JEIDA's Test-Sets for Quality Evaluation of MT Systems
-- Technical Evaluation from the Developer's Point of View --

Hitoshi ISAHARA

Kansai Advanced Research Center
Communications Research Laboratory
588-2, Iwaoka, Iwaoka-cho, Nishi-ku, Kobe, 651-24, JAPAN
E-mail. isahara@crl.go.jp

abstract

This paper describes a method of evaluating quality for developers of machine translation systems to easily check imperfections in their own systems. This evaluation method is a systematic, objective method along with test example sets in which we clarified the evaluation procedure by adding yes/no questions and explanations to the example sentences for evaluation. In March 1995, our two test-sets (English-to-Japanese and Japanese-to-English) were completed and made publicly available.

1. Introduction

Since 1992, we have been developing a method of evaluating quality for the developers of machine translation (MT) systems to allow them to easily check imperfections in their own systems [1, 2, 3, 4, and 5]. In March 1994, we made our first version of the test-sets which evaluate the quality of English-to-Japanese Machine Translation systems (the "1993 Test Sets") available to the public. This test-sets is being used by developers, researchers and users of machine translation systems inside and outside Japan. We have continued our efforts to solve the existing problems revealed by evaluation experiments on some commercial MT systems and to increase the number of example sentences so as to cover more linguistic phenomena. Also, we have started to develop our test-sets for Japanese-to-English machine translation systems. In March 1995, our two test-sets (English-to-Japanese and Japanese-to-English) were completed and made publicly available.

In this paper, we would like to describe this systematic, objective method along with the test sets in which we have clarified the evaluation procedure by adding questions and explanations to the examples for the evaluation.

The work described in this paper has been developed by the Special Interest Group on Machine Translation (Chief: Hitoshi ISAHARA) in the Natural Language Processing System Research Committee (Chairman: Prof. Hozumi TANAKA, Tokyo Institute of Technology) which is a subcommittee of the Natural Language Processing Technology Committee (Chairman: Prof. Makoto NAGAO, Kyoto University) of JEIDA (Japan Electronic Industry Development Association). The members of the Special Interest Group in 1994 are shown in Table 1.

JEIDA has formulated three criteria for evaluating MT systems: 1) technical and 2) financial evaluations for users, and 3) technical evaluation for developers. For more information on these criteria, please refer to references 1 and 6.

Table 1. Members of the Special Interest Group

Hitoshi ISAHARA (Communications Research Laboratory, MPT)
Hajime UCHINO (Nippon Telegraph and Telephone)
Shiho OGINO (Ninon IBM)
Toshiyuki OKUNISHI (Sharp Corp.)
Satoshi KINOSHITA (Toshiba)
Shogo SHIBATA (CANON INC.)
Toshiyuki SUGIO (Oki Electric Industry Co., Ltd.)
Yasuhiro TAKAYAMA (Mitsubishi Electric Corp.)
Shin'ichi DOI (NEC Corp.)
Tadashi NAGANO (Matsushita Electric Industrial Co., Ltd.)
Masumi NARITA (Ricoh Co., Ltd.)
Hirosato NOMURA (Kyushu Institute of Technology)

We will first describe how our evaluation method surpasses previous methods, with reference to the following 2 types of objectivity.

(1) Objectivity in the evaluation process
(2) Objectivity in the judgment of the evaluation results

In an evaluation method such as the one proposed in the ALPAC report, "fidelity" and "intelligibility" are employed as evaluation measures, though they are dependent on human, subjective judgment. Consequently, the results may differ according to who has made the evaluations, that is, they do not satisfy the objectivity criterion (1). Theoretically, the evaluation method in the ALPAC report satisfies criterion (2) since the evaluation results are given as numbers. The system developers, however, fail to recognize which items cannot be handled in their own system. This is because the test example in question covers various kinds of grammatical items. So, their interpretation of the evaluation result for further improvement of their system must still be subjective. Therefore, for all practical purposes, this evaluation method does not satisfy criterion (2).

On the other hand, we have been preparing test-sets that can satisfy both criteria. We have clarified how to evaluate individual examples posing yes/no questions which enable the system developers to make an evaluation just by answering them. With our method, everyone can evaluate MT systems equally, for his/her answers require only a simple yes or no. Even for imperfect translation results, judgment will not vary widely among evaluators. In addition, we have assigned to each example an explanation which gives the relationship of the translation mechanism to the linguistic phenomenon, thus enabling the system developer to know why the linguistic phenomenon in question was not analyzed correctly. Consequently, with our test-set method, the evaluation results can be utilized for improving MT systems.

There is another proposed method where example evaluation sentences are collected. Each example sentence relates to a linguistic phenomenon subject to evaluation [7. 8. and 9]. With these test sentences, if a system is evaluated as incapable of properly translating an example, the system developer can immediately recognize that his/her system cannot handle the linguistic phenomenon in question. Therefore, we can conclude that this method satisfies the objectivity criterion (2). At present, however, this method has the following two problems:

(1) The procedure for evaluating the translation output has not been clarified.
(2) Identifying the deficiencies of the MT system through the evaluation results is dependent on the linguistic intuition of the evaluator.

As long as it is based on example sentences simply collected as a set of test sentences, this method can only be used for ad hoc evaluation, and cannot be established as an evaluation method. Moreover, to enable evaluation results to be used for improving MT systems, the listing of various linguistic phenomena is not enough; it is also necessary to clarify the positioning of each linguistic phenomenon within the grammar.

In our test-sets, we have systematically sampled the grammatical items that ought to be taken up, and listed some examples for each item. The test-sets clearly describe what linguistic phenomenon should be evaluated in each example so that the developers can easily understand the problems they need to solve in their systems. The system developer can then identify causes of translation failures.

In Chapter 2, we will describe our standpoint of the evaluation method, i.e., what information should be provided to system developers as a result of quality evaluation of MT systems. Chapter 3 describes how the test examples were collected. Chapters 4 and 5 describe the details of the test-sets for English-to-Japanese MT systems and Japanese-to-English MT systems, respectively.

2. Standpoint of the Evaluation Method

The method we propose here is a quality evaluation method which is totally independent of the MT system design. Therefore, the system developer can use this method regardless of his/her system type, e.g., whether the relevant MT system is rule-based or example-based. Conversely, in this method, if it becomes clear that a specific linguistic phenomenon cannot be processed by the relevant MT system, no solution common to the various system types is indicated, so the solution is entrusted to the developer according to the specific system type.

In our test-sets, we give no information on how often the linguistic phenomenon in each test-set appears in general usage. This is because the frequency of appearance of the relevant linguistic phenomenon may differ according to the type of document to be translated. If specific linguistic phenomena regularly appear in the documents handled on a specific MT system, the evaluator needs only to select the test-set which corresponds to the linguistic phenomena in question. Wrong evaluations could be made if scoring was based merely on the frequency of individual linguistic phenomenon.

To sum up, this evaluation method is designed in such a way that the system developers, irrespective of their system type, can precisely understand linguistic phenomena which cannot be handled by their systems and thus should be taken into account when improving the system performance.

3. Collection of Example Sentences for Evaluation

The test-sets employed in our evaluation method consist of example sentences for evaluation, their model translations (human translations), and the questions by which MT outputs should be evaluated. With the test-sets, the MT system developers can make objective judgments on the translation quality just by preparing system output and answering the question assigned to each example sentence. This chapter describes how the example sentences were collected for the test-sets.

The example sentences in the test-sets were collected by researchers and engineers who have actually dealt with the development of MT systems and/or natural language processing systems. During the collection of the examples, we emphasized the following two points:

(1) Coverage of basic linguistic phenomena
(2) Selection of examples with linguistic phenomena that are difficult to handle with MT systems, especially those with ambiguity problems

In other words, (1) refers to a systematic specification of the grammatical phenomena to be evaluated (top-down approach) and collecting examples according to these phenomena. On the other hand, (2) refers to a collection of examples that are difficult to translate on MT systems (bottom-up approach). In particular, we concentrated on those linguistic phenomena whose processing difficulties may be solved in the near future. Then, we systematized the examples for evaluation of MT systems. Furthermore, we repeated the translation evaluation tests on those examples using some commercial systems, and improved the test-sets focusing on the following points. All of them are important factors for maintaining objectivity during the evaluation process.

- No ambiguity in the questions
- No unnecessary complexity in the examples
- No ambiguity in the translation of the examples

4. Test-Sets for English-to-Japanese MT Systems

Our test-sets for English-to-Japanese MT systems consist of 770 test-sets, each with an English example sentence involving important grammatical phenomena of various kinds, a model Japanese translation of the sentence, a yes/no question to evaluate the system's translation, and so on. In the first two years, we selected mainly simple English sentences as test items and compiled 309 examples of basic sentences in our "1993 Test Sets." After that, we enriched the test-sets with the basic grammatical phenomena and extended them to complex and compound sentences. We also evaluated the test-sets with 8 different English-to-Japanese MT systems in order to examine their practicability, rewriting the questions in the test-sets if necessary. The revised test-sets was entirely completed by the end of March, 1995.

Each test-set consists of: an ID number, an example, a model translation, a yes/no question, translation sample(s) by MT systems, a sentence or sentences with related grammatical phenomena, ID number(s) of the reference item(s), and explanation (See Fig. 1). In this chapter, the Quality Evaluation Process, Object's Linguistic Phenomena, and the Simulation on MT systems are described.

4.1. Evaluation Process

Evaluation of the quality of English-to-Japanese MT systems is conducted as follows.

(1) Translate each [Example] in each test-set using your English-to-Japanese MT system.
(2) Answer "yes" or "no" (O or X) to the question on each example by referring to the translation result.
(3) Check the distribution of "yes's," and "no's" in the test-sets to evaluate the system performance.

We specified the judging points in the questions (e.g. which part of the example plays the grammatical role in question, and how that part should be translated), and we posed the questions in a yes/no style, to avoid varying judgments among the evaluators. Moreover, sample answers were also assigned to each test-set which were based on the translation results of several existing commercial MT systems. By referring to them, judgment can be easily made on each question.

With the yes/no distribution, the system developer can easily pinpoint the items which his/her system did not translate properly. In the test-sets, however, differences in significance and frequency among the examples are not taken into consideration. Therefore, it is meaningless to simply count the number of "yes" answers to compare the performance of various MT systems.

```
2．1．1　多品詞（品詞認定）　(= Part of Speech Disambiguation)
2．1．1．2　名詞/助動詞　(= Noun/Auxiliary verb)

【番号】　2.1.1.2-1　(= 【ID No.】)
【例文】　The trash can was thrown away.　(= 【Example】)
【訳文】　ごみカンは捨てられた。　(= 【Translation】)
【質問】　"can" が「カン/缶」のように名詞として訳されていますか？
　　(= 【Q.】Is "can" translated as a noun?)
【訳出例】○（くず缶/ごみ容器/くず入れ）は(廃棄された/[投げ]捨てられた)。
　　× ごみは捨てられ得る。
　　(= 【Translation Samples】literally meaning:
　　yes: The (garbage can/trash bin/litter bin) was (discarded/[thrown] dumped).
　　no: The trash can be discarded. )
【関連文】The last will was opened.「最後の遺言書は開けられた。」
　　(= 【Related Examples】and the Japanese translation)
【参照項目】2.1.1.2-2, 2.1.1.2-3　(= 【Reference Items】)
【解説】　"can was" の並びから、"can" が助動詞でないことがわかる。
　　(= 【Explanation】The word order of "can was" shows that "can" is not an auxiliary
　　verb.)

【番号】　2.1.1.2-2　(= 【ID No.】)
【例文】　The trash can be thrown away.　(= 【Example】)
【訳文】　ごみは捨てられ得る。　(= 【Translation】)
【質問】　"can" が「～できる/得る」のように助動詞として訳されていますか？
　　(= 【Q.】Is "can" translated as an auxiliary verb meaning "has ability to/has a
　　possibility to"?)
【訳出例】○（くず/ごみ）は(廃棄できる/[投げ]捨てられることができる)。
　　× ごみカンは捨てられた。
　　(= 【Translation Samples】literally meaning:
　　yes: The (garbage/trash) (can be discarded/[thrown] dumped).
　　no: The trash can was discarded. )
【関連文】(= 【Related Examples】)
【参照項目】2.1.1.2-1, 2.1.1.2-3　(= 【Reference Items】)
【解説】　2.1.1.2-1とは逆に、ここでは "can" は名詞ではなく助動詞。
　　(= 【Explanation】In contrast to No.2.1.1.2-1, here, "can" is not a noun but an
　　auxiliary verb.)
```

Fig. 1. Sample of the Test-Sets for English-to-Japanese MT systems

4.2. Linguistic Phenomena as Test Object

The test-sets consist of 770 English example sentences (see Table 2).

As shown in Table 2, the quality evaluation items were collected from the following perspectives: "Structural Analysis" and "Structural Selection."

In the "Analysis" part, MT systems are checked as to whether they can correctly analyze the sentence structure of the test example. This is a top-down approach in which the comprehensiveness of MT systems is checked. This part is intended to judge whether the MT system in question meets the requirements for an MT with good general performance. We selected grammatical phenomena essential for English by referring to grammar books (see [10, 11 and 12]), and classified them into 3 levels: (1) Part of Speech (2) Partial Structure of Sentence, (3) Sentence Structure. To cover all basic uses of verbs, adjectives and nouns, we adopted sentence patterns classified by Hornby (see [10]). However, some patterns were intentionally omitted because they were judged to be unnecessary for quality evaluation of MT systems. In addition, some usages of auxiliary verbs were omitted because they were considered to appear only rarely in the documents typically translated by MT systems.

Table 2. Distribution of the test items

1 Structural Analysis Part	—Subtotal:684 examples
1.1 Part of Speech	
1.1.1 Article:	15 examples
1.1.2 Noun, Proper Noun:	27 examples
1.1.3 Pronoun:	25 examples
1.1.4 Adjective:	42 examples
1.1.5 Adverb:	54 examples
1.1.6 Preposition:	40 examples
1.1.7 Verb, Auxiliary Verb:	85 examples
1.1.8 Relative:	25 examples
1.1.9 Conjunction:	26 examples
1.1.10 Symbol:	16 examples
—Subtotal:	355 examples
1.2 Partial Structure of Sentence	
1.2.1 Infinitive:	26 examples
1.2.2 Participle:	19 examples
1.2.3 Gerund:	23 examples
1.2.4 Tense, Aspect:	63 examples
1.2.5 Numerical Expression:	28 examples
1.2.6 Idiom:	8 examples
—Subtotal:	167 examples
1.3 Sentence Structure	
1.3.1 Sentence Type:	19 examples
1.3.2 Negation:	16 examples
1.3.3 Special Construction:	19 examples
1.3.4 Comparative:	21 examples
1.3.5 Subjunctive:	16 examples
1.3.6 Voice:	10 examples
1.3.7 Narration:	4 examples
1.3.8 Insertion:	16 examples
1.3.9 Ellipsis:	9 examples
1.3.10 Inversion:	7 examples
1.3.11 Parallel Expression:	25 examples
—Subtotal:	162 examples
2 Structural Selection Part	—Subtotal: 86 examples
2.1 Structural Disambiguation:	61 examples
2.2 Semantic Disambiguation (by co-occurring word(s)):	25 examples
-TOTAL:	770 examples

In the "Selection" part, on the other hand, MT systems are checked as to whether they can identify the correct structure syntactically and/or semantically when example sentences are ambiguous. This is a bottom-up approach in which the disambiguating ability of MT systems should be checked. Thus example sentences were classified into two groups: (1) Structural Disambiguation and (2) Semantic Disambiguation.

4.3. Test-Sets Simulation on MT Systems

In order to examine the practicality of the test-sets, we conducted a translation simulation on eight MT systems. The correct answer rates of the eight systems differed greatly: from 42 to 70 percent. Though these rates alone do not have any significance, they do indicate that the eight systems are quite different in performance both in the "Analysis" and in the "Selection" parts. That is to say, our

test-sets have successfully revealed that the range of linguistic phenomena which each MT system can handle is quite different. Therefore, the method that we have proposed here allows an efficient quality evaluation of MT systems.

5. Test-Sets for Japanese-to-English MT Systems

In order to evaluate the ability of Japanese-to-English MT systems, two kinds of proposals have been made so far. The first focused on differences in the way of perception between English-speakers and Japanese-speakers and used these differences as a base to classify Japanese expressions to be used as test examples [7 and 8]. On the other hand, the second focused solely on the structure of Japanese expressions and proposed example sentences for evaluation which represent the typical structural characteristics of Japanese expressions [9].

We began to construct our test-sets for Japanese-to-English MT systems in 1993. Like the set for English-to-Japanese MT systems, it is designed so as identify what is insufficient in systems by simply answering questions. However, we have constructed the test-sets for Japanese-to-English MT systems from a slightly different perspective than we have done for English-to-Japanese. Fig. 2 shows a sample test-set for Japanese-to-English MT systems.

```
JET140000 （1－4）複合述部
JEX140000 複合述部では、並列用言の認識を行ない、また用言部と格要素・副詞句とを
JEX140000 区別して翻訳しなければならない。
JEQ141000 複合述部の並列用言としての認定
JEX141000 複合述部の並列用言を認識するには、
JEX141000    ・助詞の種類により判断する
JEX141000    ・助詞の種類と名詞の意味属性により判断する
JEX141000    ・用言性の単語が並んでいれば、並列用言と認定する
JEX141000 等といった方法がある。
JEG141001 私達は研究開発する。
JEE141001 We do research and development.
JEE141001 We are carrying out research and development.
JEC141001 (失敗例) We study it ‖ develop it.
JEC141001 (失敗例) We develop a research.
JEC141001 「私達は研究開発する」の「研究開発」が「研究し開発する」という意味に
JEC141001 訳出されるかを確認する。
    (= JET140000 (1-4)Complex Predicates
       JEX140000 In order to translate a complex predicate, the grammatical relation
       JEX140000 between the components, i.e., the verb-verb combination,
       JEX140000 complement-verb combination or adjunct-verb combination,
       JEX140000 should be accurately identified.
       JEQ141000 Identification of the Verb-Verb Combination
       JEX141000 How to identify the Verb-Verb Combination
       JEX141000    ・judge by particle type
       JEX141000    ・judge by particle type and semantic attribute of noun
       JEX141000    ・recognize as parallel verbs,if verbal component is lined up
       JEG141001 私達は研究開発する。
       JEE141001 We do research and development.
       JEE141001 We are carrying out research and development.
       JEC141001 (miss) We study it ‖ develop it.
       JEC141001 (miss) We develop a research.
       JEC141001 The words 「研究」 and 「開発」 should be identified as parallel verbs. )
```

Fig. 2. Sample of the Test-Sets for Japanese-to-English MT Systems

In our approach, we have not only employed test-sets which enable an objective evaluation of MT systems but also established an evaluation method which enables the developers of Japanese processing systems to identify the correspondence between the linguistic phenomena and the processing modules.

That is to say, in addition to the example sentences and their evaluation procedure, explanations have been assigned to each test-set so that the evaluator can check how his/her system handles the linguistic phenomenon in question.

In this way, the system developers can evaluate the processing ability of their system as a whole and also recognize the performance of each processing module within their system.

In our test-sets, linguistic phenomena in Japanese were classified into 45 categories (See Table 3). An explanation has been given to each category enable a check of how the linguistic phenomenon in question is handled. If necessary, additional explanations are provided to clarify the problem itself and how to judge the output of the process. Each linguistic phenomenon is exemplified in test sentences and provided with a model translation in English as well as an explanation about the key factors in the translation. We select 330 questions and 400 technical sentences.

Table 3. Items in the Japanese-to-English test-sets

(1) Predicates
  (1-1) Correct Translation of Predicates
  (1-2) Copulas
  (1-3) Substantive Predicates
  (1-4) Complex Predicates
  (1-5) Parallel Verbs to translate One Word
  (1-6) Verbs translated as Adverbial Phrases
  (1-7) Subsidiary Verbs
  (1-8) Correct translation of Basic Verbs
(2) Nouns
  (2-1) Correct translation of Nouns
  (2-2) Noun Compounds
  (2-3) Translation of N1 'の(no)' N2
  (2-4) Translation of N1 'の(no)' N2 'の(no)' N3
  (2-5) Parallel Noun phrases
  (2-6) Interrogative Noun phrases
  (2-7) Verbal Nouns
  (2-8) Countability and Number
  (2-9) Proper Nouns
  (2-10) Indefinite Nouns
  (2-11) Nouns that show relationships
(3) Adverbs
  (3-1) Types of adverbs
  (3-2) Adverb phrases
  (3-3) Onomatopoeic words, Imitative words
(4) Noun Modifiers
  (4-1) Non-Inflecting
  (4-2) Inflecting
  (4-3) Particle phrases
  (4-4) Relative pronouns
(5) Particles
  (5-1) Correct translation of particles
  (5-2) Identification of Cases
(6) Affixes
(7) Tense, Aspect and Modality
  (7-1) Tense
  (7-2) Aspect
  (7-4) Modality
  (7-4) Voice
(8) Special structures
  (8-1) Idiomatic expressions
  (8-2) Idiomatic four character phrases
  (8-3) Co-occurring expressions
  (8-4) Weather, climatic expressions
  (8-5) Impersonal constructions
  (8-6) {は(wa) – が(ga)} constructions
  (8-7) Comparison
  (8-8) Metaphors
  (8-9) Partial Negation, Double Negation, Inversion
  (8-10) Honorifics
  (8-11) Quotation
  (8-12) Exemplification and Enumeration

So that developers can use the test-sets simply, we fixed the format of the test-sets and attached indexes to sentences to make them easily retrievable on machines. The index attached to each item has the following structure.

JET?????? Title
JEX?????? Explanation
JEQ?????? Question
JEX?????? Explanation of Question
JEG?????? Japanese Test Sentence
JEE?????? Model English Translation Example
JEC?????? Comment on translation (check points, examples of mistakes, …)

Numeral or letters are used for the part shown as '??????'. The first 2 characters show the chapter number of the title or subtitle. The next 3 characters show the question number. Questions consist of three stages at the most. The last character shows the sentence number of the test sentence and translation examples. Explanations and comments have the same sentence number as the item they explain.

Using these indexes as a key to retrieval, a user can easily extract only the necessary parts by using the retrieval commands of various OSs; For example, the texts for machine translation or the item list. Fig. 3 shows examples of using this test-sets. (MT_EVAL_JE.doc is the filename of the test-sets.)

6. Conclusion

In this paper, we have proposed systematic and objective methods for evaluating the quality of translation of a MT system from the developer's point of view.

Our method employs test-sets in which example sentences, their model translations, questions for evaluating the system output, similar examples (if any), and grammatical explanations have been systematically aligned. The example sentences have been collected focusing on wide coverage of both basic linguistic phenomena and linguistic phenomena problematic to MT systems.

The questions in the test-sets are designed to clarify the evaluation viewpoints. Given the system outputs for each example sentence in question, the system developer needs only to answer the question assigned to the example sentence. This judgment does not vary among evaluators, thus enabling an objective evaluation. Furthermore, with our test-sets, the system developer can precisely recognize which linguistic phenomena cannot be handled by his/her system.

Our two test-sets (English-to-Japanese and Japanese-to-English) are now available to the public without charge. In conclusion, we hope our evaluation method can play a useful role in the development of MT systems.

For any inquiries contact Hitoshi ISAHARA at: isahara@crl.go.jp.

References

[1] "Survey Report on Machine Translation Systems" (in Japanese), Japan Electronic Industry Development Association (JEIDA), 1993.

[2] "Survey Report on the Natural Language Processing Technology" (in Japanese), JEIDA, 1994.

[3] "Survey Report on the Natural Language Processing Technology" (in Japanese), JEIDA, 1995.

[4] H. Isahara, et al. : "JEIDA's Proposed Method for Evaluating Machine Translation (Translation Quality) — A Proposed Standard Method and Corpus ---" (in Japanese), IPSJ SIG Report, NL96-11, 1993.

[5] H. Isahara et al.: "Technical Evaluation of MT Systems from the Developer's Point of View: Exploiting Test-Sets for Quality Evaluation", First Conference of the Association for Machine Translation in the Americas (AMTA-94), 1994

[6] H. Nomura and H. Isahara : "JEIDA's Criteria on Machine Translation Evaluation", International Symposium on Natural Language Understanding and AI, 1992.

[7] S. Ikehara and S. Shirai : "Function Test System for Japanese to English Machine Translation" (in Japanese), IEICE SIG Report, NLC90-43, 1990.

[8] S. Ikehara et al. : "Criteria for Evaluating the Linguistic Quality of Japanese to English Machine Translations" (in Japanese), Journal of Japanese Society for Artificial Intelligence, Vol. 9, No. 4, 1994.
[9] H. Narita:"A Criteria of Processing Ability for Sentence Structure"(in Japanese), IPSJ SIG Report, NL69-1, 1988.
[10] A. S. Hornby : "Guide to Patterns and Usage in English, Second edition", Oxford Univ. Press, 1975.
[11] Y. Ogawa, et al. : "The Wonder Book of English Grammar" (in Japanese), Obun-sha, Tokyo, 1991.
[12] Y. Egawa: "Explanation on the English Grammar" (in Japanese), Kaneko Shobo, 1964.

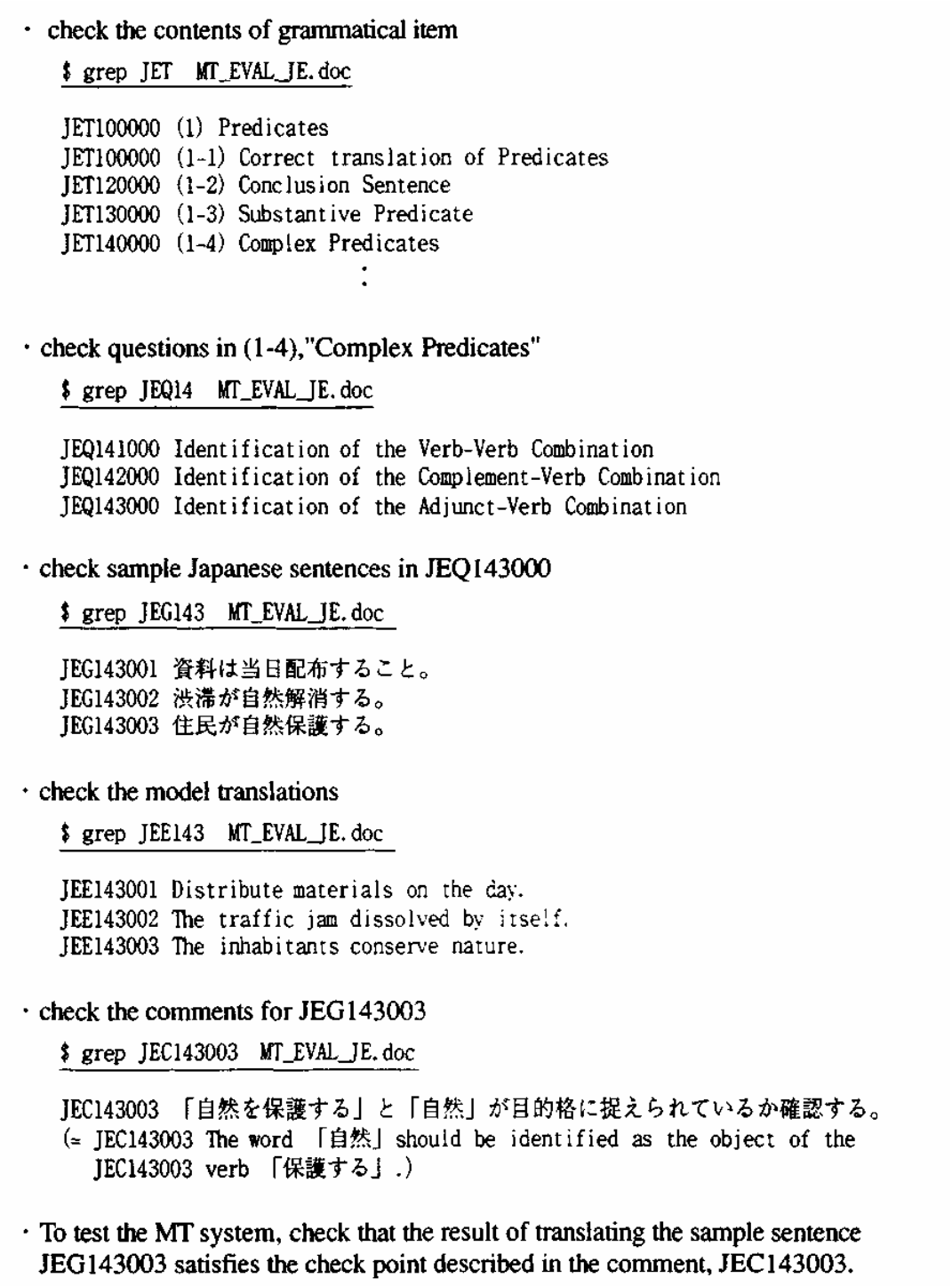

Fig. 3. Example of using this Test-sets