Figure 4 on page 7 and Figure 5 on page 8 can be constructed. As can be seen from these figures the work for transfer is quite minimal. Other than lexical selection, transfer must differentially process the Japanese CRS' for adversity passives. If the subordinate clause in an adversity-passive construction is *intransitive*, then the two clauses undergo ICU, with the upper predicate being unexpressed. On the other hand, if the subordinate clause is *transitive*, then the adversity predicate

rare is replaced by have and Inversion is applied to the upper (have) clause. The reader, no doubt, has noticed that the intransitive case involves both the *inverse* of ICU to arrive at the Japanese CRS (J-CRS) and then a subsequent application of ICU in transfer to arrive at the English CRS (E-CRS). In implementation, one might consider removing this "yo yo" effect, by not actually constructing the postulated J-CRS, but truly independent construction of the Japanese relational grammar and the English one would lead to the results shown here. The advisability of taking this "practical" short cut could only be determined by examining the consequences for transfer more generally. As can be seen from Figure 4 on page 7 and Figure 5 on page 8, the rules which work for the simpler cases function together in the more complex case (5) to produce the desired results. A summary of the rules used by the various components is shown in Figure 6 on page 8

1.3 Concluding Remarks

The key point of the above discussion is that the relational rules used are, by and large, those that would be arrived at *independently* of the machine translation task by applying the principles and methodology of RG to *each language separately*. More generally, by providing theoretically justified and empirically well-founded framework for natural-language syntactic analysis, RG supports modular design and sophisticated linguistic analyses.

1.4 References

- Dubinsky, S. and C. Rosen (1983) "A Bibliography on Relational Grammar through April, 1983 with Selected Titles on Lexical-Functional Grammar," distributed by Indiana University Linguistics Club, Bloomington, Indiana.
- Johnson, D. E. 1974. **Toward a Theory of Relationally Based Grammar**. University of Illinois PhD Thesis. Published by Garland Publishing Co., New York, 1979.
- Johnson, D. E. 1988a. "A Modular Design for Transfer and Generation in the JETS Japanese/English Machine Translation System," IBM Tokyo Research Laboratory Report TR87-1020.
- Johnson, D. E. 1988b. "The Linguistic Design of Post-Analysis in the JETS Japanese/English Machine Translation System," to appear in the **Proceedings of the International Conference on Fifth Generation Computers 1988**, Tokyo, Japan.
- Johnson, D. E. and P. M. Postal. 1980. **Arc Pair Grammar** Princeton University Press, Princeton, NJ.
- Perlmutter, D. M. and P. M. Postal. 1974. **Lectures on Relational Grammar** LSA Linguistic Institute, University of Massachusetts, Amherst, MA.
- Schindler, P. A. In prep. **An Object-Oriented System Shell for Relational-Grammar Based Natural Language Processing**, MS Thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

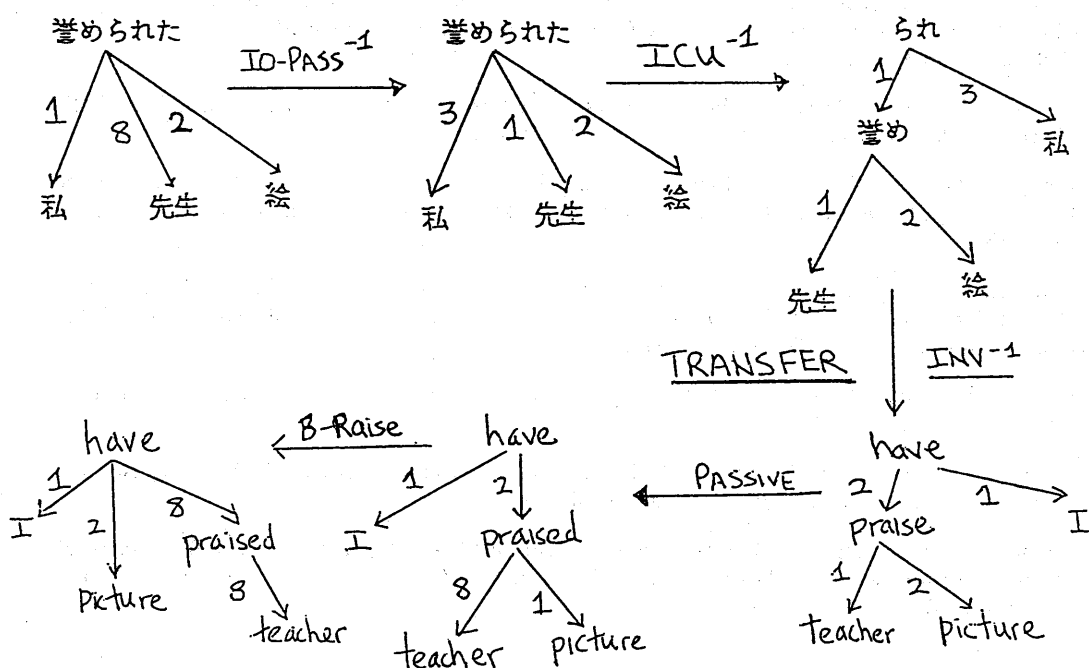
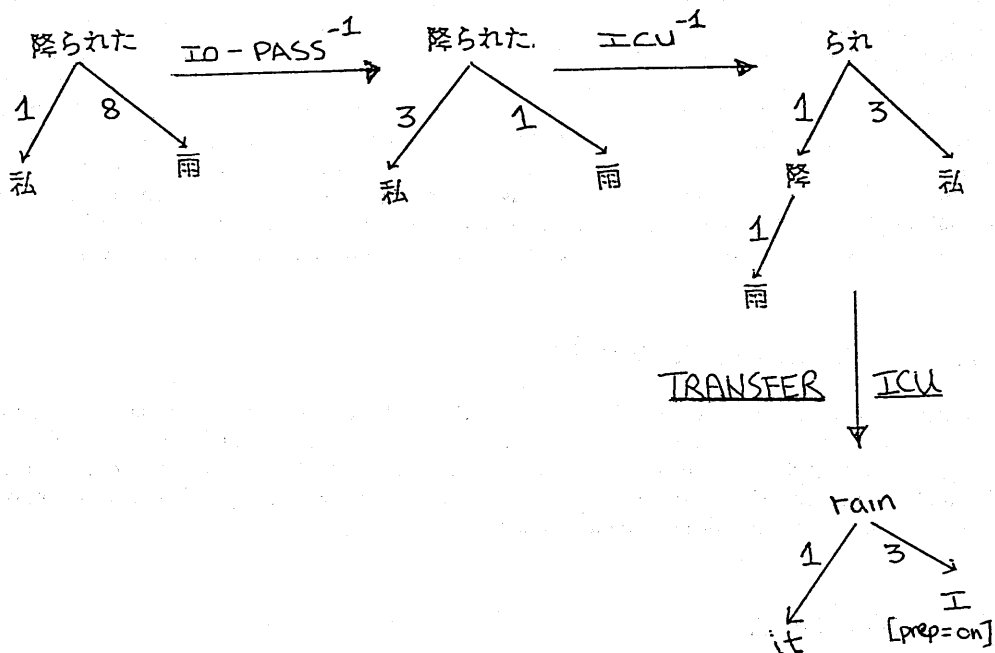


Figure 2. Adversity Passive (Sentences 1 and 2): 1 = subject, 2 = direct object, 3 = indirect object, 8 = chomeur, superscript -1 means "inverse".

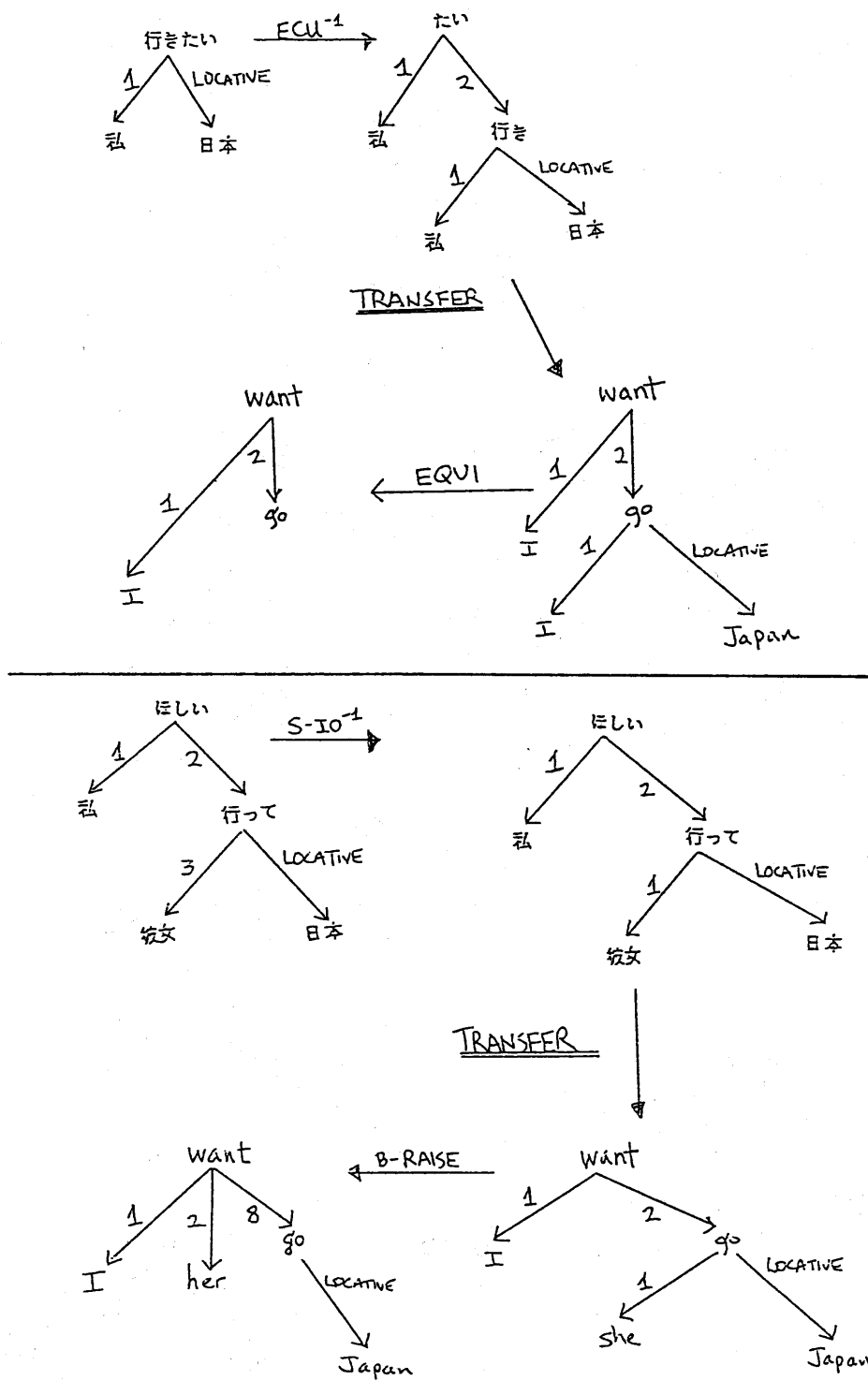


Figure 3. -tai/te hoshii (Sentences 3 and 4): 1 = subject, 2 = direct object, 3 = indirect object, 8 = chomeur, superscript -1 means "inverse".

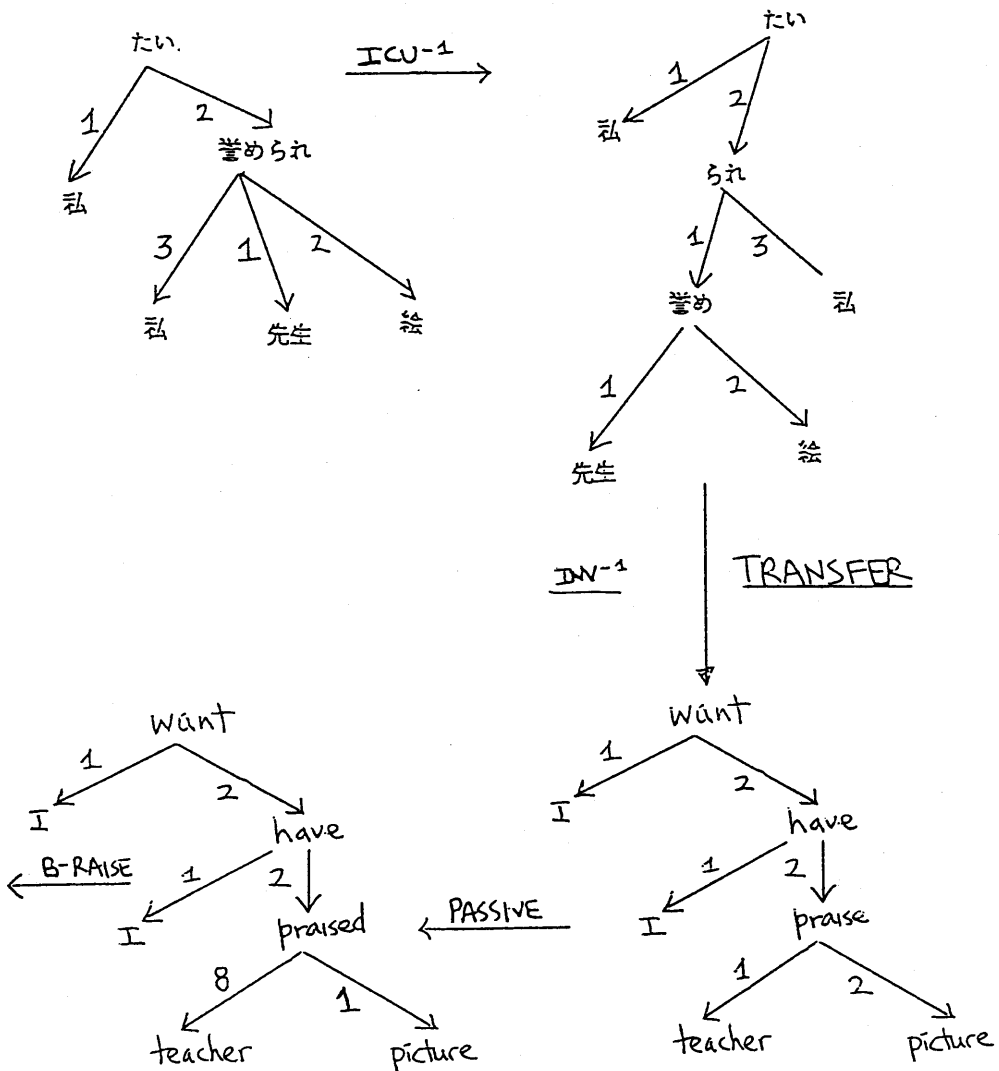
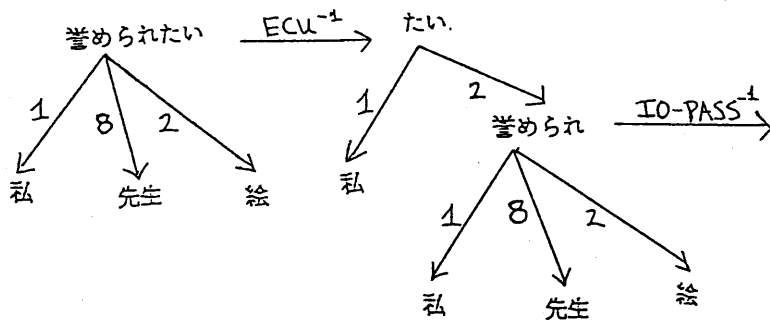


Figure 4. Sentence 5: 1 = subject, 2 = direct object, 3 = indirect object, 8 = chomeur, superscript -1 means "inverse".

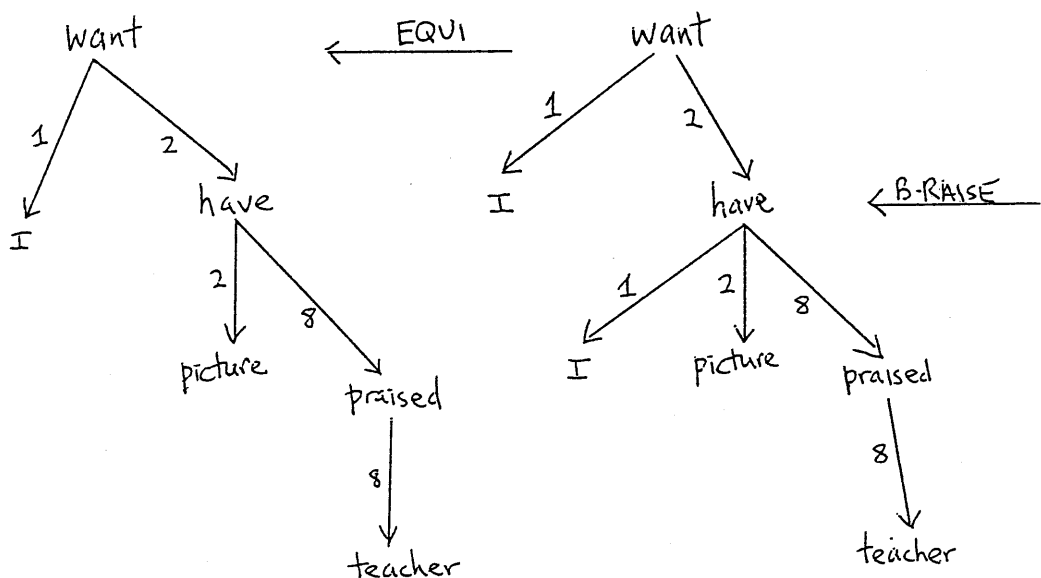


Figure 5. Sentence 5 Continued: 1 = subject, 2 = direct object, 3 = indirect object, 8 = chomeur, superscript -1 means "inverse".

Sent	Japanese Analysis	J/E Transfer	English Generation
1	IO-PASS ⁻¹ ICU ⁻¹	ICU	
2	IO-PASS ⁻¹ ICU ⁻¹	INV	PASSIVE B-RAISE
3	ECU ⁻¹		EQUI
4	S-IO ⁻¹		B-RAISE
5	ECU ⁻¹ IO-PASS ⁻¹ ICU ⁻¹	INV	PASSIVE B-RAISE EQUI

Figure 6. Rule Summary by Component: superscript -1 means "inverse"