ISSN 0918-2802

Technical Report

MT R&D IN ASIA [1]

Hozumi TANAKA
and Virach SORNLERTLAMVANICH

TR98-0001 January

Ôokayama 2-12-1 Meguro Tokyo 152, Japan
http://www.cs.titech.ac.jp/

[1]A revised version of the paper submitted to MT Summit VI

Abstract

There has been a big shift in MT R&D in the Asian region after many large-scale projects being conducted over the past ten years. The Multi-lingual Machine Translation (MMT) project is a significant R&D project that has increased the number of NLP-related researchers and the quantity of research activities undertaken by research institutes in recent years. This project has provided considerable insight through collaborative research across languages, and is hoped to catalyze future MT R&D in this region. Though MT systems are still far from the ultimate goal of perfect translation, at present many MT systems are being applied to the support of information retrieval from the Internet.

1 Introduction

Machine translation R&D activities were vitalized by the Mu-project [7], a national Japanese MT project. The Multi-lingual Machine Translation (MMT) project was a multi-national R&D project that made a great contribution to the Asian natural language processing community, and led to the establishment of many MT R&D institutes in this region. Both the public and private sectors realized the necessity of developing MT and its possible spin-offs.

Though there is a bottleneck in raising the translation accuracy of MT, users can be satisfied by the speed of translation and user-friendliness which can often support the retrieval of information over the Internet. In the Japanese market, users also use MT to extract an outline English source texts for quick reference, despite the fully functional, high accuracy translation being a distant goal. With current MT technology, advantages of hardware and networked environments, MT is an essential tool for surfing the Internet.

Based on experiences with the MMT project, many countries in the Asian region started developing more practical MT systems, bilingual MT between English and local languages, especially partly as a result of the realization that there is a real demand for in the existing Internet environment.

We give a brief introduction of the MMT project and discuss some problems of multi-lingual system development in the next section. Section 3

presents the current face of MT in Asia and state of R&D in some individual countries worthy of note. Section 4 gives information about NLP research activities and some available NLP resources.

2 Multi-lingual Machine Translation Project (1987-1994)

The Japanese government launched the Multi-lingual Machine Translation (MMT) project in the late 1980s, in collaboration with China, Indonesia, Malaysia and Thailand. The MMT project was a ODA project focusing on R&D conducted by the Japanese Minstry of International Trade and Industry (MITI). The past Japanese success of commercializing machine translation systems seemed to be a strong driving force of the project, which was initiated by cooperation between seven leading private companies and the Center of the International Cooperation for Computerization (CICC) of Japan, organized as given in Figure 1 [5].

The system introduced an interlingua as the intermediate representation between Chinese, Indonesian, Malay, Thai and Japanese. For each language, an analysis system to derive an interlingua representation for a given input sentence, and a generation system for generating a sentence in that target language from an interlingua representation, were developed, (see Figure 2). Description of the interlingua presented itself as the main challenge of the system, in terms of implementation.

From the very different technological starting points of each country, the project shared the common theme of developing an interlingua. The project was aimed at developing a language-independent knowledge representation scheme, compatible with at least the five languages mentioned above. The interlingua [4] is composed of two parts, namely, the concept classification component and the concept relation component. Within a language, it seems possible to define a set of concepts descriptively or by sets of synonyms, such as the synsets used in Wordnet. Problems occur when concepts defined by different languages must be linked. The project used the set of concepts contained in the EDR concept dictionary as the initial concept set to link the languages. Only about 25% of concepts can be linked for any two languages, a figure which drops to about 15% for multiple languages. One of the typical

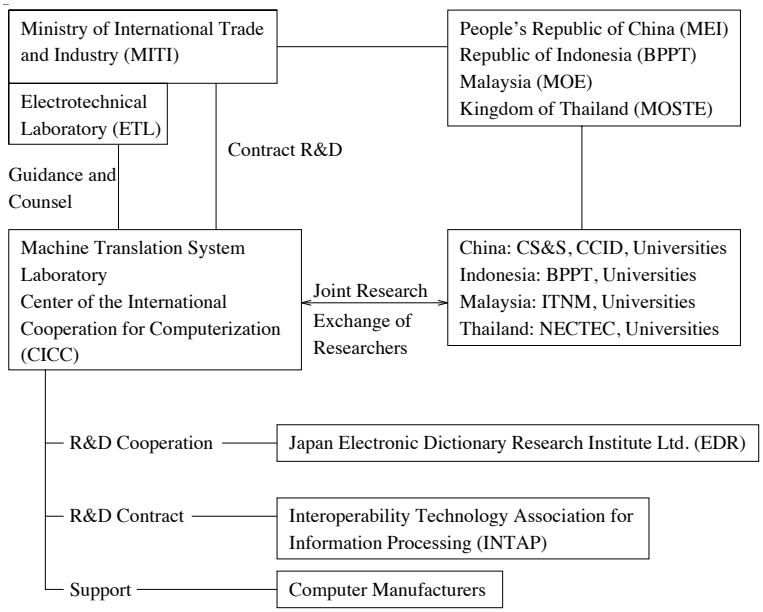

Figure 1: Collaboration in the MMT project

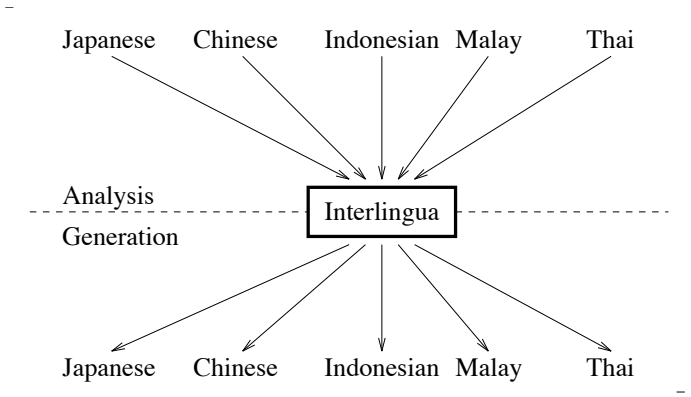

Figure 2: Interlingua-basis of the MMT system

problems of concept linking across languages is the inequality of concepts, which brings about further adjustment of the concept linking procedure.

Concerning the set of concept relations, the project proposed a set of 24 concept relation types as the final version of the interlingua [3]. This was then used in the development of the MMT system. Mismatches between analysis and generation systems occurred when the definitions for concept relations were unstable. The project tried to re-adjust the set of concept relations and finally proposed a new set of 17 concept relation types, combined in a hierarchy. Together with the definitions, a set of case-frames to constrain the concept relations were also stipulated. Using this approach, concept relations could be fixed by each verb case-frame. However the cost of extracting prototypical verb case-frames is very high. In a feasibility study, we prepared about 100 basic verbs in each language. See [4] for the details.

The project ended in early 1995. The interlingua approach looked to be the appropriate solution for developing a multi-lingual machine translation system. Idealistically, any target languages can share resources simply by being linked to the interlingua. But in practice, in addition to the problems faced in designing an interlingua, as mentioned above, developing a multi-lingual machine translation system has its own particular problems:

- **Resource availability**:- This includes human and language resources. Human resources, in terms of system developers, are limited. Grammar developers for the MMT system had to be fluent in the target language to be able to adjust the grammar to suit the interlingua. One reason for this is that no reliable inference engine which supports the interlingua exists currently. Turning to language resources, large quantities of corpora are required for language study. The process of building a collection of language corpora and large-scale dictionary have just begun. This is a labor intensive task because most Asian languages still require manual data input with a keyboard.

- **System evaluation**:- It is hard to evaluate the system performance of a multi-lingual system. The evaluator has to be fluent in all the treated languages. Most fine-detailed problems cannot be solved because they are unrecognizable to the evaluator.

- **Linguistic issues**:- There are various language types using in the Asian region. Each has its own particular problems in language process-

ing. Some languages need a process for word segmentation at the morphological analysis stage, such as Chinese, Japanese and Thai. Some have no word inflection, such as Chinese and Thai. Therefore, we need to have separate language-specific modules from which the parser can be shared.

3 State-of-the-Art MT R&D in Asia

There is no doubt that a technology shift occurred in Asian countries as a refult of the MMT project. Especially for countries participating in the MMT project, many MT-related research activities have been initiated and significant government support has been forthcoming. Most of the countries started with limited numbers of MT-related researchers and language resources. Constructing machine-readable dictionaries was one of the most labor intensive tasks. Though there were a lot of classical printed dictionaries available, the number of word entries was very limited, and most of the word information was omitted and left to the reader's intuition. The project succeeded not only in realizing a multi-lingual machine translation system but also in building the infrastructure for MT R&D in this region. The project derectly resulted in a basic term dictionary of about 50,000 word entries and information science-related technical term dictionary of about 25,000 word entries in electronic form for each language. In addition, corpora for grammar acquisition and evaluation were also gradually built up.

Thanks to the widespread availability of the Internet of late, the expectation placed on MT technology has shifted to greater application to networked environments and larger amounts of information. The potential for English language tasks has burgeoned through the widespread use of computer networks, which has had a great effect on MT R&D in Asia. The following issues are raised in connection with remarkable paradigm shifts in MT R&D, especially in Asia:

- **Knowledge representation to knowledge acquisition**:- Study of interlingua as a universal representation in knowledge-based MT has been a major topic over the past ten years. The availability of information in an electronic form has produced interest in the field of knowledge acquisition. In addition to research on knowledge represen-

tation, much recent research has been conducted within the framework of statistics or probability, as well as machine learning.

- **Multi-lingual MT to Bilingual MT**:- As a result of the momentum built up in developing MMT system, many countries have looked to a more practical approach to meet the present needs of MT. With the growing demand for English, especially through the Internet, priorities in MT development has been altered to focus on translation systems between English and local languages.

- **Total to partial use of MT technology**:- Instead of applying NLP technologies to MT as a whole, some partial technologies have been applied directly to language specific problems, such as word processing, word segmentation, spell checking and grammar checking.

- **Workstation to personal computer**:- Rapid leaps in hardware performance have offered another choice of computer platform, with a cost performance in line with personal use. Many commercial systems are now being developed for the Windows system, at the personal computer level in addition to workstations.

- **Translation accuracy vs cost, speed and user interface**:- High translation accuracy is no longer the only goal of MT development. Current technology barely assures translation accuracy. On the other hand, it seems to be difficult to make a significant improvements on current technology. Responding to present needs, systems must be cost effective, able to translation in real time and have a user-friendly interface in addition to being accurate.

3.1 Japan

As per report [1] on commercial MT products at the end of 1996, 20 companies had released 59 products less than 1,000,000 yen (about 8,500 US dollars) in price, as listed in Table 1.

The lowest price is 6,000 yen (about 51 US dollars), unconceivable only ten years ago when prices were around 7,000,000 yen (about 60,000 US dollars). According to report [1], the demand for MT in Japan is as high as that for word processors. But, because of the surprisingly high price and the

Table 1: MT product list as surveyed by AAMT in 1997

No.	Product Name	Language	OS	Developer
1	Yakushi Nyorai	E → J	Windows	CSK
		E → J	Mac 7.1	
2	BRAVIS J/E	J → E	Windows	Pacific I&MT Lab
	BRAVIS J/E Sasuga	J → E	Windows	
3	TransLand	J → E	Windows	Brother
		J → E	Mac 7.1	
4	Transporter EJ/C	E → J	Windows	Sanyo Information Business
	Transporter EJ/W	E → J	Windows	
	Transporter JE	J → E	Windows	
5	JK Machine Translation System	J → K	MS-DOS 5.0	Hitachi Information Network
	KJ Machine Translation System	K → J	MS-DOS 5.0	
6	PIVOT/JE	J → E	EWS-UX/V	NEC
	PIVOT/EJ	E → J	EWS-UX/V	
	EJ Translation Adapter	E → J	Windows	
	EJ Jisho Biki Kun	E-J	Windows	
	JE Jisho Biki Kun	J-E	Windows	
7	ATLAS EJ for Windows	E → J	Windows	Fujitsu
	ATLAS JE for Windows	J → E	Windows	
	Translation Surfing	E → J	Windows	
	Denjikai	E-J, J-E	Windows	
8	Pensee/V EJ	E → J	Windows	Oki Software & Ogis Lab.
	Pensee/V JE	J → E	Windows	
	Pensee for Internet	E → J	Windows	
	Pensee WWW Server Model	E → J	Solaris 1.X	
9	X-EJ2/W	E → J	Windows	Spirit
	X-EJ	E → J	MS-DOS	
	X-DIC	E → J		
10	Logo Vista E to J	E → J	Windows	Catena & Logo Vista
		E → J	Mac	
	Logo Vista E to J Personal	E → J	Windows	
		E → J	Mac	
	The Translator Mini	E → J	Mac	
	Korya EJ	E → J	Windows	
		E → J	Mac	
11	Uchi no Honyaku Yasan	E → J	Windows	Nagase Sangyo
		E → J	Mac	
		J → E	Windows	
		J → E	Mac	

Table 2: MT product list as surveyed by AAMT in 1997 (cont'd)

No.	Product Name	Language	OS	Developer
12	ASTRANSAC Sun WS	E-J	UNIX	Toshiba
	ASTRANSAC C/S	E-J	OS/AS	
		E-J	Windows NT	
	ASTRANSAC for Windows	E → J	Windows NT	
		J → E	Windows NT	
		E-J	Windows NT	
	ASTRANSAC/Web	E-J	OS/AS	
13	J · London/EJ	E → J	Windows	Koudensha
	J · London/JE	J → E	Windows	
	J · London Twin Set	E-J	Windows	
	J · Seoul JK	J → K	MS-DOS 3.1	
	J · Seoul KJ	K → J	MS-DOS 3.1	
	WorldNet/EJ	E-J	Windows	
14	Dr. Surf for Windows	E → J	Windows	Media Vision & Kyushu
	Hon Yaku Kakume	E → J	Windows	Matsushita Denki
15	Transpad	E → J	Windows	Kameshima Sangyo
	E-J BANK for Windows	E → J	Windows	
16	Translation Manager/2	E → J	OS/2	IBM Japan
	Honyaku no Ohsama	E-J	Windows	
17	Songoku	J → C	MS-DOS 3.1	Create Osaka
18	Net Surfer/ej	E → J	Windows	Nova & Yamatane Sangyo
	PC-Transer/ej	E → J	Windows	
		E → J	Mac	
		E → J	DOS(PC98)	
		E → J	DOS(DOS/V)	
	PC-Transer/je	J → E	Windows	
		J → E	Mac	
		J → E	DOS(PC98)	
		J → E	DOS(DOS/V)	
	My Transer EJ	E → J	DOS(PC98)	
		E → J	DOS(DOS/V)	
	Super Transer/ej	E → J	DOS(PC98)	
		E → J	DOS(DOS/V)	
19	Power Translator Pro.	G, F, S	Windows	Iris International
			DOS	
			Mac	
		R-E	DOS	
	Power Translator	G, F, S	Windows	
			DOS	
			Mac	
	Power Translator Chinese	E → C	DOS	
	Power Translator CD-ROM	G, F, S	Windows	
20	EJ Translation Support System	E → J	Windows	Hitachi Seisakusho
	EJ Translation Support System		UNIX	
	JE Translation Support System	J → E	UNIX	

unreliability of translation, MT applications had previously failed to gain general use. Not only lower prices but also the following factors have led to widespread of MT in the Japanese market:

- Translation accuracy raised to a nearly acceptable level.
- Interface with web browsers saves trouble in inputting data.
- Translation speed is acceptable due to present day high-performance hardware.
- Translation can be adjusted to user requirements.
- Fault tolerance of ungrammatical input.
- Many systems are bundled together with hardware packages.

It is noticeable that almost all of the mentioned products are for the translation between Japanese and English. Few of them are focused at other European languages, Korean or Chinese.

Other than the above home-user oriented MT systems which have to be low-priced and compact, there are also some noteworthy middle-range systems developed and used within organizations. The MT system developed by NHK is used for broadcasting-related tasks such as news capture, superimposition and so on. The system has to be able to handle various specialist fields, including the frequent used of proper nouns. NTT has developed a system called ALT-J/E focusing on natural translation. Introducing a case-frame transfer method restricting the translation scope in the target language. ATR is working towards developing a speech-to-speech translation system which is expected to be able to support interpretation at international conferences or over a telephone. JICST, through a new organization called JST, follows on from the MT system of Mu-project in providing a translation service for Japanese research paper abstracts.

The needs of English-to-Japanese language translation are clearly laid out in a JEIDA (Japan Electronics Industry Development Association) investigation on the requirements for English use over the Internet in Japan. The committee on text processing technology at JEIDA is a subcommittee of JEIDA's committee on natural language processing, and has been developing a bilingual corpus for research on machine translation systems since

the 1996 Japanese fiscal year [6]. The committee has carried out a survey on the current standing and requirements of information use over the Internet, through a questionnaire.

To accurately analyze the needs of users new to computer networks, according to use of the Internet and assistance functions, the committee designed a questionnaire on (1) the use of document data, (2) electronic newspapers, (3) digital libraries, (4) English use over networks, and (5) information retrieval. The survey was carried out from April 1996 to May 1996. Of the 214 answers, 50% were from engineers or scientists and 35% were from teaching staff and students. All questions in JEIDA's questionnaire, results and their analysis can be found in the JEIDA homepage,
`http://www.jeida.or.jp/committee/textsyori/sec-0.html`.

Concerning the results of asking, "How do you use English on the Internet?", we can summarize the responses as follows:

- Ninety percent of the responses stated they used the WWW for information gathering, and 80% of WWW users answered as accessing it several times a week. This indicates that due to the growing use of the WWW, frequency of access to information written in English is increasing. On the whole, access for information gathering purposes is much more frequent than access for information dissemination purposes, in the case of Japan.

- As for the purpose of using English, "to understand the basic content of the text" got the highest score. People tend not to and do not want to simply surf the network. They gather necessary information from the huge amount of information written in English.

- As for inconvenience in using English, less than 15% of responses indicated that they did not feel inconvenienced in having to use English.

- As for the way of overcoming the inconvenience of using English, printed or electronic dictionaries are frequently used; however, the percentage of persons who use machine translation systems is less than 10%.

Concerning machine translation systems, almost half of the answers indicated experience in using a machine translation system. Though this survey was carried out one and a half years ago, when there were not as many

low-cost machine translation software packages available as at present, the percentage of having experience with machine translation systems was higher than we had expected.

Of the people who had used machine translation applications, no one replied as having "no dissatisfaction (or being completely satisfied)". This is embarrassing to MT researchers in that it implies that machine translation systems have not reached a level at which users are fully satisfied with them.

Because many users come across English text on some occasions over the Internet, and almost all users experience difficulty in using English, the demand for machine translation systems is reasonably high. Indeed, the percentage of users satisfied with the summarization functionality of English reading assistance tools is high. At the same time, because printed or electronic dictionaries are used to resolve problems with English, current machine translation systems are insufficient even for "assistance". We have to identify the items with which each user is satisfied in each situation in which English is necessary, and identify user needs with respect to machine translation systems.

3.2 China

In additiona to the R&D for the MMT project, many bilingual MT projects, such as English-Chinese, German-Chinese, Russian-Chinese and Japanese-Chinese are being conducted separately. Specialized development aimed at a particular pair of languages seems to be more practical and straightforward for high demand language pairs.

The Chinese government plays an important role in supporting national MT R&D projects. Support is focused on follow-up research to the MMT project, in both improving translation quality and evaluation methods, aimed at boosting performance to practical levels. The government is also concentrating on constructing bilingual MT systems for such pairs of languages as Chinese and any of English, Japanese or German. Dynamic translation through a network is also within the scope of support.

MT R&D activities have increased in recent years. This can be observed from the increasing number of research institutes. Most of these are aimed at the development of a practical bilingual MT systems. Some are extending the achievements of MT R&D in part to other applications, such as word segmentation, part-of-speech tagging, phrase boundary identification and bilingual

parallel text alignment.

Regarding commercial MT activities, there are at least three systems on the market at present. All of them are English to Chinese MT systems aimed at the personal computer platform. Unsatisfactory translation accuracy (70-80%) is the main reason for the ambivalent market attitude towards MT in China.

3.3 Indonesia

BPPT (Agency for the Assessment and Application of Technology) of Indonesia has individually expanded the achievements of the MMT project by developing a tool for linguistic research called LOP (Linguistic Operation Panel). LOP is developed for unix-based PCs. The system provides a user friendly interface for grammar and dictionary development. Access to the MMT system prototype is now available on-line.

Though there is no commercial MT system at present, MT-related technologies have been applied to some applications, such as hyphenation in word processor and grammar checking.

English is reported as the highest priority for MT development. Many other MT R&D activities are being conducted in universities, most of them concerned with Indonesian-English MT.

3.4 Malaysia

ITNM (Malaysian National Institute of Translation), a previous partner in the MMT project, has re-organized itself to focus on translation services. MT R&D activities have thus been transferred to other research institutes attached to universities. Some NLP-related research exists focusing on constructing bilingual dictionaries (Malay and English) and thesauri.

Many commercial products are being derived from applied NLP technology for example spell checkers, grammar checkers, word processors, bilingual dictionaries between Malay, English, Chinese and French. A talking dictionary for Malay, Chinese and English has also been released.

Though there is no evidence of MT R&D, requirements of NLP technology are obvious through the widespread dictionary development and grammar research.

3.5 Thailand

LINKS (Linguistic and Knowledge Science Laboratory) of NECTEC (National Electronics and Computer Technology Center) has prepared Thai language resources for language processing research purposes, intended to be developed into a full MT system or other NLP application. So far it has been reported that there are two projects to develop English-to-Thai MT systems. One is to be conducted in LINKS and the other is in collaboration with a university and an organization from Singapore. However, there are currently no commercial plans for MT development.

LINKS, in collaboration with CRL (Communications Research Laboratory) of Japanese Ministry of Posts and Telecommunications, have constructed a tagged corpus for the Thai language, called the ORCHID corpus. The corpus is tagged with an original part-of-speech tagset, an improved version of the tagset used in the MMT system. The corpus consists of about 2MB (or about 400K words) of the proceedings of the NECTEC annual conference. It is scheduled for release at the end of 1997.

The wide use of computers in the present day has stimulated the need for standardized processing of the Thai language. As a national project, NECTEC has set up a plan for developing a Thai standard software library. It covers all fundamental technologies for processing Thai, including character coding, font design, word segmentation, and extending to speech processing.

3.6 Singapore

ISS (Institute of Systems Science) of the National University of Singapore has made a significant progress in the development of MT for English, Chinese and Malay. Implementation has been via a transfer-based architecture, and the MT service is currently available on the market. ISS also plans to increase the scope of language pairs for translation. Besides its success in making MT available, ISS has also developed a multi-linguage support system called MASS (Multi-lingual Application Support Service) which is now available under both the Unix and Windows environments.

Many other kinds of multi-lingual processing projects are going on at the institute. Service through the Internet is also within the scope of development.

3.7 Korea

Much effort has been made toward in the development of Japanese-Korean MT systems. There are 5 such systems released on the market, only one system of which is for mainframe computers with the other four systems being for personal computers. All systems have been developed using the linguistic similarities existing between the two languages. While this has met with limited success, in order to gain higher accuracy in translation, close-study needs to be made of the differences between the languages [2].

Other than Japanese-Korean MT systems, there are also many systems for English-Korean MT developed in Korea.

3.8 Republic of China (Taiwan)

Most of the MT systems available in Taiwan are bilingual systems for Chinese-English translation. One of the significant research institutes is the National Tsing Hua University, which has already succeeded in commercializing a Chinese-English MT system.

Other than the above countries, MT R&D activities are also being conducted in areas such as India and Hong Kong. India has succeeded in developing an MT system for tranlation between Hindi and English.

4 NLP Related R&D and the Available NLP Resources

The numbers of research papers and conferences show that research activities in this region are increasing year by year. Conferences and workshops are held both locally and regionally. The following is a selection of the more prominent conferences held in Asia:

- NLP, The Association for Natural Language Processing, Japan. Held annually since 1995.
  `http://www.kyutech.ac.jp/nlp/index.html`
- PACLING, Pacific Association for Computational LINGuistics, Japan-Australia. Held every 2 years since 1989.

- AAMT, The Asia-Pacific Association for Machine Translation, Japan. MT activities and a supporting journal.
  http://www.jeida.or.jp/aamt

- ROCLING, Research on Computational Linguistics, Taiwan. Held annually since 1987.

- SNLP, Symposium on Natural Language Processing, Thailand. Held every 2 years since 1993.

- NLPRS, Natural Language Processing Pacific Rim Symposium - regional. Held every 2 years since 1991.

Recent afforts have been made to promote standardization and data exchangability, such as UPF (Universal Platform), an AAMT activity aimed at setting up a standard format for dictionary encoding, and MLIT (Standardization of Multi-lingual Information) aiming at setting up a standard for character codes, keyboards, input/output methods, fonts and so on.

NLP Resources A number of electronic dictionaries and corpora are available for research use. The followings are collections of resources. Further details are available through the indicated sites.

- CICC - achievements from the MMT project. Besides the technical reports on interlingua and the MMT system, there are also electronic dictionaries and corpora available for Chinese, Indonesian, Malay and Thai. For details see
  http://tyo-cc-server.cicc.or.jp/homepage/english/about/act/mt/mt.htm.

- EDR - electronic dictionaries and corpora are available for both English and Japanese. In addition to word dictionaries, large-scale also co-occurrence dictionaries and concept dictionaries are also available.
  For details see http://www.iijnet.or.jp/edr.

- IPA - both a Japanese electronic dictionary and corpus are available. For details see
  http://www.ipa.go.jp/STC/NIHONGO/IPAL/ipal.html.

Many other resources can also be found developed for NLP research purposes such as ATR, RWC and so on. From these, it is evident that NLP research in this region is growing and will make considerable contributions to the MT community. Some available language databases are shown in Tables 3 and 4[1].

Table 3: Language databases

Name	Organization	Language	Type	Size
EDR Corpus	EDR	J,E	Text	500k sent.
KAIST Tree Bank	KAIST	K	Text	1M words
RWC Corpus	RWC	J	Text	30M words
IPA Corpus	IPA	J	Text	60k sent.
JEIDA Corpus	JEIDA	J	Text	N/A
ATR Dialogue Database	ATR	J,E	Speech	1M words each
NHK News Database (1)	NHK	J	Text	37M char.
NHK News Database (2)	NHK	J,E	Text	1.2M words
AP News Database	NHK	E	Text	59M words
Kyoto Univ. Corpus	Kyoto Univ.	J	Text	10k sent.
ORCHID Corpus	LINKS	T	Text	2M char.

Table 4: Types of language databases

Name	Text	Tagged	Bracketed	Parallel
EDR Corpus	◯	◯	◯	-
KAIST Tree Bank	◯	◯	-	-
RWC Corpus	◯	◯	-	-
IPA Corpus	◯	◯	-	-
JEIDA Corpus	◯	-	-	-
ATR Dialogue Database	◯	◯	◯	◯
NHK News Database (1)	◯	-	-	-
NHK News Database (2)	◯	-	-	◯
AP News Database	◯	-	-	-
Kyoto Univ. Corpus	◯	◯	◯	-
ORCHID Corpus	◯	◯	-	-

[1]Presented by Ehara Terumasa, NHK.

5 Conclusions

It is evident that MT R&D in every Asian country involves consideration of the English language. The demand for English language compatibility will increase further when it becomes possible to freely connected to the Internet from almost anywhere. Most information flowing over the Internet is written in English. To absorb this flood of information, we need efficient tools to extract what we want, presented in a form we can understand quickly. Translation from English to our mother tongue is vital in this respect. Though the accuracy of fully automatized translation is barely acceptable, other factors such as low-price, high-speed and user friendliness have led to widespread use of MT in the Japanese market. Countries other than Japan are now starting to develop bilingual MT for the English language, after gaining impetus from the MMT project over the past ten years.

NLP resources such as online dictionaries and language corpora are also increasing year by year. They are expected to be a valuable resource for language study and MT research in the future. Providing on-line dictionary over the Internet can be another element that fills the needs of MT for both the gathering and dissemination of information.

6 Acknowledgments

We would like to thank Dr. Hitoshi Isahara of CRL for the helpful JEIDA results presented in Section 3.1 and Timothy Baldwin for many helpful comments.

References

[1] AAMT. 1996 *AAMT Journal*, The Asia-Pacific Association for Machine Translation, No.16.

[2] Choi, K. S. and Kim, T. W. 1996 Present Status of Japanese-Korean MT Systems and the Analysis, *Proceedings of The Second Annual Meeting of the Association for Natural Language Processing*, pages 433-443, (in Japanese).

[3] CICC. 1995 *Technical Report on Interlingua Final Edition*, 6-CICC-MT36, Center of the International Cooperation for Computerization.

[4] CICC. 1997 *Technical Report on Interlingua and Concept Classification*, 8-CICC-MT31, Center of the International Cooperation for Computerization.

[5] Komurasaki, M. 1994 Profile of International R&D Cooperation Project on Multi-lingual Machine Translation (MMT) System, *MMT'94*, Center of the International Cooperation for Computerization, pages 11-23.

[6] Isahara, H. 1997 The Present Situation and Requirements for Information Usage on the Internet - Based on a Questionnaire by JEIDA -, *MT Summit VI.*

[7] Nagao, M., Tsujii, J. and Nakamura, J. 1985 The Japanese Government Project for Machine Translation, *Computational Linguistics 11*, 2-3, pages 91-110.