Backgrounds

Japan

Population: 127,768,000 (2005 Census)

Inhabitants of other nationalities (over 100,000 people)

Korean:	467,000
Chinese:	350,000
Brazillians:	100,000
Fillipinos:	100,000

Ainu: spoken by less than 20 natives

Status of Japanese in Japan

Official Language, but no legal statement except the court law states so.

In edicational and administrative services, ability of Japanese is presupporsed.

Literacy rate of Japanese is 99.8% (UNESCO 2005)

Internet Connectivity

Households:
Connected with the wired media: 57%
Connected with the mobile phone: 57%

85% of population can access the Internet by at least one medium.

Schools and Companies:

Almost all the schools and companies are connected to the Internet.

PC user

Households:

80% of households have at least one PC.
72% of population use mobliephones.

Schools and Companies:

Almost all the schools have PCs for the educational purpose.

Local Language Policy

Localization of Japanese:

Three main themes:

1. Standardization of character sets
2. developing fonts for Japanese
3. developing the input software for Japanese

These are all requested by the nature of Japanese writing system which use many Kanji letters.

Character sets

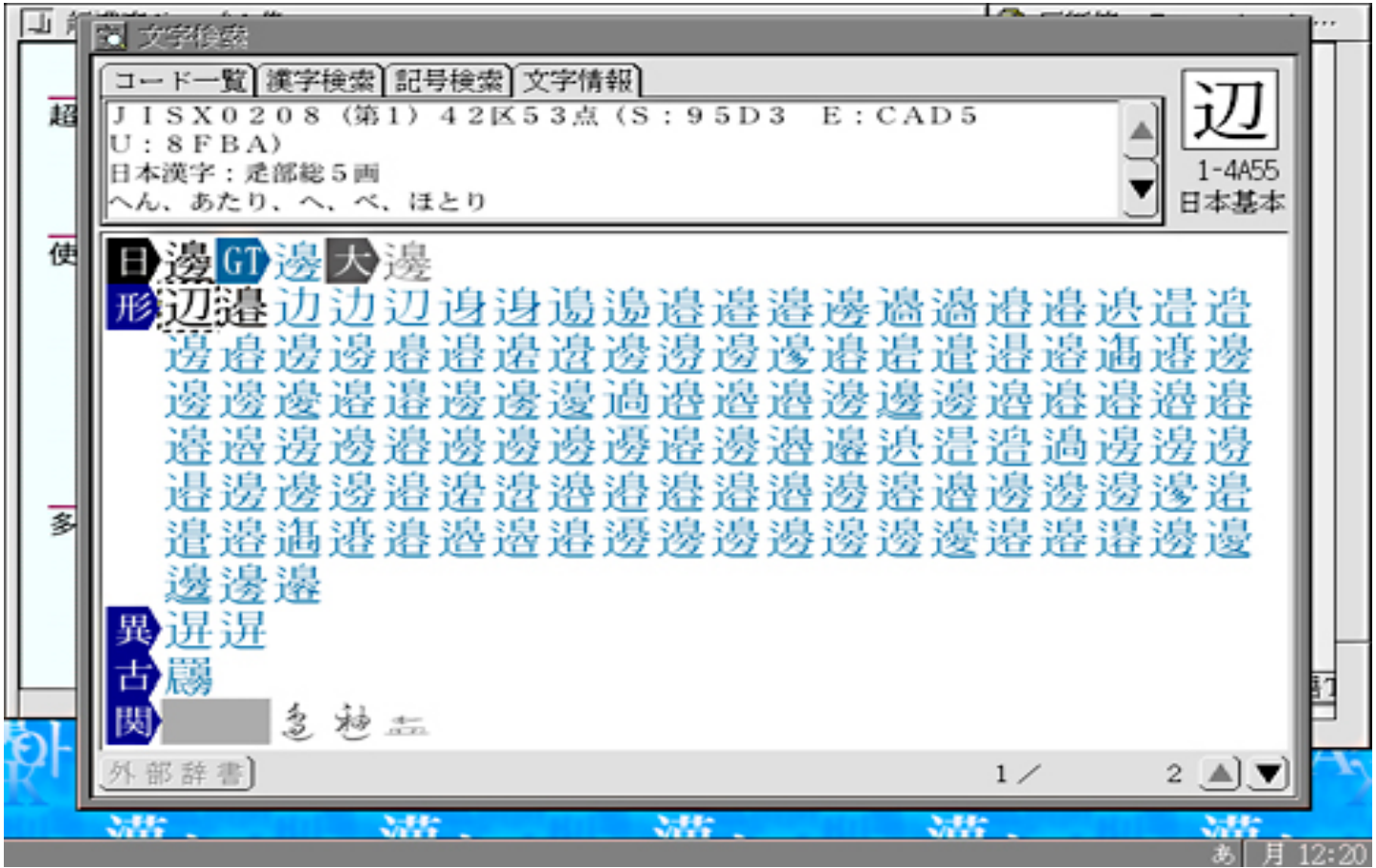
Standardization started in 1960's

	N. of Kanji	
1969		JIS C6220 (JIS X0201)
	0	
1978		JIS C6226-1978 (JIS X0208:1997)
6,802		
1990		JIS X0212
	6,067	
2000		JIS X0213 (JIS X0213:2004)
11,233		

Glyph Problem

Many of Kanji letters are expressed with more than one glyphs.

An extreme case:



Encoding Problem

Three different encoding schemes:

- S-JIS
- EUC
- JIS

Needs for identifying which encoding scheme is used in a document.

MOJI-BAKE problem: occures when encoding scheme used is not properly judged.

Standardization of Japanese

Kokugo Shigikai (Council of Japanese) under MEXT

Theme: Simplification and restriction of Kanji and reformation of spelling

Established in 1934

1946 Toyo Kanji (Kanji of daily use) and Gendai Kanadzukai (Current spelling of Japanese)

1981 Joyo Kanji (Kanji of usual use)

Font

Proprietary Fonts

as a product
as a part of Operating Systems

Free Font:

Kochi font
IPA font
Some other fonts under OSS movement

Some other fonts not along the national standards

Imput Software

Kanji input to computers:

Kana-Kanji translation

Kana-input → dictionary→show candidates→user's choice→Kanji-input

Proprietary softwares was already developed in 80's (8 bit machines were widely used)

Free softwares: SKK, Wnn, Kanna....

m17n libraries by FSS, IPA, ETL and AIST

Relevant Government Ministries and Departments

Ministry of Economy, Trade and Industry (METI)

General support for development of IT,

Ministry of Education, Culture, Sports, Science and Technology (MEXT)

Language Policy of Japanese and other languages

Ministry of Internal Affairs and Communications (MIC)

General Research activity

Other Public Organizations

The National Institute for Japanese Language

Research on the Japanese language

National Institute of Advanced Industrial Science and Technology (AIST)

Research and development of multilingual libraries and ontology-aware software

National Institute of Information and Communications Technology (NiCT)

Research and development of ICT including natural language processing

Center of the International Cooperation and Computerization (CICC)

International cooperation on IT

Japan Standards Association (JSA)

Standardization including Japanese

Information-Technology Promotion Agency (IPA)

Financial support for development of software

Other Non-Profit Organizations

The Association for Natural Language Processing

Academic community on natural language processing

Gengo Sigen Kyokai (GSK)

Promotion of sharing, reusing, and spreading linguistic resources (corpora and software)

Conclusion

- Localization of Japanese has almost completed.
- No governmental organiation plans local language computing policy.
- Relative rich human resource can be used.
- Japanese government shares resources to support or assist local language computing in othe countries.
- Development of m17n libraries by AIST is one of the examples
- ALRN is the another one.