Formal Description of Grammar and Translation (I)

Akihiro NOZAKI

Institute of Mathematics, College of General Education, University of Tokyo.

(Received February 18, 1965)

Introduction

In dealing with mechanical translation of Japanese scientific papers into English, the author realized the utility of a formal description of grammar and of the translation procedure. It will facilitate the software building adequate for mechanical translation, and once such a software is built up, the coding of the translation will be made much easier.

In the following, we shall describe an attempt of formalizing the procedure of translation in the scheme of Chomsky's mathematical theory of language. After the general description, simple models of translation procedure will be illustrated.

1. Formal Description of Grammars

Let us start by a brief summary of the description of the grammar after Chomsky [1, 2] with some slight modifications.

A language L is a set of strings of *terminal symbols*. One considers words or occasionally letters of alphabet as terminal symbols, but in the following we shall consider only words as these symbols, and call them rather *basic symbols* of L. The set of all these symbols will be denoted by B.

We consider grammatical concepts such as: sentence, subject, predicate etc. as *non-basic symbols*, and denote by N the whole set of these symbols. We shall denote by $\sum$ the union of B and N.

We have the rules of grammar such as

"S (meaning "sentence")→Subject+Predicate".

Expressions like "S", "Subject+Predicate" will be called $\sum$-strings, and denoted in general by Greek letters, α or β. A $\sum$-string consisting solely of the elements of B is a *B-string*. (Ex. "I hit him.")

The whole set of rules is denoted by F. We shall denote by $f_1, f_2, \cdots$ the elements of F and write like

$$f\colon \alpha \longrightarrow \beta,$$

$$\alpha \overset{f}{\longrightarrow} \beta$$

or occasionally $\beta \overset{f}{\longleftarrow} \alpha$.

Examples: f_1: S→Subject+Predicate

f_2: Predicate→Transitive Verb+Object

$S \xrightarrow{f_1}$ Subject+Predicate $\xrightarrow{f_2}$ Subject+Transitive Verb+Object.

The *grammer* G over the set B is the triple $G=(B, \sum, F)$. A *derivation* in G is a diagram of the form:

$$\alpha_1 \xrightarrow{f_1} \alpha_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-1}} \alpha_n \qquad (1)$$

where α_i are $\sum$-strings and f_i are rules of F.

If $\alpha_1=S$ (special symbol contained in $\sum$, meaning "sentence"), the derivation (1) is called *generative*. A generative derivation ending with a B-string α_n is called a *terminated generative derivation* and α_n will be called a *derived string*.

The set of all derived strings in G will be denoted with $D(G)$. If $L=D(G)$, G is called a *generative grammar* of L. Chomsky described solely this kind of grammars.

Now, if α_1 is a B-string, the derivation (1) will be called *analytical*. An analytical derivation ending with $\alpha_n=S$ will be called a *terminated analytical derivation*, and the initial term α_1 of such a derivation an *admissible string*. The set of all admissible strings of G will be denoted by $A(G)$.

If $L=A(G)$, G is called an analytical grammar of L.

Duality.

Let f: $\alpha \to \beta$ be a rule of F. Then the rule $\beta \to \alpha$ is called the *dual* of f and denoted by f^*. We shall write

$$F^*=\{f^*: f \in F\},$$

and $G^*=(B, \sum, F^*)$ when $G=(B, \sum, F)$.

Obviously we have $G^{**}=G$, $D(G)=A(G^*)$, $A(G)=D(G^*)$, and G is a generative grammar of L if and only if G^* is an analytical grammar of L. This duality transfers every fact in generative grammars to a corresponding fact in analytical grammars.

2. Formal Description of Translation

Now we can give a formal description of translation procedure from a language L to another language L'.

Let G be an analytical grammar of L, and G' a generative grammar of L'. Then a *translation procedure* Φ will be a mapping from a terminated analytical derivation Δ of G to a terminated generative derivation Δ' of G':

$$\Delta: \alpha_1 \xrightarrow{f_1} \alpha_2 \longrightarrow \cdots \xrightarrow{f_{n-1}} \alpha_n \qquad (\alpha_n=S),$$

$$\Delta': \beta_m \xleftarrow{g_{m-1}} \beta_{m-1} \longleftarrow \cdots \xleftarrow{g_1} \beta_1 \qquad (\beta_1=S).$$

$$\Delta'=\Phi(\Delta).$$

When we are to translate a given B-string α_1 in L into L', we have first to find the derivation Δ ("parsing"), and then apply Φ to Δ. The derived string β_m of $\Delta'=\Phi(\Delta)$ will be the translated sentence.

It would be most convenient, if the grammars G, G' are so constructed that Φ can be given by a "contravariant functor" φ: $F \to F'$:

$$\Delta:\ \alpha_1 \xrightarrow{f_1} \alpha_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-1}} S,$$
$$\Phi(\Delta)=\Delta':\ \beta_n \xleftarrow[\varphi(f_1)]{} \beta_{n-1} \xleftarrow[\varphi(f_2)]{} \cdots \xleftarrow[\varphi(f_{n-1})]{} S.$$

This is in fact possible in simple cases as the following:

Example 1. L: Japanese, L': English.

The set F of rules for L used here:

f_1: watashi→Noun
f_2: Noun+wa→Subject
f_3: anata→Noun
f_4: Noun+ni→Indirect Object
f_5: hon→Noun
f_6: Noun+wo→Direct Object
f_7: Subject+Indirect Object+Direct Object +ataeru→S.

Using these rules, the sentence

α_1=watashi+wa+anata+ni+hon+wo+ataeru

will be analyzed as follows:

$\alpha_1 \xrightarrow{f_1}$Noun+wa+anata+ni+hon+wo+ataeru
$\xrightarrow{f_2}$Subject+anata+ni+hon+wo+ataeru
$\xrightarrow{f_3}$Subject+Noun+ni+hon+wo+ataeru
$\xrightarrow{f_4}$Subject+Indirect Object+Hon+wo+ataeru
$\xrightarrow{f_5}$Subject+Indirect Object+Noun+wo+ataeru
$\xrightarrow{f_6}$Subject+Indirect Object+Direct Object+ataeru
$\xrightarrow{f_7}S.$

This is Δ.

Now we set:

$\varphi(f_1)$: Noun→I
$\varphi(f_2)$: Subject→Noun
$\varphi(f_3)$: Noun→you
$\varphi(f_4)$: Indirect Object→Noun
$\varphi(f_5)$: Noun→book
$\varphi(f_6)$: Direct Object→Noun
$\varphi(f_7)$: S→Subject+give+Direct Object+to+Indirect Object

Then $\Phi(\Delta)=\Delta'$ will be:

$S \xrightarrow{\varphi(f_7)}$Subject+give+Direct Object+to+Indirect Object
$\xrightarrow{\varphi(f_6)}$Subject+give+Noun+to+Indirect Object
$\xrightarrow{\varphi(f_5)}$Subject+give+book+to+Indirect Object

$\xrightarrow{\varphi(f_4)}$Subject+give+book+to+Noun
$\xrightarrow{\varphi(f_3)}$Subject+give+book+to+you
$\xrightarrow{\varphi(f_2)}$Noun+give+book+to+you
$\xrightarrow{\varphi(f_1)}$I+give+book+to+you

The output sentence lacks in the article. A more correct English sentence would be:

I give a book to you.

This final stage is within the realm of transformational grammar of L'. This final transformation will be denoted by Φ'.

Example 2. L: English, L': French.

The set F of rules for L used here:

f_1: a+book→Book
f_2: Book→Noun
f_3: you→Pronoun
f_4: I+give→I+Give
f_5: Give+Pronoun+Noun→Predicate
f_6: I→Subject
f_7: Subject+Predicate→S

The image of the mapping φ:

$\varphi(f_1)$: Livre→un livre
$\varphi(f_2)$: Noun→Livre
$\varphi(f_3)$: Pronoun→vous
$\varphi(f_4)$: Donner→donne
$\varphi(f_5)$: Predicate→Pronoun+Donner+Noun
$\varphi(f_6)$: Subject→Je
$\varphi(f_7)$: S→Subject+Predicate

An example of translation using φ:

Δ: I+give+you+a+book
$\xrightarrow{f_1}$I+give+you+Book
$\xrightarrow{f_2}$I+give+you+Noun
$\xrightarrow{f_3}$I+give+Pronoun+Noun
$\xrightarrow{f_4}$I+Give+Pronoun+Noun
$\xrightarrow{f_5}$I+Predicate
$\xrightarrow{f_6}$Subject+Predicate
$\xrightarrow{f_7}S$.

$\Phi(\Delta)$: $S\xrightarrow{\varphi(f_7)}$Subject+Predicate
$\xrightarrow{\varphi(f_6)}$Je+Predicate

$\xrightarrow{\varphi(f_5)}$Je+Pronoun+Donner+Noun
$\xrightarrow{\varphi(f_4)}$Je+Pronoun+donne+Noun
$\xrightarrow{\varphi(f_3)}$Je+vous+donne+Noun
$\xrightarrow{\varphi(f_2)}$Je+vous+donne+Livre
$\xrightarrow{\varphi(f_1)}$Je+vous+donne+un+livre

The output sentence: Je vous donne un livre.

Remarks

In the preceding, we have just given simple models, and there are many difficult problems left to solve. E.g. the "firal transformation" in Example 1 involves the following problem.

It is not clear that whether "Hon (book)" in Japanese is in singular or in plural. In our Φ translation, it was translated into "book" (in singular). Then it would be possible to supplement the article "a" mechanically, but to obtain a good translation which takes also the context into account, we should have recourse to pre-editing. (cf. [3])

The vernacular Japanese sentences are generally not written following the rigid grammar; they are often lacking in subjects, etc. To apply our Φ-method, such sentences have to be pre-edited and brought to some standardized form. It would be most desirable to extend our method, so that we can apply it also to incomplete sentences to obtain "good translations".

However, our method applies to a good many examples of translation of (pre-edited) Japanese (in a scientific paper) into English, as we are intending to show in subsequent papers. (cf. [4], [5], [6])

Acknowledgement

The author wishes to express his sincerest gratitude to Prof. S. Iyanaga for his kind support in elaboration of this paper. Thanks are also due to Dr. T. Murata and other members of Hitachi Central Research Laboratory for their kind help in the experiment of mechanical translation using an electronic computer HITAC 5020.

References

[1] Chomsky, N., *Syntactic structures*, Mouton & Co., the Hague (1957).
[2] Chomsky, N., On certain formal Properties of grammars, *Information and Control*, **2**, 137-167 (1957).
[3] Iyanaga, S., Procedure for pre-editing Japaneses, *Preprints for Seminar on Mechanical Translation*, April 1964 TOKYO (1964).
[4] Nozaki, A., Mechanical translation of pre-edited Japanese, *ibid.*
[5] Nozaki, A., Report on the MT experiment (I) (in Japanese), *MT Research Committee, Information Processing Society of Japan*, Report No. 2-1 (1964).
[6] Nozaki, A., Report on the MT experiment (II) (in Japanese), *MT Res. Comm., Information Processing Society of Japan*, Report No. 3-2 (1965).