

Multilingual Translation

Dr. Eiichiro Sumita

Associate Director General
Universal Communication Research Institute,
NICT, Japan

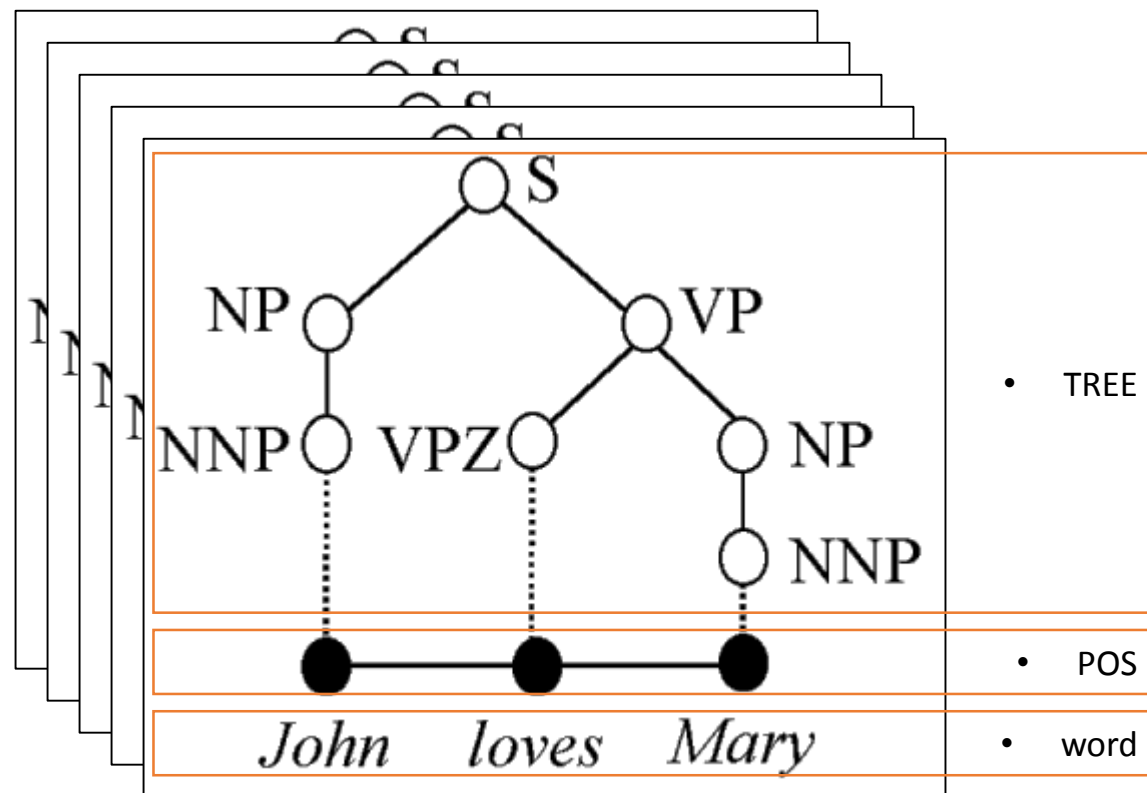
Outline

- 1. Asian Language Treebank (ALT)**
- 2. Workshop on Asian Translation (WAT)**
- 3. Universal Speech Translation Advanced Research (U-STAR)**
4. Recent Topics in NICT's Translation Research

Asian Language Treebank (ALT)

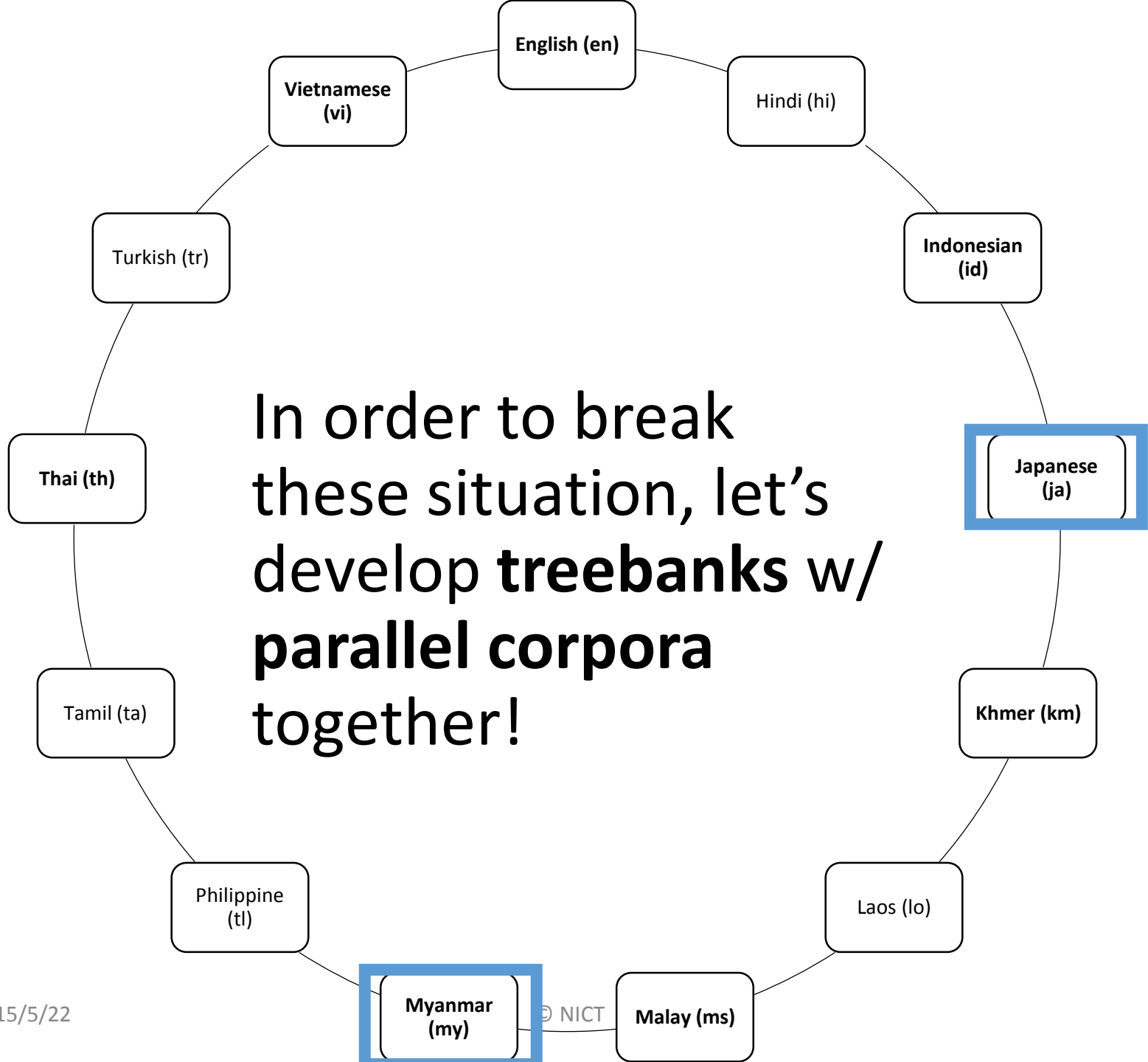


- Treebank **accelerates** research of NLP for each language
- **No** publicly available **POS-tagged** and **constituency tree corpora** for **many Asian** languages.





- **Parallel corpus accelerates research of Machine Translation among the languages**
- **No big parallel corpora among many Asian languages**



In order to break these situation, let's develop **treebanks w/ parallel corpora** together!

4 steps for developing

1. **Translating** the common English sentence into the Asian Language
2. **Aligning words** between the sentence pair of English and Asian Language
3. **Tagging the POS (Part-Of-Speech)** for the sentence of the Asian Language
4. **Building the tree** of the sentence of the Asian Language

WEB-based UI for ALT's 4 steps

Site

タスク

ワーク

Task

Translation

http://

task desc

カナダ

Translate

Tr

カナ

Can

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

カナダ

Translate

Site

タスク

ワーク

Task

Translation

http://

task descr

カナダ

Translate

Alig

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

Site Tit

タスク

ワークス

Task Ti

Translation

http://www.gc

task description

カナダ

Translate

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

Site Title

タスク

ワークスペース

Task Title

Translation

http://www.google.co.jp

task description XXX

カナダ

Translate

Alig

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

表示言語/language: 日本語 ヘルプ help

タスク

ワークスペース

Task Title

Translation

http://www.google.co.jp

task description XXX

カナダ

Translate

Alig

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

カナ

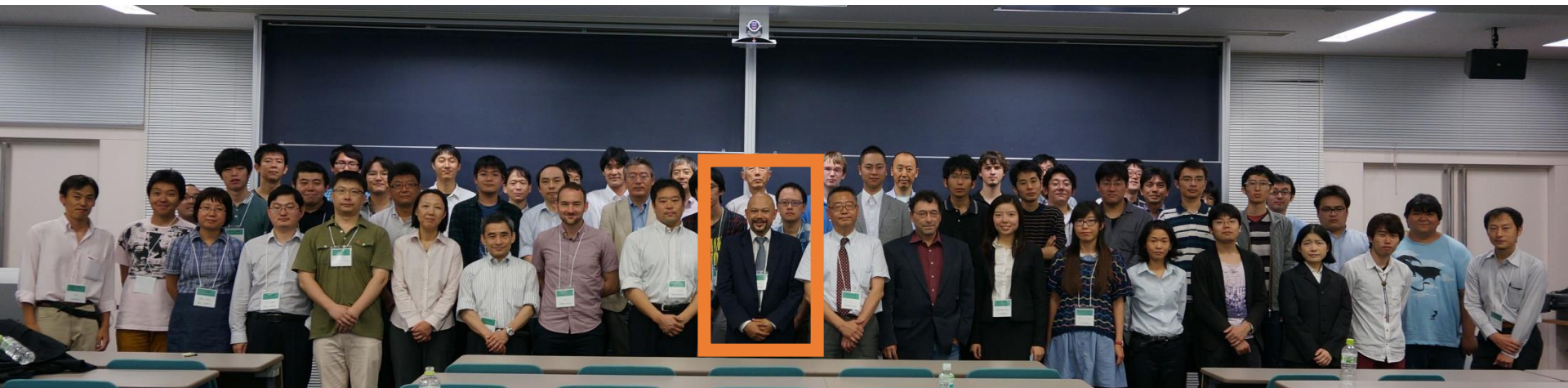
カナ

カナ

カナ

カナ

Workshop on Asian Translation (WAT)



1st WAT (Workshop on Asian translation)

<http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2014/index.html>

- Was held in **Tokyo on October, 2014**
- Had over **fifty participants**
- MT evaluation campaign focusing on **scientific documents in Japanese-English/Japanese-Chinese**

2nd WAT <http://lotus.kuee.kyoto-u.ac.jp/WAT/patent/index.html>

- Will be held on October, 2015
- New Task: **translation of patent document** with bilingual Japanese-Chinese corpus (1 M) provided by Japan Patent Office (JPO)

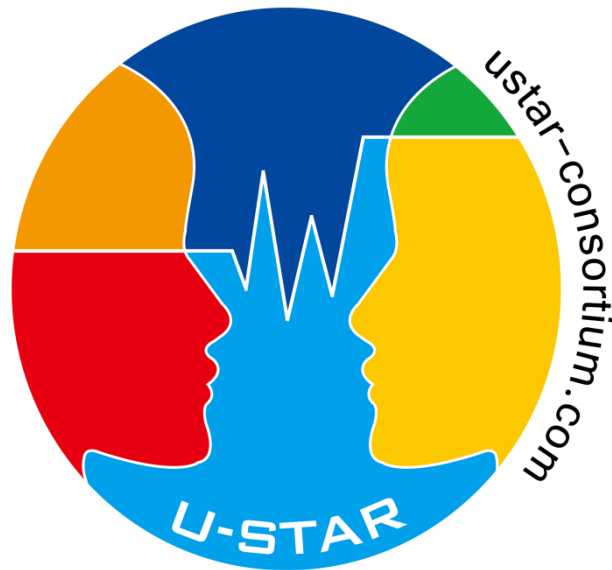
We'd like to ask you to **join 2nd WAT.**

3rd WAT

- Will be held in 2016, possibly being collocated by **Coling 2016** (Osaka, Japan).
- New Task:

We'd like to ask you **provide 3rd WAT with parallel corpus of other Asian languages** and join 3rd WAT.

Universal Speech Translation Advanced Research (U-STAR)



30 Institutes from 25 countries/regions
 formed under the “Memorandum of Understanding” (Current MOU: valid until March 31st, 2016)



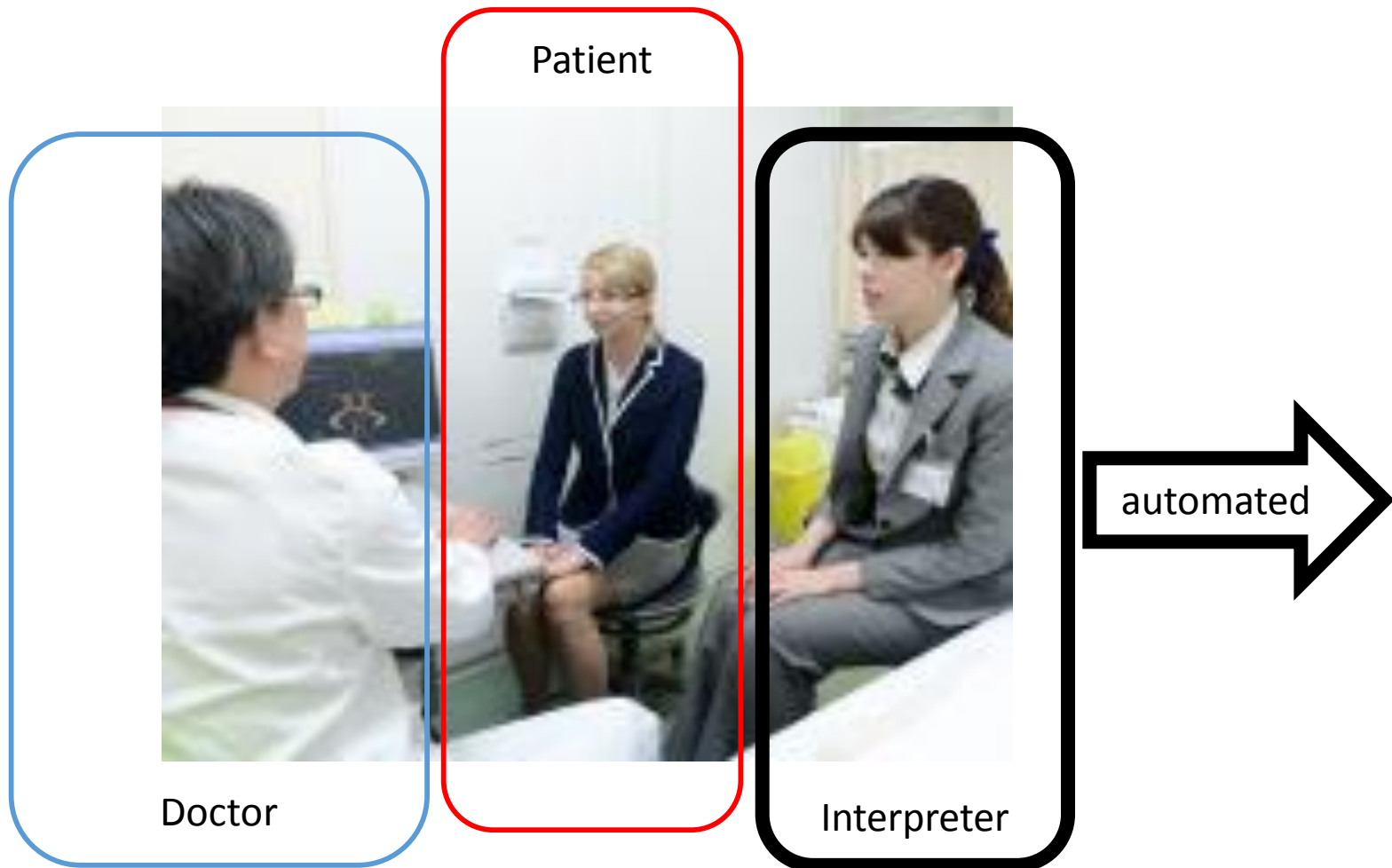
World-wide experiment using smartphone application of speech translation



Machine **Translation** for 27 languages
Speech **Recognition** for 17 languages
Speech **Synthesis** for 14 languages

Recent Topics in NICT's Translation Research

Speech Translation for **Medical** Domain



Text Translation for **Patent** Domain



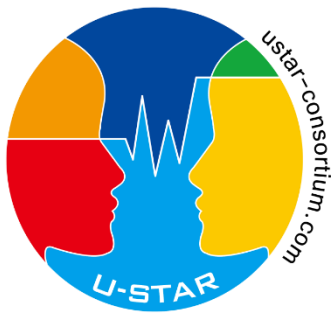
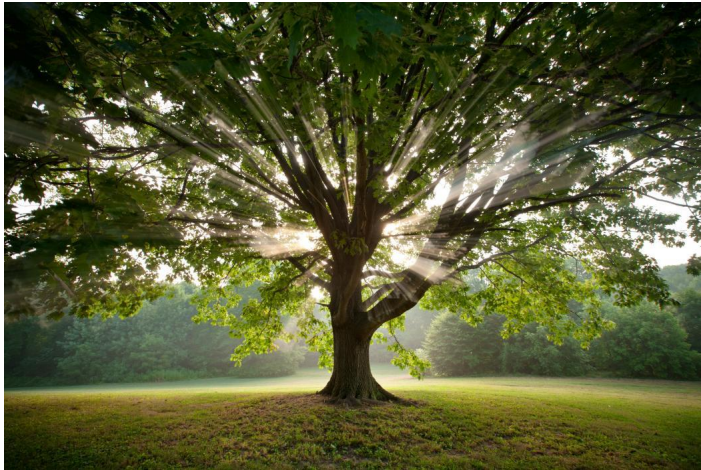
Advanced SMT (ML-based Preordering)

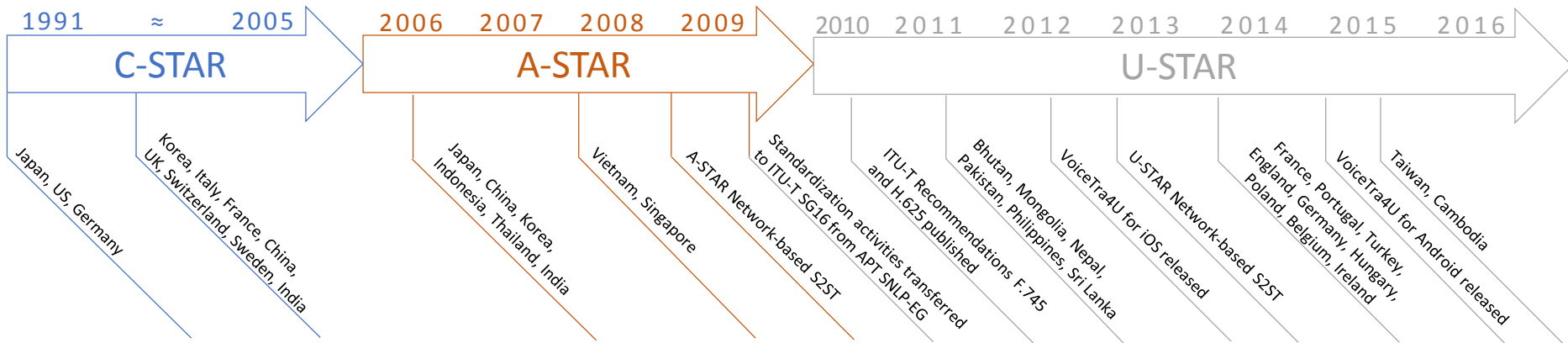
- i. Parse** the input English sentence
- ii. Preorder** the English sentence to Japanese word order
- iii. Translate** the Japanese-ordered English sentences in monotone

Development Situation by Language Pair

Japanese <-> English, Chinese	Commercialized
Japanese <-> Korean	Under development
Japanese <-> Thai, Indonesian, Vietnamese	Under investigation

Wrap up



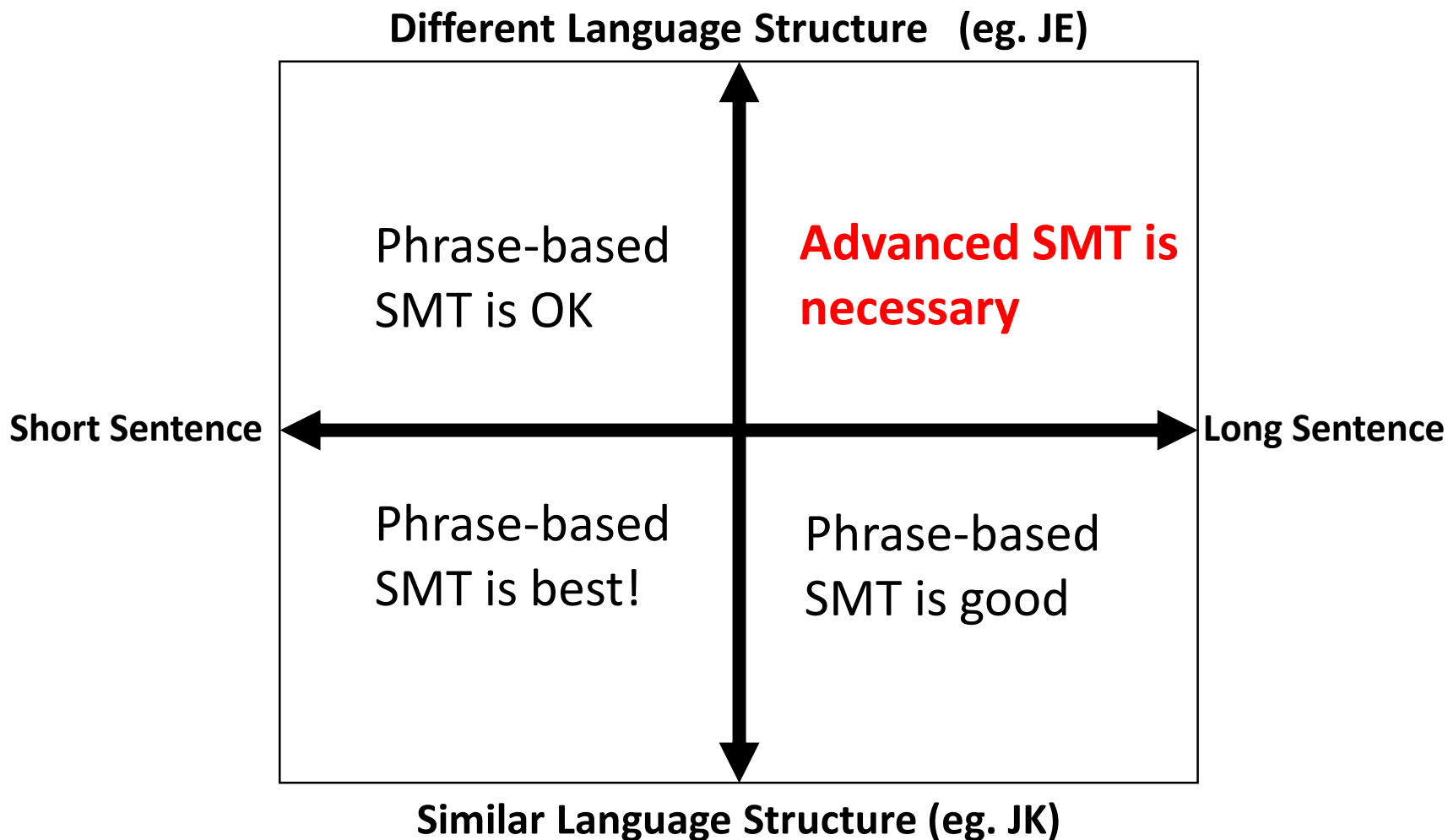


The *Consortium for Speech Translation Advanced Research* (C-STAR) started out over 20 years ago to develop multilingual speech translation systems. Numerous post activities and workshops derived from C-STAR such as the *International Workshop on Spoken Language Translation* (IWSLT).

The *Asian Speech Translation Advanced Research* (A-STAR) was formed in the Asian regions to develop a network-based speech-to-speech translation (S2ST) system. A-STAR initiated in standardizing international communication protocols, especially in the S2ST field, in association with the *Asia-Pacific Telecommunity Standardization Program* (ASTAP) and launched "the first Asian network-based speech-to-speech translation system", on July 29th, 2009. The system enabled real-time, location-free, multi-party communication between speakers using different Asian languages, and confirmed the feasibility of network-based S2ST protocol.

In 2010, the standardizing procedures at ASTAP were transferred to *International Telecommunication Union Standardization Sector* (ITU-T) as A-STAR shifted to U-STAR, transforming not only its name but its organization to a worldwide consortium with the aim of establishing a more global system. U-STAR's network-based S2ST is developed based on the ITU-T Recommendations F.745, and H.625, which were both published in October, 2010.

Technologies differ in conditions



Input:

FIG. 3C is graph illustrating a simulation that includes the effects of resonance, cyclic clocks, and a change in logic current.

Result of STEP i. (English sentence in Japanese order)

FIG. 3C _va1_ resonance of effects , cyclic clocks , and logic current in change _va2_ includes that simulation _va2_ illustrating graph is .

Result of STEP ii. (translation without change of order)

図3Cは、共振による効果、環状のクロック、および論理電流の変化を含むシミュレーションを示すグラフである。

An ideal translation:

図3Cは、レゾナンス効果、周期クロックおよび論理電流の変化を含むシミュレーションを示すグラフである。