

# A Report on the Machine Translation Market in Japan

**Setsuo Yamada**

NTT Corporation

[yamada.setsuo@lab.ntt.co.jp](mailto:yamada.setsuo@lab.ntt.co.jp)

**Syuuji Kodama**

Fujitsu Limited

[xiaoyu@jp.fujitsu.com](mailto:xiaoyu@jp.fujitsu.com)

**Taeko Matsuoka**

Toshiba Solutions Corporation

[Matsuoka.Taeko@toshiba-sol.co.jp](mailto:Matsuoka.Taeko@toshiba-sol.co.jp)

**Hiroshi Araki**

Inter Group Corporation

[haraki@intergroup.co.jp](mailto:haraki@intergroup.co.jp)

**Yoshiaki Murakami**

Navix Co., Ltd.

[murakmy@navix.jp](mailto:murakmy@navix.jp)

**Osamu Takano**

Sonderhoff & Einsel Law and Patent Office

[trademark@sonderhoff-einsel.co.jp](mailto:trademark@sonderhoff-einsel.co.jp)

**Yoshiyuki Sakamoto**

Tsukuba Women's University

[ysaka@cs.kasei.ac.jp](mailto:ysaka@cs.kasei.ac.jp)

## Members of the Market Research Committee

Asia-Pacific Association for Machine Translation (AAMT)

### Abstract

When conducting market research on machine translation, we research the volume of sales continuously in order to determine the scale of the machine translation market in Japan. We have officially announced these figures every year. Furthermore, since 2003, we administered questionnaires regarding the Web translation

service and MT package software to general users through AAMT homepage.

In this paper we will report on the two kinds of execution conditions outlined above and present the results, and introduce to the case study of translation workflow using MT and TM by the Japanese translator<sup>1)</sup>.

### Keywords

Machine translation, Market research, Questionnaire for users and vendors, Workflow

### 1 Introduction

With the spread of the Internet, machine translation has become an essential item for the global network. The arrival of a networked society has resulted in many documents being written in a language other than the writer's native language. The time to utilize machine translation for the receipt and sending of information is now. The "Market Research Committee" was started in 1996; this is its 10th year in our society. Our committee is mainly composed of manufacturers who deal with machine translation systems, users, and neutral organizations. Their activities focus on researching

the changes in the trends of machine translation systems and related issues. Our committee works towards the development and spread of machine translation systems. We plan to create an allied industry. Computer usage and computer environments have changed greatly in Japan.

Machine translation systems and translation software have been vastly improved. Following the spread of the Internet, significant changes in attitude were seen in regard to machine translation and evaluation. Additionally, the use of machine translation systems and evaluation issues has also changed.

## 2 Survey of MT Vendors

We investigated the shipments and sales of MT systems, the source and target languages that MT systems process, and the platforms of MT systems by collecting questionnaires from vendors. In order to keep the raw data confidential, we asked the office of AAMT to collect and count the data. The results were discussed at the joint meeting of the Market Research Committee. (Shipping trends were identified, and the next investigation was conducted in order to create new interest in the foundation of MT business in order to estimate the scale of the machine translation market in Japan.)

### 2.1 Methodology

We sent the questionnaires to MT vendors (included manufacturers) on March 15, 2004 and had them returned by the end of March, 2004.

#### Scope

The target systems of our investigation were limited to MT systems sold in Japan. Programs that were pre-installed in computers and OEM supplies were also included.

#### Language Combinations

Translation in Japan is usually between English and Japanese. We set the four patterns of language combinations in the questionnaire, as follows:

- English to Japanese

- Japanese to English
- Bi-directional (E to J and J to E)
- Other languages

**NOTE:** The last category includes translation from Japanese to Chinese, from Korean to Japanese, etc.

#### Software Packages / Pre-installed Programs

We analyzed the shipments and sales of MT software packages and pre-installed MT programs separately. The software packages are sold independently of hardware whereas the pre-installed programs are part of computer systems.

#### Shipments and Sales

Vendors provided information about the actual shipments and sales in 2001 and 2002, provisional shipments and sales in 2003, and forecasts for 2004 (in this paper all years are Japanese fiscal years: April 1 to March 31).

### 2.2 Results

#### Return Rate

The return rate was 44%. We sent the questionnaires sheets to 27 companies and received completed questionnaires from 12 of them.

### 2.3 Analysis on Shipments

Shipments of one-system (E to J) units decreased slightly while J to E units have drastically decreased. Two-way systems had even more total

Language Combinations		Numbers of Shipments			
		2001	2002	2003 (Provisional)	2004 (Forecast)
Package	E to J	7,761	6,600	6,001	3,901
	J to E	3,774	1,385	813	550
	E to J, J to E	138,731	337,131	401,014	139,240
	Others	31,132	23,121	25,137	30,180
Pre-installed	E to J	1,810,760	1,433,620	1,200,000	1,800,000
	J to E	-	-	-	-
	E to J, J to E	1,300,000	1,218,000	1,080,000	1,102,000
Total shipments		3,292,158	3,019,857	2,712,965	3,075,871

**Table 1: Shipments Each Year**

shipments. Two-way translation systems comprised 93% of all shipments. Total shipments increased steadily year by year, and were 1.6 million units in 2003, but a tendency towards a decrease is apparent. We believe that one of the reasons this decrease occurred is because the total shipments of personal computers in Japan have also decreased.

## 2.4 Analysis of Sales

Sales of two-way systems have been growing rapidly. The sales of the J to E translation units have decreased by 50% every year. In 2003, two-way system units accounted for 94% of the whole, and sales were growing, too.

Language Combinations	Sales (¥1,000)			
	2001	2002	2003 (Provisional)	2004 (Forecast)
E to J	32,381	26,390	26,325	21,260
J to E	178,768	75,548	39,503	27,227
E to J, J to E	420,875	888,201	1,100,121	1,178,819
Other	21,900	10,000	10,000	14,560
Total sales	653,924	1,000,139	1,175,949	1,241,866

**Table 2: Sales Each Year**

## 3 Investigation for MT General Users

### 3.1 Purpose

We have to consider how to investigate the usage of and the attitude towards MT software for MT general users. Moreover, we must also consider what kind of users should be selected for the investigation. The method of investigation depends on whether the target individuals use a personal computer or the Internet. The method also depends on whether the target individuals have the experience or the desire to use MT software.

In August 1999 we conducted an online questionnaire and issued prizes. Approximately 5,000 people completed the questionnaires.

Because the Internet has become such a popular new communication media, we encounter foreign languages more frequently. Considering this global situation, we decided to develop questionnaires that inquired about how people use MT software on the Internet and to explore what problems are encountered when using MT software on the Internet.

We issued questionnaires to Internet users via the AAMT homepage at the same period (12 days) and the same time (from the end week of January to the first week of February) in 2003, 2004 and 2005. Prizes (MT packages) were offered as an incentive all times. Analyzing the three results are shown in the next section.

### 3.2 Results and Analysis

The number of completed questionnaires each year is as follows:

1,413 in 2003      1,720 in 2004      1,018 in 2005

#### 3.2.1 Result of Every Questionnaire from 2003 to 2005

Overall, the frequency of use of translation services offered on the web translation services has slightly but steadily increased to 53.05 % (Table 7 and 8). This shows the convenience of web translation services has increasingly been recognized. The percentage of the people who do not know the availability of web translation services “do not know of MT” in Table 14 or who “do not know how to use MT,” which are major reasons for

not using such a service has slightly decreased (Table 14). If web translation services are advertised more in the future, they will gain greater recognition and the number of users will increase accordingly.

The percentage of female users (Table 4) and that of advanced-age users in the 50s or older (Table 3) have risen to 16.21 % and 31.04 %, respectively.

Other trends include;

- An increased number of professional translators (5 %) are using the services. (Table 5)
- The percentage of people who use the services just to get some idea “find a clue” or “know an outline” has reached 44%, clearly showing the

main use of machine translation. In addition, people who use the services to make a draft translation “get a draft for translation” are increasing, and the percentage of such use has exceeded 10 %. (Table 13)

- Users who score 500 points or more on the TOEIC test are increasing and their percentage has reached 24 %. (Table 6)

From the above-mentioned trend, we can conclude that people with a certain degree of English ability, including translators, are well using web translation services.

	10s	20s	30s	40s	50s	60s	70s	80s or over	No data
2003	3.32	23.21	36.37	23.99	8.84	1.41	0.92	0.14	1.76
2004	5.81	28.02	30.93	22.50	9.65	1.51	0.23	0.06	1.29
2005	4.52	22.40	31.43	23.18	11.20	3.44	1.57	0	2.26

**Table 3: Age (%)**

	Male	Female	No data
2003	74.38	25.05	0.56
2004	70.64	29.13	0.23
2005	68.17	31.04	0.79

**Table 4: Sex (%)**

	Student	Company employee	Translator	Homemaker	Self-employed	Teacher	Other	No data
2003	9.27	54.77	1.41	9.97	7.99	2.19	12.80	1.55
2004	15.47	50.29	0.64	9.83	7.67	1.40	13.90	0.80
2005	11.20	48.92	5.40	8.74	8.74	2.36	12.97	1.67

**Table 5: Occupation (%)**

TOEIC	-199	200-499	500-749	750+	No data
2003	57.82	24.62	11.53	5.09	0.92
2004	55.64	28.49	11.86	3.14	0.87
2005	47.25	27.60	13.85	9.92	1.38

**Table 6: English skill (%)**

	Not know	Never used	Have used	No data
2003	24.20	30.00	44.16	1.62
2004	23.02	25.87	49.65	1.46
2005	20.14	25.34	53.05	1.47

**Table 7: Experience of any web translation services (%)**

	Every day	A week	A month	A year	No data
2003	1.91	11.18	23.49	14.01	49.39
2004	3.60	15.81	22.91	14.30	43.38
2005	4.22	15.23	24.07	16.90	39.58

**Table 8: Frequency of web translation services (%)**

**NOTE:** All questions in 2003 under Table 9: Purpose can be answered more than once. Every rate is shown by the following: the number of answer divided by total number of answer.

	Personal	Work	No data	Total answers
2003	554	240		794
(%)	69.77	30.23		
2004(%)	41.80	15.70	42.50	
2005(%)	38.80	20.73	40.47	

**Table 9: Purpose**

	HP on the Internet	Text	E-mail	No-data	Total answers
2003	435	450	92		977
(%)	44.52	46.06	9.42		
2004(%)	28.37	18.43	7.79	45.41	
2005(%)	27.50	24.36	7.96	40.18	

**Table 10: Translation type**

**NOTE:** Because questions on translation articles were changed in 2004, Table 11 and Table 12 are shown.

	Government	Company	Computer	News	Sport	Sightseeing	Shopping	Hobby	Total answers
2003	56	162	252	179	61	68	51	342	1,171
(%)	4.78	13.83	21.52	15.29	5.21	5.81	4.36	29.20	

**Table 11: Translation articles in 2003**

	Social science	Natural science	Industry	Culture & Art	Sport	Sightseeing	Daily information	No data
2004	5.29	8.20	13.37	8.14	2.97	4.24	10.00	47.79
2005	6.09	8.15	17.88	7.47	4.72	5.01	10.71	39.97

**Table 12: Main translation article in 2004 and 2005 (%)**

**NOTE:** All questions in all years under Table 13 and Table 14 can be answered more than once. Every rate in the tables is shown by the following: the number of answer divided by total number of answer.

	To find a clue	To know an outline	To get a draft for translation	Don't have MT software	Easier usage	No charge	Total answers
2003	310	451	121	313	106	484	1,785
(%)	17.37	25.27	6.78	17.54	5.94	27.10	
2004	420	583	168	443	145	585	2,344
(%)	17.92	24.87	7.17	18.90	6.19	24.95	
2005	201	357	130	168	75	353	1,284
(%)	15.66	27.80	10.12	13.09	5.84	27.49	

**Table 13: The reason why people use web translation services**

	Do not need	Do not have MT information	Do not know of MT	Do not know how to use MT	Poor quality	Use MT software package	Can ask someone to translate	Can translate	Total answers
2003	109	105	262	311	143	71	29	58	1,088
(%)	10.02	9.65	24.08	28.58	13.14	6.53	2.67	5.33	
2004	124	166	297	335	150	65	35	56	1,228
(%)	10.10	13.52	24.19	27.28	12.21	5.29	2.85	4.56	
2005	59	74	143	146	116	26	15	48	627
(%)	9.41	11.80	22.81	23.29	18.50	4.15	2.39	7.65	

**Table 14: The reason why people do not use web translation services**

### 3.3 About the investigation into the packaged MT software

In addition to the questionnaires about web translation services, this time we issued some other questionnaires and administered an investigation into the actual usage and the attitude towards packaged MT software too. We plan to make the same investigation again next year and then with comparing both year's results, we will present them.

## 4 SATILA

Integrated Translation Workflow Using Translation Memory and Machine Translation

### 4.1 Introduction

SATILA stands for Software Assisted Translation Workflow Involving Linguistic Analysis.

SATILA is neither a tool nor something that depends on a specific piece of software or environment. In SATILA, active term unification and its automatic application can be accomplished by the combination of translation memory and machine translation software.

### 4.2 Translation Workflow

Five years have passed and still the basic functions of translation software are not used by translators and only a handful of translators have been able to effectively use translation software. The reason why translation software has not been used does not lie with the software itself. It is because how it can best be used is not understood and appropriate workflows do not yet exist, rather than any type of insufficiency in the software.

Translation software's strongpoint is most easily realized in large localization projects and other large translation projects. As opposed to human translation, the larger the project the more efficiency increases. The SATILA workflow assists translators to accomplish high-quality translations. In traditional human translation there is much wasted effort and mistakes are numerous. The translator's concentration is not an automatic function and can begin to fade in the presence of monotonous substitution work.

When translation memory is used, efficiency goes up but involvement by human translators will also be necessary.

### **4.3 Software Dominance, User Dominance**

Communication between translation software vendors and translators was insufficient in that it did not adequately convey the needs of the translators.

It is necessary to differentiate fully-automatic translation from translation support. There are numerous cases where neither the manufacturers nor the users understand this. It is not yet understood what is necessary for translation software to support translation as a commodity.

Fully-automatic translation is not a problem for the ordinary user or the in-house translation work user. But, professional translators need a translation support system that has quality control functions. For instance, an English to Japanese translator should have skill in both Japanese and English, translation capability and minimum levels of skill.

Natural and correct Japanese and a certain level of English language skill, such as a score of 850 points in TOEIC, is necessary to be able to avoid syntax interpretation errors. Moreover, reliable translation skills that are equal to the translation software processing speed and personal computer skills that make full use of translation memory, translation software and other available tools are

necessary. The automation of post processing, impossible for current translation software alone, becomes especially important.

The ordinary user is satisfied with only the understanding of the translation software as a 'black box.' If he doesn't even want to be bothered to press a button to tell the software where to stop a certain Web translation, then the software has a dominate role in the translation work and we call that 'software dominance.'

On the other hand, if the translator/user understands and manages the selection of what words are translated how, then we call this 'user dominance.' The role of translation software in this instance is translation support throughout the project.

The translator/user completes a high-quality translation using translation software in those places where it is adequate and for those places where it is not he will deal with it himself. In this case, the translation software is a strict and accurate term-translation tool for the translator.

### **4.4 Llewelyn Reaction**

Llewelyn reaction is the key to how we can get the user to use translation software more frequently. The Llewelyn reaction is an emotional rejection on the part of the human translator to the software's recognition failure, especially to the failure of the recognition of natural language processing. Distrust of and dissatisfaction with the software are put ahead of rational judgment so that use of the software is rejected. Even when the software performs translation, the user feels that the software has failed to perform and that reinforces the distrust and dissatisfaction.

In many cases, the Llewelyn reaction consists of four stages. The first stage: Deviating from common sense in the natural language resulting in laughter, the *Laughter Stage*.

The second stage: *Anger Stage*—the user experiences stress. The possibility that the user will abandon the software rises if the Llewelyn reaction cannot be overcome. These users form the largest group that hinders the acceptance of translation and voice recognition software. The number of users that distrust software increases considerably here.

It then shifts to the third stage: *Acceptance Stage*—when a translator/user gets past the second stage. The user adjusts to and accepts the mis-processing of the natural language processing software, and is able to make an effort avoid the errors.

Additionally, the translator/user is likely to advance to stage four: *Action Stage*—the stage for a so-called power user. This stage corresponds with the acknowledgment of the mis-processing patterns of software, and voluntarily performing corrective programming.

The Llewelyn reaction is one of the causes of the ‘digital divide,’ and the spread of advanced IT, including natural language processing, is hindered. Even if a correction is easily accomplished, the user refuses emotionally without benefit of a calm analysis of the error and correction.

*Llewelyn* of the Llewelyn reaction was a character in a child’s tale who killed his loyal dog because he thought the dog had failed to protect his child, but in fact the dog had killed a wolf to protect the child.

#### **4.5 Suitable Translation**

In normal system dictionaries there are a lot of expressions that the translator finds lacking. A complete mechanism (a systematic collection of excellent translations and translator feedback) doesn’t yet exist. Every day a large amount of excellent translated expressions are buried by the present system. If phrase dictionaries are improved in the future, translation software will be more

favorably received by translators and the general public.

The translation memory function of translation software is not up to par. Because the translation software does not protect format tags in the text, the tags must be removed from the text prior to translation.

Suitable translations are always required for actual work, but present software cannot always provide such suitable translations. If the area of knowledge of the field is limited and a human translator can check the translation, suitable translation settings can be applied and automatic translation becomes possible.

If the processing of the suitable translation assumes a difficult case in the translation engine, and the standard for the suitable translation can be given to the translation dictionary as a comment, the translation dictionary becomes a useful tool.

Great improvements in the system can be expected in automatic translation, to say nothing of translation support.

#### **4.6 Conclusion**

Maximum effectiveness and quality can be achieved by the effective combination and use of translation software and translation memory as presently marketed. When translation memory is combined with translation software, quality is improved as the merits of both raise the maximum effectiveness of each. Moreover, if recommended terms are decided, the labor involved in term selection and unification is greatly reduced.

The larger the scale, the more SATILA demonstrates its effect and the improvements in quality and efficiency will be remarkable.

#### **Reference**

- 1) Yuji Yamamoto. 2005. “SATILA - Integrated Translation Workflow Using Translation Memory and Machine Translation” (in Japanese). AAMT Journal No.37, 2005.