Network-based Speech-to-Speech Translation Technology

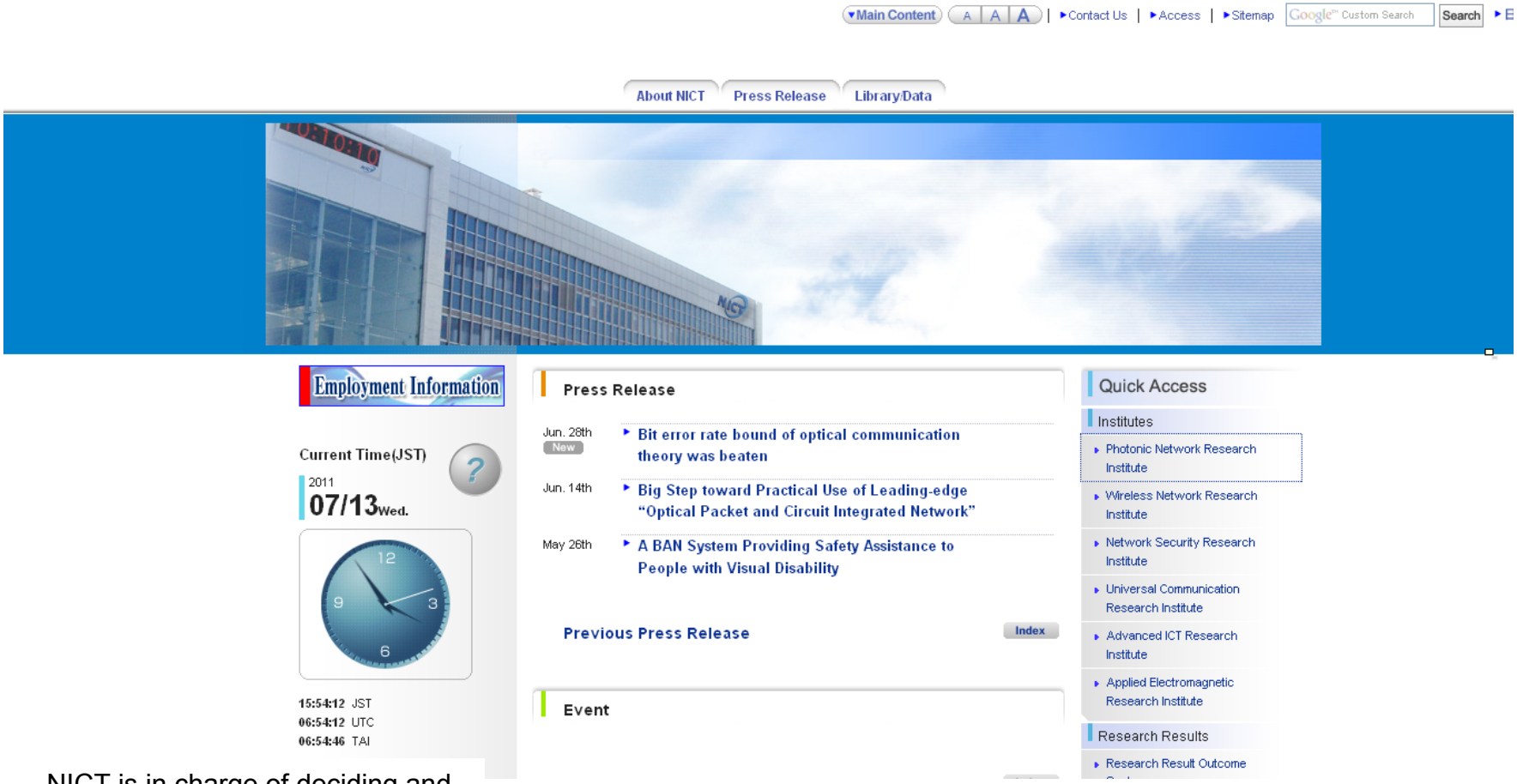
Chiori Hori

Spoken Language Communication Laboratory

MASTAR Project

National Institute of Information and Communications Technology

The National Institute of Information and Communications Technology (NICT)



NICT is in charge of deciding and maintaining Japan Standard Time.

National Institute of Information and Communications Technology

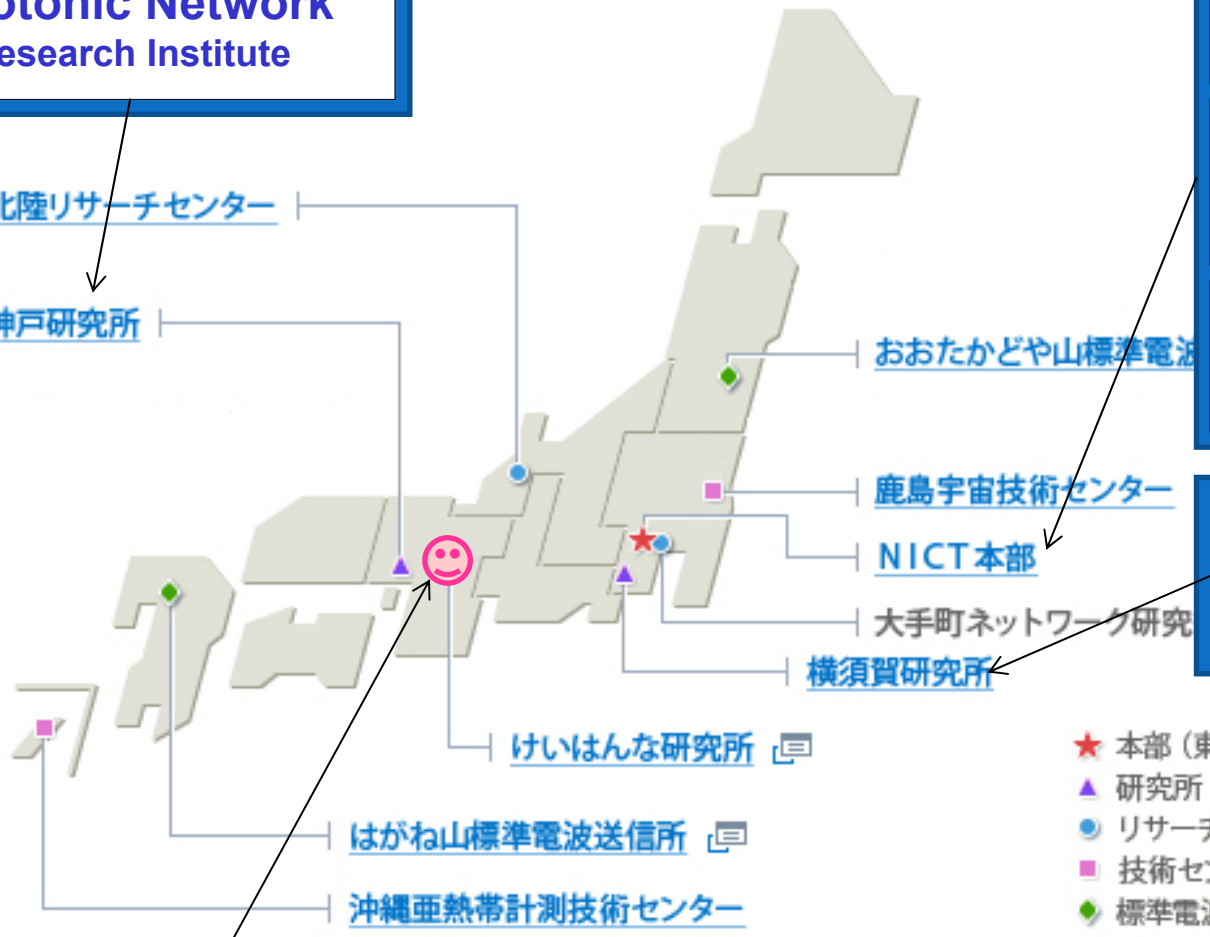
Photonic Network
Research Institute

Network Security
Research Institute

Applied
Electromagnetic
Research Institute

Advanced ICT
Research Institute

Wireless Network
Research Institute



▲ このページの先頭へ

Universal Communication
Research Institute
-Spoken Language Communication laboratory

MASTAR

Research Target

1. Spoken language communication technologies
   - human and machine
     -> Spoken Dialog system
   - humans who speak different languages
     -> Speech-to-Speech Translation system

2. Rich transcription technologies
   for speech/video archives
   -> Information Retrieval system

Goal of Speech-to-Speech Translation

- Many different languages in the world
- Overcoming the language barrier is a long-held dream of mankind.

- Speech-to-Speech Translation technology

Breaking the language barrier

http://en.wikipedia.org/wiki/List_of_language_families

Speech-to-Speech Translation(S2ST)

Japanese
「私は学校に行く」

Speech Recognition (ASR)

Machine Translation (MT)

Speech Synthesis (TTS)

English
“I go to school”

w a t a sh i
w a g a xtu
k o o n i…..

私は
学校に行く

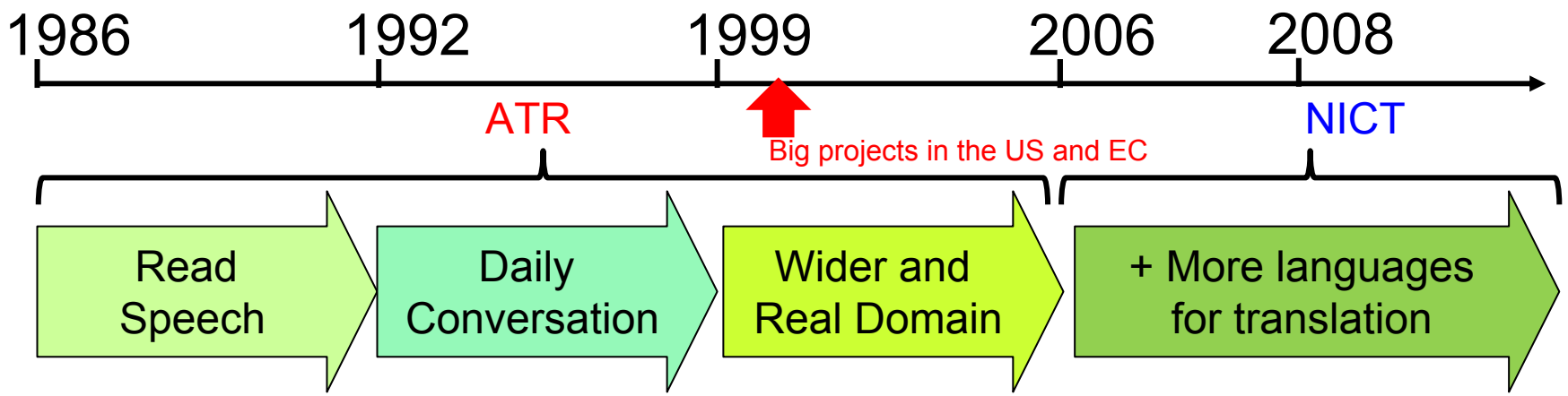
I to
school go

I go to
school

Convert to Japanese phoneme sequence “w”, “a”, “t”..	Convert to word sequence using lexicon and grammar	Convert to English word sequence 「私は」⇒ “I” 「学校に」⇒“to school 「行く」⇒“go”	Reorder word sequences according to English grammar “I” ⟷ “I” “to school” ⤫ “go” “go” ⤫ “to school”	Select appropriate waveform for English text

Corpora

History of S2ST research

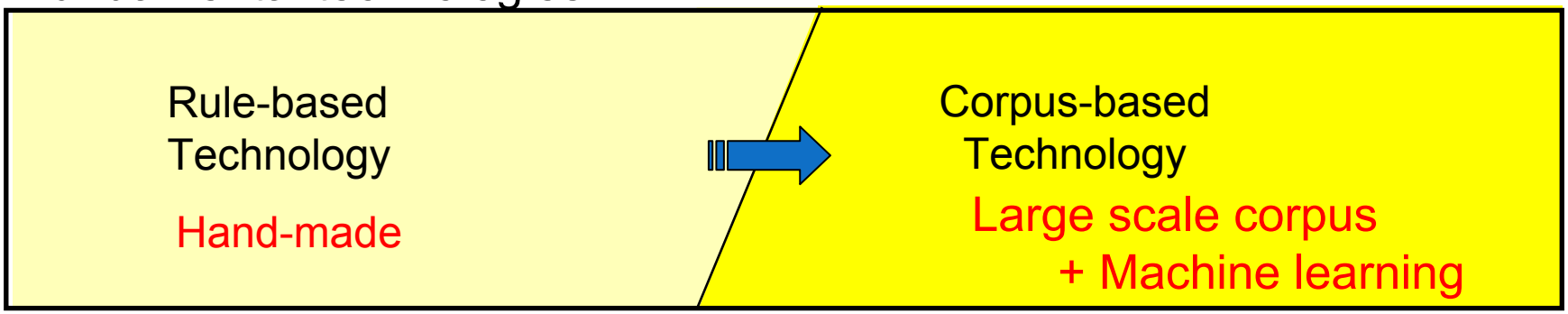


- Syntactically correct
- Clear utterance
- Limited domain
  "Conference Registration"

- Standard expression
- Unclear utterance
- Limited domain
  "Hotel Reservation"

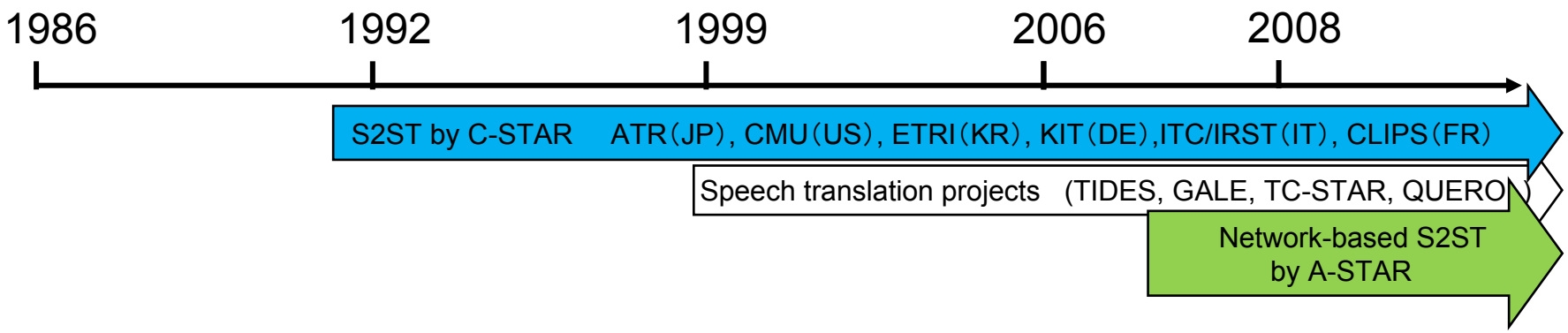
- Wider and real domain
  "International Travel"
- Realistic expressions
- Noisy speech
- J-E, J-C speech translation

Fundamental technologies



Network-based S2ST

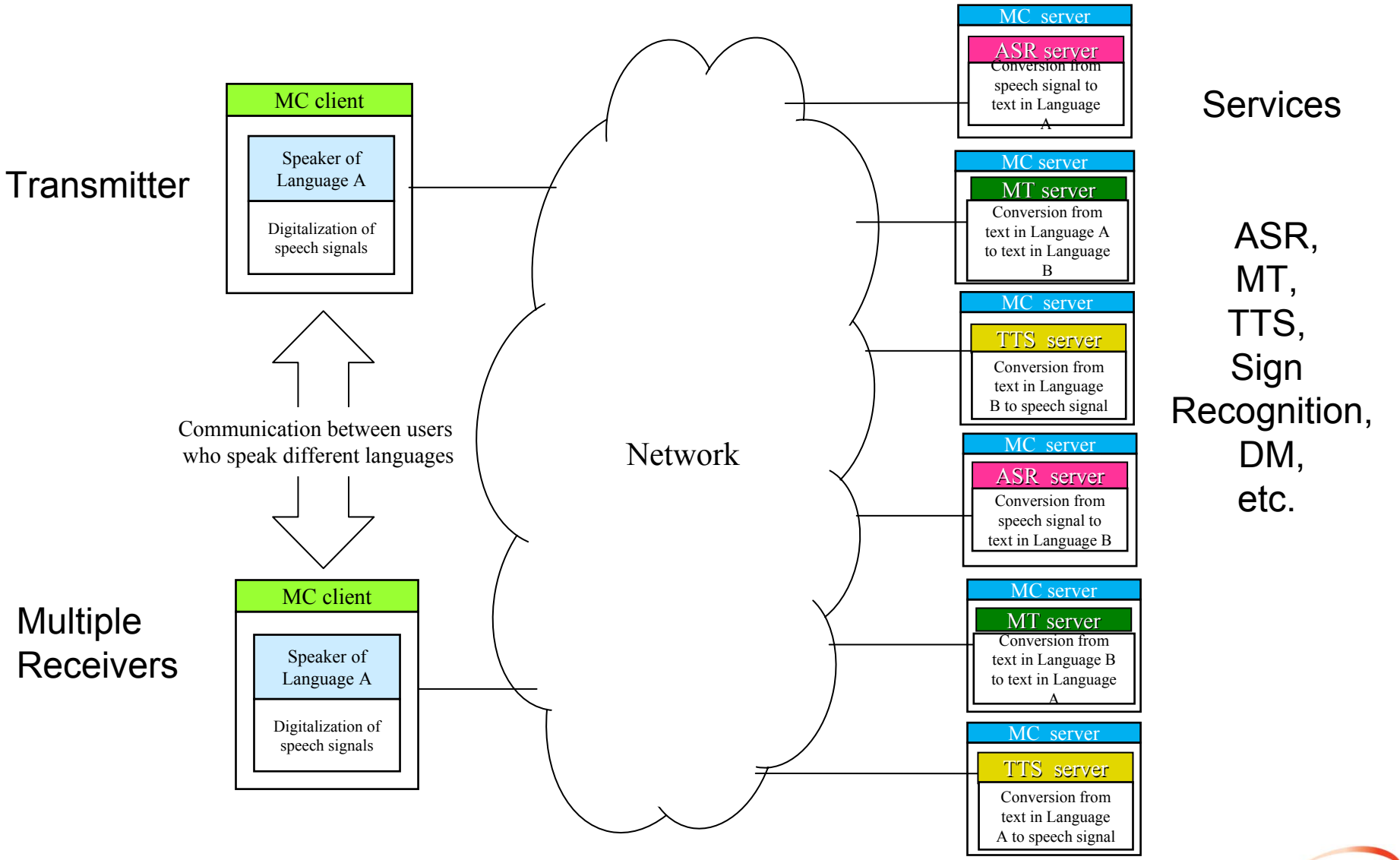
- Realizing S2ST for multi languages in the world, in a collaboration with the international society



- Connecting ASR, MT, and TTS modules existing in the world, via Network.

- Definition of communication protocol and data format between modules is needed.

Architecture of Network-based S2ST system



MCML

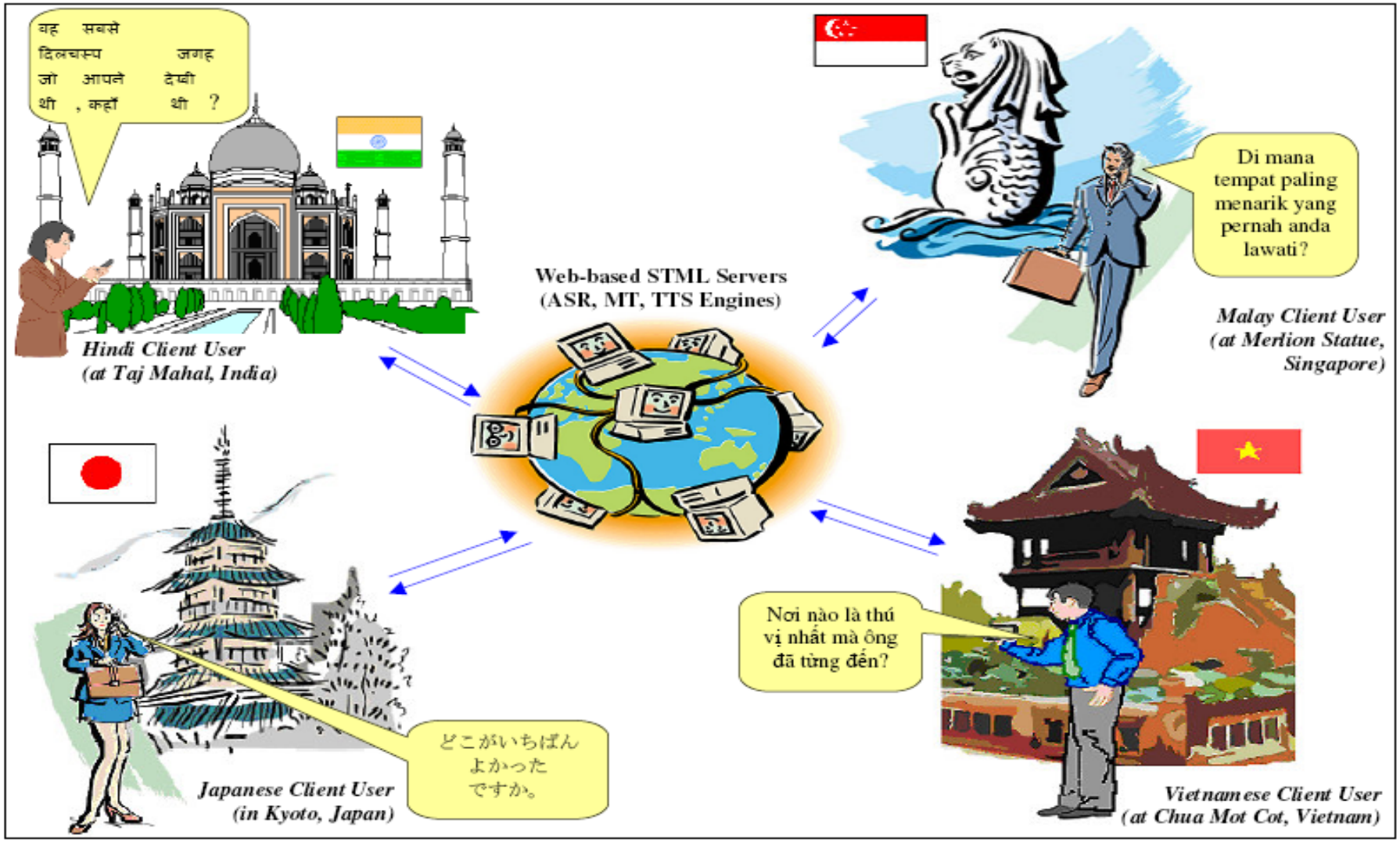
- Modality Conversion Markup Language
- XML schema
- MCML includes

  information for communication

  among multiple persons

  who use different modalities

Conversation of multiple Language speakers

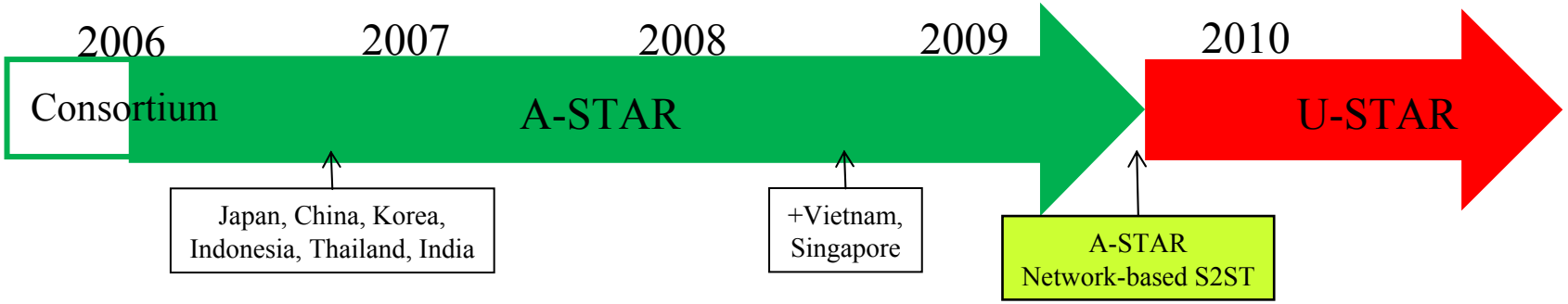


English meaning: "Where was the most interesting place you've visited?"

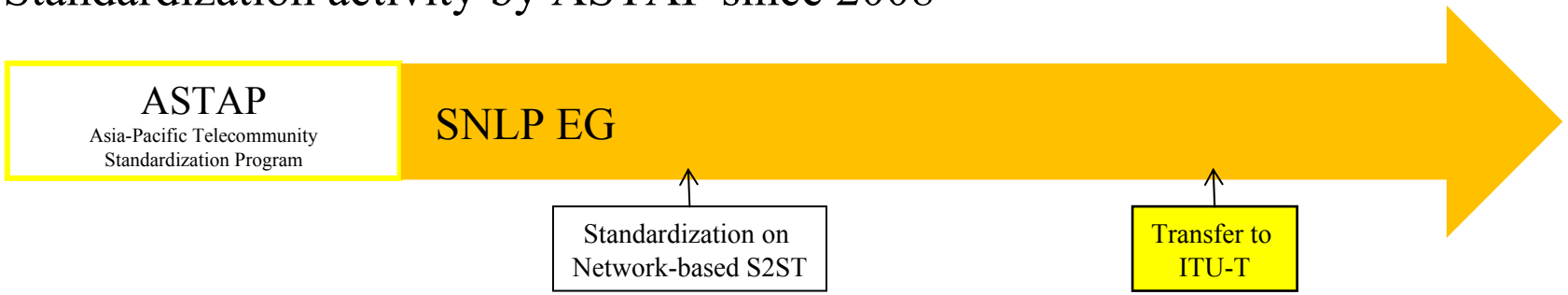
Multiparty conversation system

Standardization Activities

- Network-based S2ST research by ASTAP consortium



- Standardization activity by ASTAP since 2008



- Shift of standardization activity to ITU-T in 2009

Standardization activity of Network-based S2ST protocol at ITU-SG16

- Activity start for standardization of Network-based S2ST at ITU-T SG16
- Session period : October, 2009 to the present
- NICT is the editor for S2ST standardization at ITU-T SG16, WP2 Q21/22

Document	Title	Scope
F.745	Functional Requirements for Network-based S2ST	- Definition of Network-based S2ST - Functions and service requirements of network-based S2ST
H.625	Architectural Requirements for Network-based S2ST	- Requirements of S2ST architecture - Definition of interface for Network-based S2ST

- Not only language conversion but also the potentially added module like sign language are taken into account :

  S2ST -> Modality conversion

International Telecommunication Union

ITU-T F.745

TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU

(10/2010)

SERIES F: NON-TELEPHONE TELECOMMUNICATION SERVICES

Audiovisual services

Functional requirements for network-based speech-to-speech translation services

Recommendation ITU-T F.745

ITU-T F-SERIES RECOMMENDATIONS

NON-TELEPHONE TELECOMMUNICATION SERVICES

TELEGRAPH SERVICE	
Operating methods for the international public telegram service	F.1–F.19
The gentex network	F.20–F.29
Message switching	F.30–F.39
The international telemessage service	F.40–F.58
The international telex service	F.59–F.89
Statistics and publications on international telegraph services	F.90–F.99
Scheduled and leased communication services	F.100–F.104
Phototelegraph service	F.105–F.109
MOBILE SERVICE	
Mobile services and multidestination satellite services	F.110–F.159
TELEMATIC SERVICES	
Public facsimile service	F.160–F.199
Teletex service	F.200–F.299
Videotex service	F.300–F.349
General provisions for telematic services	F.350–F.399
MESSAGE HANDLING SERVICES	F.400–F.499
DIRECTORY SERVICES	F.500–F.549
DOCUMENT COMMUNICATION	
Document communication	F.550–F.579
Programming communication interfaces	F.580–F.599
DATA TRANSMISSION SERVICES	F.600–F.699
AUDIOVISUAL SERVICES	**F.700–F.799**
ISDN SERVICES	F.800–F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850–F.899
HUMAN FACTORS	F.900–F.999

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Name space

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    targetNamespace="http://www.itu.int/xml-namespace/itu-t/H.645/MCML.xsd "
    elementFormDefault="qualified" attributeFormDefault="unqualified">
    <xs:element name="MCML">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="User" />
                <xs:element ref="Server" maxOccurs="unbounded" />
                <xs:element ref="History" minOccurs="0"
    maxOccurs="unbounded" />
            </xs:sequence>
            <xs:attribute name="Version" type="xs:string" use="required" />
        </xs:complexType>
    </xs:element>

...
```

Field Experiment by VoiceTra

VoiceTra(Speech to Speech Translator by NICT) By NICT View More By This Developer

Open iTunes to buy and download apps.

View In iTunes

Description

You can translate the contents of your spoken words to a foreign language using this application developed by National Institute of Information and Communications Technology (NICT). The target conversation content for the application is travel-related

NICT Web Site › VoiceTra(Speech to Speech Translator by NICT) Support › ...More

What's New in Version 3.4.0

(Feature Additions)
•Now available for use on the iPod touch (4G) and iPad2.
•Speaking can now be performed by tapping the screen.

http://www.youtube.com/watch?v=v6ULvxsPh54&feature=player_detailpage

Further expansion of
Asian languages
and
plus European languages

Growth of
multi-lingual S2ST network