Machine Translation Technology:
Evaluation and Selected Sentences
on Which It Is Based

Shun ISHIZAKI and Hitoshi ISAHARA
Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba, JAPAN   Zip Code 305

1. INTRODUCTION

In Japan many corporations including computer related firms are engaged in research of machine translation systems. Recently there are already some developed and marketed.
It appears that machine will soon or later take over man in this field too.

In reality, however, few machines are nearly practical. Even plain accounts such as newspaper reports, computer can hardly handle them, not to mention artistic writings like novels. One may say that so far it is still man that gets the job done mostly. With pre-editing and post-editing human work, machine does only what it can door, you would say, helps man do the job.

To be a fully qualified translator, machine must acquaint itself with different types of sentences as taken from the following sources:
•School textbooks •Newspapers •Patent, industrial standards, and legal documents
•Academic papers and bibliographical descriptions •Manuals, catalogs, ads, and publicity brochures •Letters •Contracts and agreements •Interview articles and speeches •Novels, essays, and poems •Miscellaneous
Many genres of writings above include a vast amount of technical terms:
 information processing, medicine, mathematics, physics, politics, economy, law, etc.
This means a good machine translation requires appropriate dictionaries of technical terms on hand. In addition to the genre and field, machine translation must also be guided by the level of difficulty of sentence to translate.

2. COMPLEXITY IN TEXTBOOK SENTENCES—GRADES AND SUBJECTS

Those sentences discussed in this paper are from textbooks used at grade through high schools covering the different subjects of language, mathematics, science, and social study. They are also found at the varied levels of difficulty.

We first discuss the quantitative characteristics of sentences:
 1)the length of sentences; 2)the number of verbs, adjectives, and adjectival verbs;
 and 3)the numbers of modifying phrases and parallel structures.
These units are relatively easy to identify, thus facilitating analysis of sentences when
their complexity varies among the subjects of learning as the grade of the learner advances. The analyzed results are shown in Fig.l.

Second, we discuss the quality characteristics;
 1)homonyms; 2)morphological ambiguities; 3)sentence styles; 4)parallel structures;
 5)ellipses; and 6)anaphora.

At present it is not easy for the natural language processing technology to successfully deal with these quality elements: they are no doubt awkward factors in machine translation too.

We have chosen 100 sentences at random from each of the 26 textbooks for the quantitative analysis. Then we finally selected five sentences identified as those having

the average characteristics from each of different grade and different subject textbooks in the hopes that they will be used for data to evaluate machine translation systems. Some of these sentences are shown in Table 1.

3. EVALUATION

Efficiency of machine translation systems can be evaluated against criteria established in ALPAC reports or Japanese Mu projects. With one of these conventional methods applied, however, evaluation tends to be subjective instead of objective because of the evaluator, which is human, and is general rather than concrete because of the approach, which is overall. This paper examines every detail of major technological factors in machine translation systems. In addition, sentences on which evaluation is based are arranged to coordinate with the course of intellectual growth of school children.

We are to make full use of these sentences to evaluate machine translation systems. A series of our experimental machine translation efforts so far has indicated that our approach can be enough objective and effectively responsive to varied degrees of complexity of sentences.

Translation should start only after the meanings of the source language have been grasped. We may need to establish long-term objectives to achieve a truly good, practical translation system. Sentences collected here are then hoped to serve the purpose all the way.

REFERENCES
Ishizaki,S. and Isahara.H,"Extraction of Qualitative and Quantitative Characteristics of Complexity Included in Japanese Sentences," Information Processing Society, Report of Natural Language Processing Group 67-6(1988.7).
Ishizaki.S. and Uchida, H, "On Interlingua for Multilingual Machine Translation," Information Processing Society, Report of Natural Language Processing Group 70-3(1989.1).

Table 1   Selected Sentences from Textbooks
(The English are human translations from the Japanese)

|	
Science, 2nd Grade Compare a place in sunshine and one in shadow. Is this why the sun moved? Science, 6th Grade Light a big candle to see how it burns. Only a single dry cell is used for a device like the photographed one. Science, Jr. High School Both are colorless liquid and look equally transparent regardless of the volume. See if your measured volume agrees to your calculated figure. Science. High School However, today, we know that "a running car would not stop promptly" if the engine is turned off. Then,when attempt to consolidate force,we must consider not only its size but its direction as well.	新訂小学理科2 日なたと日かげをくらべよう。 たいようがうごいたからかな。 新訂小学理科6(下) 大きなろうそくに火をつけて、燃えている 　様子を調べてみよう。 写真のようなしかけには、かん電池が一個 　だけ使われている。 改訂理科1分野上 どちらも無色の透明な液体であるが、水の 　分量が少ないのにつりあっている。 こうしてはかった体積が、計算で求めた結 　果と一致するかどうか確かめてみよう。 高等学校物理改訂版 しかし、現代の私たちは、走行中の自動車 　のエンジンを切っても「クルマは急には 　止まらない」ことを知っている。 それで、力を合成するときには、大きさだ 　けでなく、力の方向・向きも考えなけれ 　ばならない。

Fig. 1   Quantitative Analysis of Elementary and Secondary Schools Textbooks

Language

Mathematics

Science

Social Study
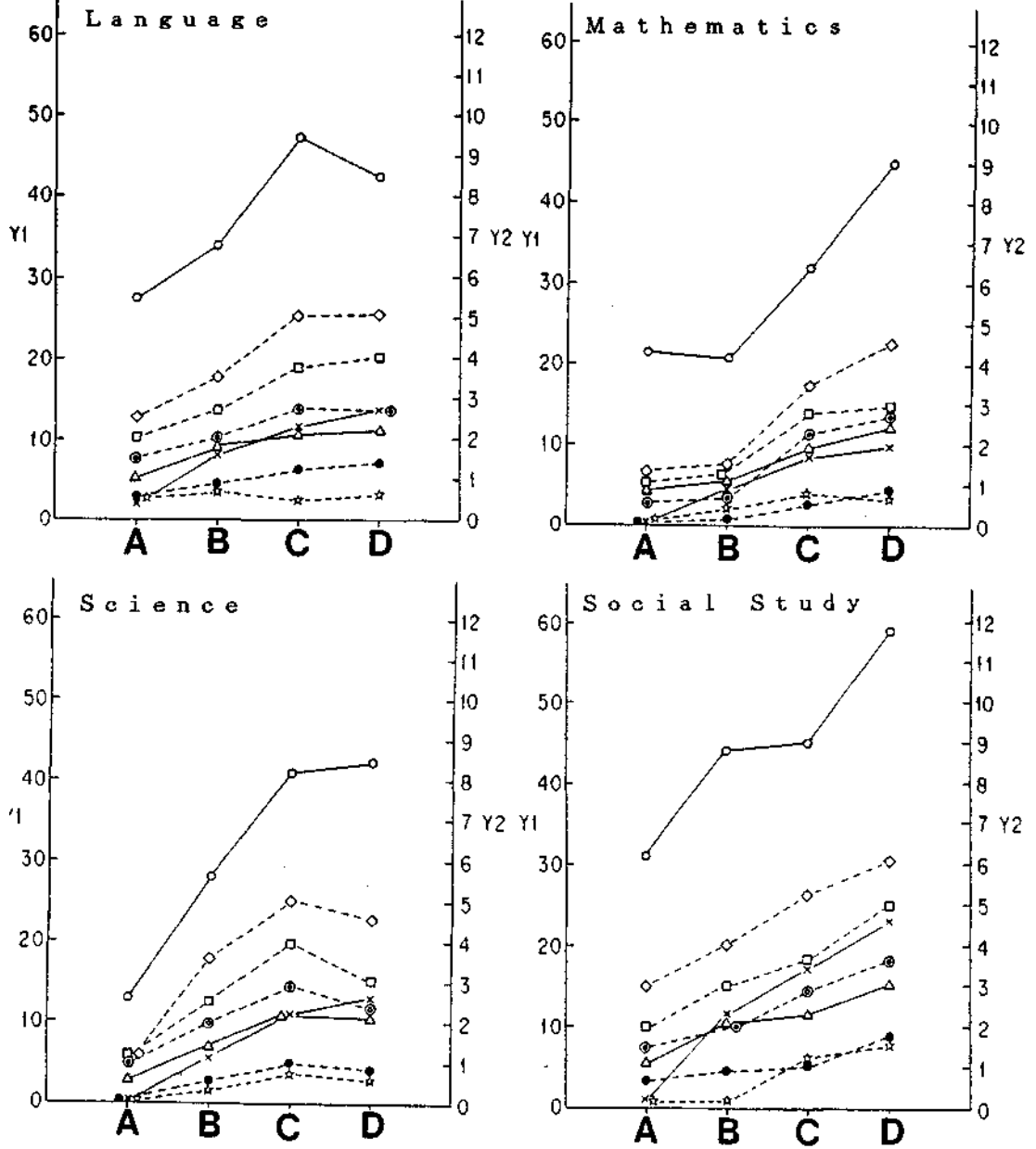
Y 1   O : No. Characters
      X : No. Kanji
      △ : No. Phrases
Y 2   □ : No. Yogen
      ◎ : No. Verbs
      ● : No. Embedded clauses and phrases
      ◇ : No. Verbal kakari
      ☆ : No. Parallel structures

A : Elementary school lower grade
B : Elementary school upper grade
C : Jr. high school
D : High school