Presenting the Predicate Information in the Corpus of Input Language for English-Korean Machine Translation

Kim Song Jon

The great leader Comrade **Kim Jong Il** said as follows.

"We are now faced with the heavy task of developing national science and technology to a new and higher stage by making a revolutionary turn in scientific research work."
("**KIM JONG IL** SELECTED WORKS" Vol. 12 P.198)

We are now faced with the heavy task of developing national science and technology to a new and higher stage by making a revolutionary turn in scientific research work.

Today when the whole country is waging the intense drive of breaking through the cutting edge, it is also one of the important problems in developing national science and technology to a new and higher stage.

Machine translation (MT below) needs an electronic corpus since it deals with an actual text which contains a huge amount of vocabulary, expressions and various kinds of sentences.

Such an electronic corpus for machine translation as ideal linguistic resources should be furnished with required information which is effectively used for translating input actual language materials.

This essay aims at clarifying how to present input sentences' predicate information in the corpus for input language. Presentation of the predicate information in the corpus and effective use of it in MT wants, first of all, predicates in English sentences to be set up in conformity with MT features.

Main parts of English sentences are subject and predicate and so discernment of subject and predicate is important in conducting the whole processes of English-Korean(E-K below) MT. That is because subject and predicate which is framework of sentence are closely related.

Subject search of a sentence is more difficult than that of its predicate in English whose morphological marks are quite limited. So, we set up predicate as a pillar unit and search subject and other parts of sentence depending upon the information of predicate after its search.

Let us consider the role of predicate as a mainstay in the following example.

e.g.: Electronic measurement and control pervade all the corners of science and engineering.

In the above example, a word, *pervade*, has the only verbal property and a word, *control*, has the potential properties of both noun and verb. The word, *control*, is surrounded with the word, *pervade*, on its back side and with a coordinative conjunction, *and*, on its front side. In such a contextual condition, the word, *control*, can by no means be realized as a verb. Therefore, the word, *pervade* alone can be predicate and the word, *control*, is confirmed as a noun. And a word, *all*, that can be used as a verb or an indefinite pronoun, is placed in such a context that before it there is a predicate and behind it there is the definite article, *the* and therefore, the word, *all*, is confirmed as an indefinite pronoun.

As can be seen in the example, based on the given sentence structures, conducted are most translation processes in EKMT and here a predicate plays a role as a mainstay.

We established a predicate in a sentence unlike in the traditional grammar so as to make clear the

decision of predicate which is a mainstay for all translation processes. We never regarded a predicate as an integrated part of sentence which bears its predication but confined a predicate to a lexical unit with verbal properties which completes its predication when a predicate is realized in an analyzed form.

This means that a predicative with the lexical meaning of an integrated predicate is separated from such a predicate and after all this means that a link verb is considered to be an independent predicate. And this also means that an auxiliary verb alone is regarded as a predicate out of a compound verb predicate which is composed of an auxiliary verb and an independent verb. We also regard as a none-predicate an independent verb which was called a partial predicate or an omitted one in a traditional grammar when an auxiliary verb which was used in the front part of sentence is omitted in front of an independent verb placed in the back part of the sentence.

That is because it confuses the search by computers which act mechanically according to instructions if a different unit is regarded as another variety of predicate since an auxiliary verb is decided as a predicate. In this case, we deem an omitted auxiliary verb to be an “omitted predicate.”

We severed only the first component from the traditional grammar’s compound predicate which is composed of one or several auxiliary verbs and one independent verb to set up a predicate suitable to the MT processes and named it an “auxiliary verb predicate.”

e.g.: He can speak English.

They will come to us tomorrow.

The Korean equivalent to an auxiliary verb predicate was especially established as that corresponding to an auxiliary verb severed from the whole Korean equivalent to the compound verb predicate. By so doing, could it be smoothly settled in conformity with the peculiarities of MT to set up an auxiliary verb predicate and complete the translation of independent verb by means of this predicate.

English predicates composed of link verb and predicative belong to a compound nominative predicate, a type of traditional English predicate.

We severed only such a link verb from the combination structure for a traditional predicate unit and set up this link verb as an independent predicate. Modern English has such a typical link verb as *be*. Besides this, some verbs such as *get, become,* etc are used as a link verb, preserving somewhat their proper lexical meaning.

e.g.: They were boys

They are diligent.

Taking it into consideration that this link verb predicate grammatically completes a predicative which keeps its lexical meaning for an explanation of a subject and connects such a predicative to it, we regarded this predicate as one type of predicate different from an independent verb predicate and named it a “link verb predicate.”

The Korean equivalent to a link verb predicate was established as that to an auxiliary verb severed from the whole Korean equivalent to the compound nominal verb predicate. By so doing, establishment of a link verb predicate and translation of the part led by it could be briefly settled in accordance with the peculiarities of MT.

We took into consideration the corresponding peculiarities of the components of both English and Korean predicates and set up an English predicate consisting of one independent verb unit as an "independent verb predicate" in favour of compiling Korean equivalents. Of course, such an independent verb predicate should go through its decision process because it has no sure predicate mark which always shows a predicate qualification.

e.g.: The first lesson begins at 8 o'clock.

I met him in the street.

When a predicate can always be guaranteed as a sure predicate without going through its decision process, we set up such a predicate as an "absolute predicate."

Absolute predicate involves, above all, the verbs of singular form and third person after their decision of part of speech. The words can and will are treated as an absolute predicate only when they are decided as an auxiliary verb because they are multipart-of-speech words.

Absolute predicate also involves the form words of verb *be* which show tense, person and number such as *is, am, are, was, were*, the form words such as *has, does* and *did* only when they are used as an auxiliary verb, the auxiliary verbs which are used in a fixed form such as *could, may, might, shall, should, would, must*. And absolute predicate also includes the phrasal auxiliary verbs such as *has to, had to, aught to, etc*. and the special past forms of verbs which are clearly different from the past participle forms of verb (e.g. went, came).

A mark P which represents "predicate information" is, first of all, placed in front of the very lexical unit which can be decided as predicate without going through predicate decision process. A relevant information is presented in front of other predicate units when predicate decision processes are completely finished in the input sentence.

- ＼P＼::predicate(one verb or one phrasal verb)
- ＼P1＼::singular predicate
- ＼P2＼::plural predicate
- ＼EP＼::extended predicate
- ＼HR1P＼::singular predicate has
- ＼HR2P＼::plural predicate have
- ＼HDP＼::past predicate had
- ＼BR1P＼:: present singular predicate is(of be)
- ＼BR2P＼:: present plural predicate are(of be)
- ＼BRAP＼:: past singular predicate am (of be)
- ＼BD1P＼:: past singular predicate was (of be)
- ＼BD2P＼:: past plural predicate were (of be)
- ＼OP＼::auxiliary verb-omitted predicate(the place where the auxiliary verb was)
- ＼OOP＼::several elements-omitted predicate
- ＼CP＼::imperative-form predicate
- ＼MP＼:subjunctive-mood predicate
- ＼AVP＼::auxiliary verb predicate without any change of verb form
- ＼VR1P＼::verb present singular third personal predicate

- ＼VR2P＼::verb present plural indefinite predicate
- ＼VDP＼::verb past predicate
- ＼DR1P＼::verb present singular predicate, does
- ＼ABP＼::absolute predicate

Here a figure 1 indicates singular and a figure 2 plural.

The necessary ones out of these marks for predicate information should be placed in front of the relevant predicate unit in each input sentence so that all these marks for predicate information are shown in the corpus for KEMT.

e.g.: • ＼SB＼The problem ＼BR1P＼is finding the time to get things done.

- ＼SB＼The troops ＼BD2P＼were to cross the river.
- ＼SB＼That room ＼HR1P＼has only two windows.
- ＼SB＼Flax ＼DR1P＼does well after wheat.

Here a mark ＼SB＼ indicates the beginning of a sentence.

The verb present indefinite plural predicate which coincides with a verb indefinite in its form indicates predicate information according to contextual conditions like this.

- …they/I/we/you(nominative case)＼VR2P＼…
- ＼SB＼~N2 which ＼VR2P＼VI ‿ ＼VR2P＼…
- ＼SB ~N2＼CJS＼that PN2/~N2 ‿＼VR2P＼…
- ＼SB＼~N2＼VR2P＼VI ‿ and ＼VR2P＼…
- ＼SB＼[AD/PP][,]~N2[AD]＼VR2P＼…
- They ＼VR2P＼learn Korean from an excellent teacher.
- We＼VR2P＼go to school together.

In the above contextual conditions, the verb indefinite always plays a role of a verb present indefinite plural predicate.

I, a pronoun of first person, is always used as a nominative case and plays a function of predicate. When a verb infinitive form is a predicate for that pronoun, VR2P (a present indefinite plural predicate), an information mark, is placed in front of that predicate. *You*, a pronoun of second person, can play a function of subject or an object in a sentence when it is not used together with a preposition. Therefore, contextual conditions should be properly confirmed and only when *you* are a subject, a plural predicate is placed behind it and a suitable predicate information mark can be presented in front of such a predicate.

When a verb past form plays a function of predicate, such a predicate is used together with singular or plural nominative cases of first, second and third persons.

The contextual condition of verb past predicate is formalized and thereby an information mark for such a predicate is set forth like this.

- …they/I/we/you(nominative case)/he/she/it(nominative case)＼VDP＼…
- …that(CJS) ＼VDP＼…
- …which(CJS) ＼VDP＼…
- ＼SB＼~N‿ Prep Wh-Clause ~N‿ ＼VDP＼ …
- ＼SB＼ The customer ＼VDP＼claimed there ＼BD2P＼were

problems with the software, and problems that ＼VDP＼resulted in missing or incorrect data used by medical personnel.

Verb past predicates are used in all the above contextual conditions regardless of a singular doer or plural doers.

If an auxiliary verb predicate is omitted, an omission information is given like this.

e.g.: • ＼SB＼They require that he＼OP＼appear.

• ＼SB＼He command that it＼OP＼ be done.

Like this, we placed the predicates information in the corpus used in EKMT and can compare the input sentences with the corpus furnished with English predicate information sentence by sentence to find predicate units with ease. And we can confirm the other information, too, briskly and correctly based on predicate information, thus guaranteeing a smooth translation of given text.

We will make a contribution to building a socialist powerful nation with a practical success of MT by rationally settling difficult problems arising from EKMT.