THE JAPANESE FRAMENET PROJECT:
A Preliminary Report

KYOKO HIROSE OHARA, SEIKO FUJII*, HIROAKI SAITO,

SHUN ISHIZAKI, TOSHIO OHORI*, RYOKO SUZUKI

Keio University, Japan
University of Tokyo, Japan

1. WHAT IS JAPANESE FRAMENET?

An attempt to create an online lexical resource of Japanese basic verbs

- Based on corpus evidence
- Analysis based on frame semantics

In collaboration with Berkeley **FrameNet** project.
(http://www.icsi.berkeley.edu/~framenet/)
However,
we are interested in the question of whether Japanese words can be described in **FrameNet** style, i.e., along the same lines as English words, employing the same frame semantic approach.

First Phase: July 2002 – March 2005
Headquarter: Hiyoshi Campus of Keio University
Japanese FrameNet team: Researchers from
Keio University and
University of Tokyo

Current Status:

- A corpus has been chosen.
- A search tool has been developed.
- A pilot study is being undertaken to analyze motion and communication verbs.

2. THE CORPUS AND THE TOOL FOR ANALYZING JAPANESE

The Japanese FrameNet **Corpus**

- Currently contains approximately 3-million sentences.
- Taken from **CD-Mainichi Newspaper** and **Kyoto University Annotated Text Corpus**.

The KWIC **Search Tool**

- Searches for both the root form and conjugated forms of a keyword at the same time.
- Can be used with a dependency structure analyzer **CaboCha** (developed at Nara Institute of Science and Technology), which performs syntactic parsing as well as morphological analysis. Using our KWIC search tool together with CaboCha thus enables us to add any text to our corpus.
- Three display modes:
  1) Parse Tree Mode
  2) Morphological Analysis Mode
  3) Context Display Mode
- Written in Ruby script language; Runs on Linux, Solaris, and Windows platforms.

A Screenshot of Parse Tree Mode

3. AN EXAMPLE: THE JAPANESE MOTION VERB *wataru* 'cross'

What is a semantic frame?
- Frame semantics characterizes the semantic and syntactic properties of words by relating them to semantic frames.
- These are schematic representations of situations involving various participant and other conceptual roles, each of which is a frame element.

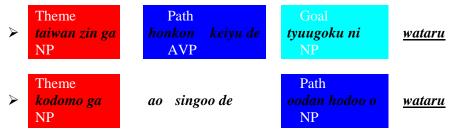An example frame: **Path_shape frame** (from Berkeley FrameNet homepage)
"The words in this frame describe motion in terms of the shape of the path traversed by the entity that moves."
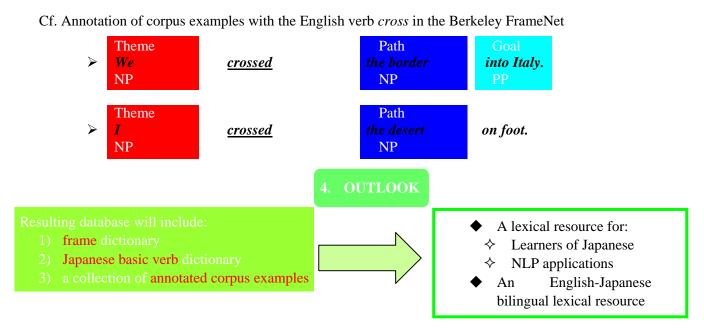- Some of the frame elements of the Path_Shape frame:
  - Theme: The object which moves.
  - Path: Any description of a trajectory of motion which is neither a Source nor a Goal.
  - Goal: Any expression which tells where the Theme ends up as a result of the motion.
  - etc.

A Japanese verb which evokes the Path_shape frame: *wataru* 'cross'

- *taiwan zin ga honkon keiyu de tyuugoku ni wataru*
  Taiwanese NOM Hong Kong via China GOAL cross
  'Taiwanese enter China via Hong Kong.'
- *kodomo ga ao singoo de oodan hodoo o wataru*
  kids NOM green light LOC crosswalk ACC cross
  'Kids cross a crosswalk on the green light.'

Annotation of corpus examples with *wataru* in the Japanese FrameNet

Theme	Path	Goal	
taiwan zin ga	*honkon keiyu de*	*tyuugoku ni*	*wataru*
NP	AVP	NP	

Theme		Path	
kodomo ga	*ao singoo de*	*oodan hodoo o*	*wataru*
NP		NP	

Cf. Annotation of corpus examples with the English verb *cross* in the Berkeley FrameNet

Theme		Path	Goal
We	*crossed*	*the border*	*into Italy.*
NP		NP	PP

Theme		Path	
I	*crossed*	*the desert*	*on foot.*
NP		NP	

4. OUTLOOK

Resulting database will include:
1) frame dictionary
2) Japanese basic verb dictionary
3) a collection of annotated corpus examples

- A lexical resource for:
  - Learners of Japanese
  - NLP applications
- An English-Japanese bilingual lexical resource

Japanese FrameNet details: http://www.nak.ics.keio.ac.jp/jfn/index.html