MMT with ICONS and ACE

Sanghoun Song
Univ. of Washington
sanghoun@uw.edu

DELPH-IN 2013
July 30, 2013

Overview

- A Grammar Library for Information Structure
  - Information Structure
    - meanings: focus, topic, contrast, background
    - markings: prosody, lexical markers, syntactic positioning
  - Individual Constraints (ICONS): MRS-based representation
  - ACE
    - regression tests
    - multilingual machine translation: text/sentence-based processing

1. Individual Constraints
   - Motivation
   - Type Hierarchies
   - Sample Representations
2. MMT
   - ACE
   - Basic Machinery
   - Grammars
   - Evaluation
3. Progress & Plan

1. Individual Constraints
   - Motivation
   - Type Hierarchies
   - Sample Representations
2. MMT
   - ACE
   - Basic Machinery
   - Grammars
   - Evaluation
3. Progress & Plan

Motivation

Using ICONS is motivated by three necessities.

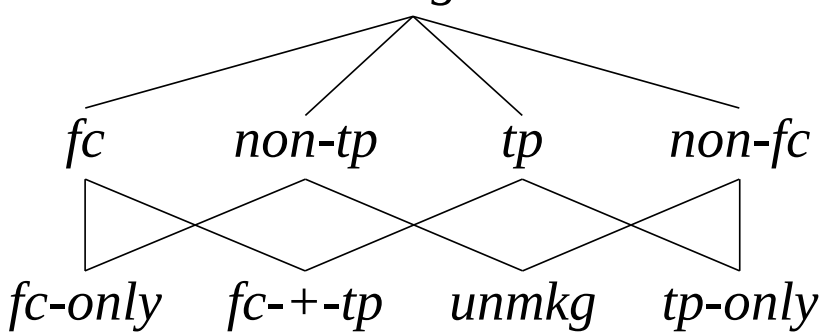
- morphosyntactic **markings** (*mkg*) *vs.* semantic **representation** (*info-str*): resolving discrepancies between forms and meanings in information structure
- **underspecification**: facilitating underspecifiability for allowing flexible and partial constraints

  inu wa hoeru.
  dog TOP bark
  'The **dog** barks.' [jpn]
- **multiclausal constructions**: capturing a binary relation of information structural components

  The dog that Kim chases barks.

mkg

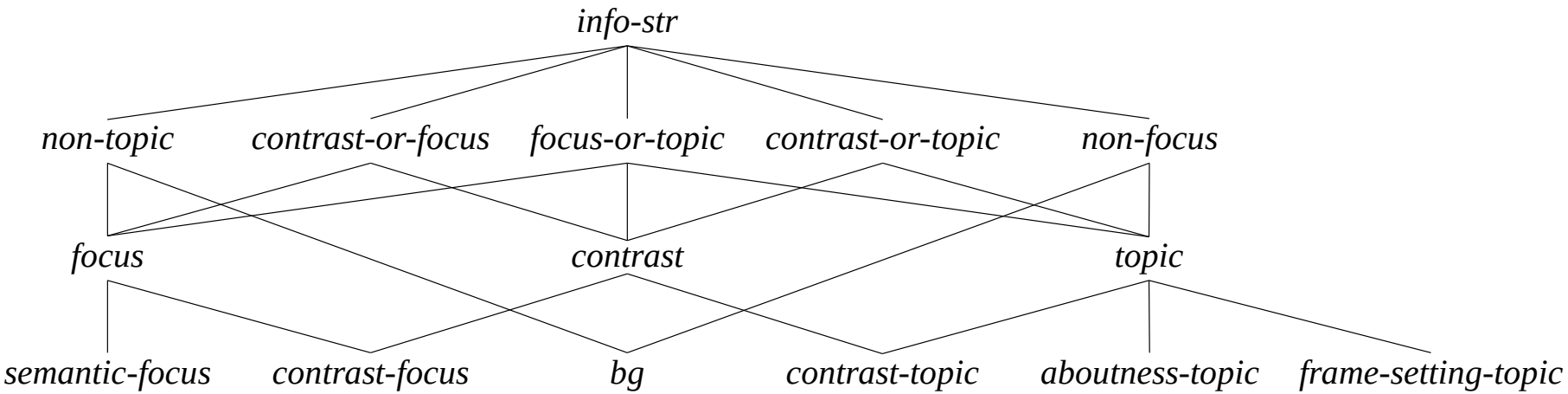
$$\left[\text{MKG} \quad \begin{bmatrix} \text{FC} & luk \\ \text{TP} & luk \end{bmatrix}\right]$$



icons

$$
\left[\begin{array}{ll}
\textit{mrs} & \\
\text{HOOK} & \left[\begin{array}{ll}
\textit{hook} & \\
\text{GTOP} & \textit{handle} \\
\text{LTOP} & \textit{handle} \\
\text{INDEX} & \textit{individual} \\
\text{XARG} & \textit{individual} \\
\text{ICONS-KEY} & \textit{info-str} \\
\text{CLAUSE-KEY} & \textit{event}
\end{array}\right] \\
\text{RELS} & \textit{diff-list} \\
\text{HCONS} & \textit{diff-list} \\
\text{ICONS} & \left\langle ! \ldots, \left[\begin{array}{ll}
\textit{info-str} & \\
\text{CLAUSE} & \textit{individual} \\
\text{TARGET} & \textit{individual}
\end{array}\right], \ldots ! \right\rangle
\end{array}\right]
$$

info-str


info-str
non-topic
contrast-or-focus
focus-or-topic
contrast-or-topic
non-focus
focus
contrast
topic
semantic-focus
contrast-focus
bg
contrast-topic
aboutness-topic
frame-setting-topic

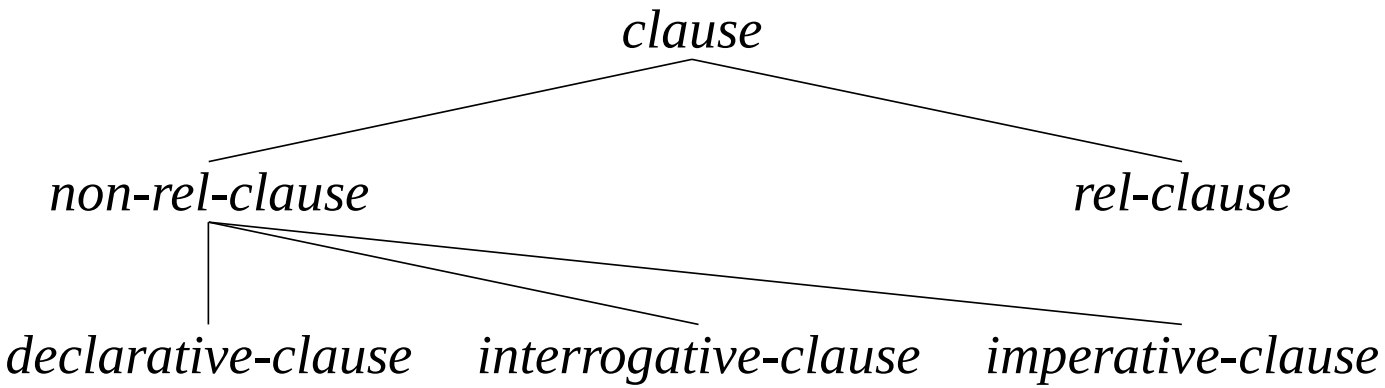
icons-lex-item

$$\left[\begin{array}{l} \textit{no-icons-lex-item} \\ \text{HOOK} \left[\begin{array}{ll} \text{ICONS-KEY}|\text{CLAUSE} & \boxed{1} \\ \text{CLAUSE-KEY} & \boxed{1} \end{array}\right] \\ \text{ICONS} \left\langle ! \quad ! \right\rangle \end{array}\right]$$

$$\left[\begin{array}{l} \textit{single-icons-lex-item} \\ \text{HOOK} \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{ICONS-KEY} & \boxed{2} \end{array}\right] \\ \text{ICONS} \left\langle ! \; \boxed{2} \left[\text{TARGET} \quad \boxed{1}\right] ! \right\rangle \end{array}\right]$$

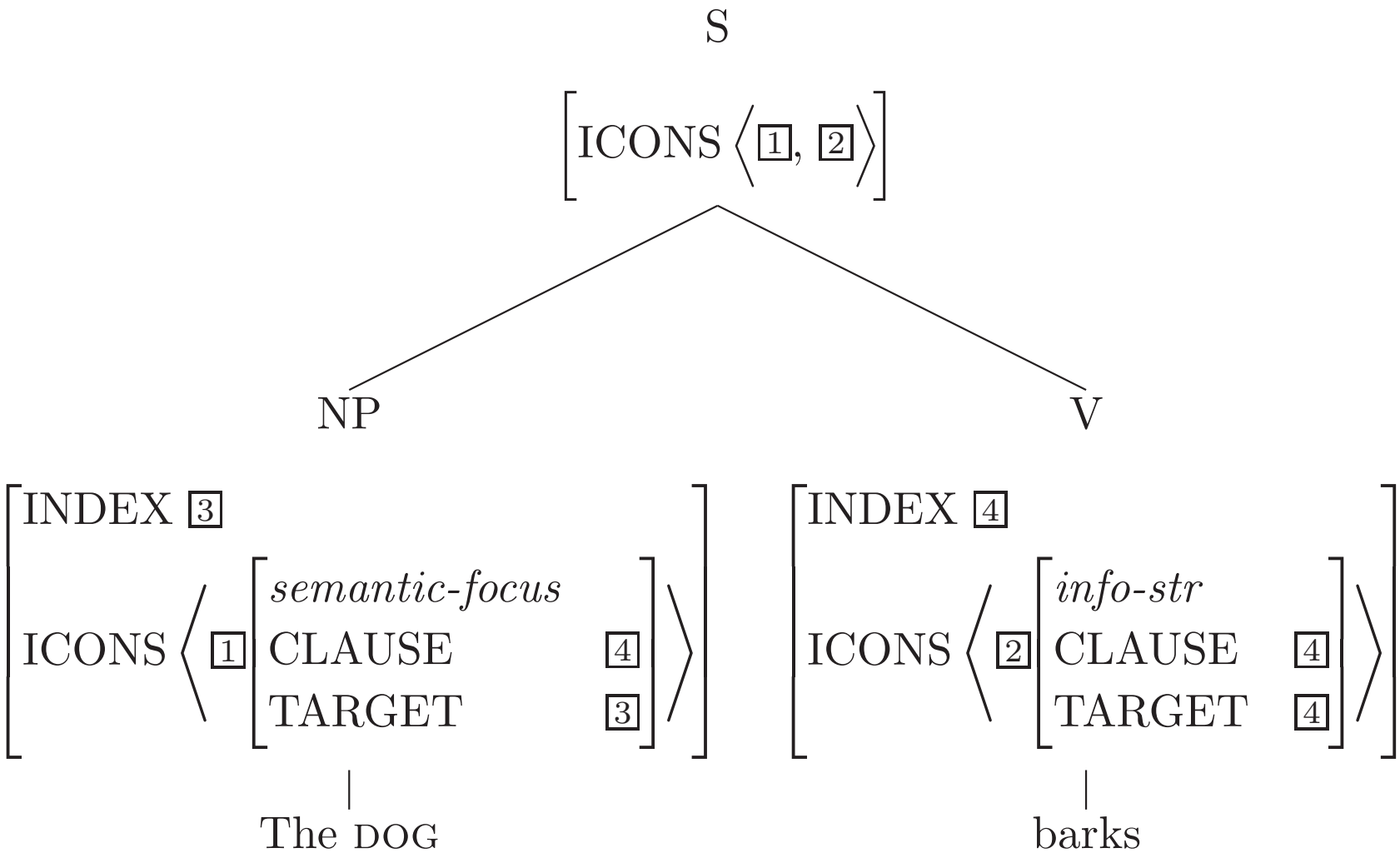
$$\left[\begin{array}{l} \textit{double-icons-lex-item} \\ \text{HOOK} \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{ICONS-KEY} & \boxed{2} \end{array}\right] \\ \text{ICONS} \left\langle ! \; \boxed{2} \left[\text{TARGET} \quad \boxed{1}\right], \left[\;\right] ! \right\rangle \end{array}\right]$$

clause



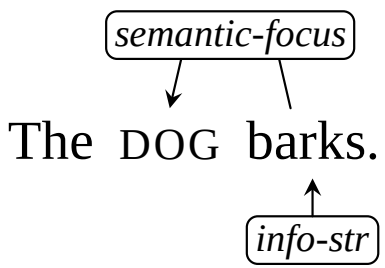
$$\left[\begin{array}{l} \textit{non-rel-clause} \\ \text{NON-LOCAL} \mid \text{REL} \;\; \textit{0-dlist} \\ \text{HD} \left[\begin{array}{ll} \text{HOOK} & \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{ICONS-KEY} \mid \text{CLAUSE} & \boxed{1} \\ \text{CLAUSE-KEY} & \boxed{1} \end{array}\right] \\ \text{NON-LOCAL} & \left[\begin{array}{ll} \text{QUE} & \textit{0-dlist} \\ \text{REL} & \textit{0-dlist} \end{array}\right] \end{array}\right] \end{array}\right]$$

The DOG *barks.*

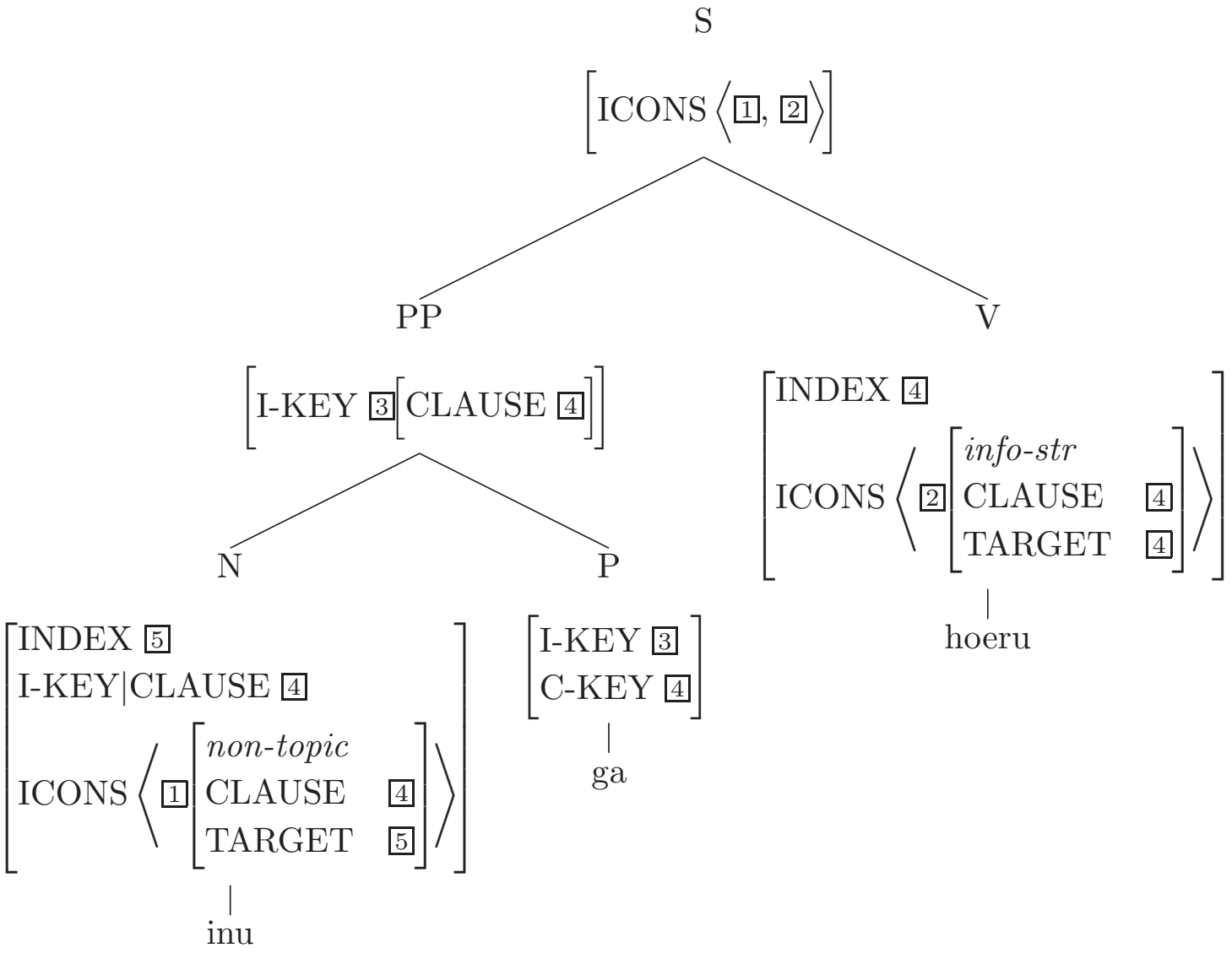


MRS & Dependency Graphs

$$
\left[\begin{array}{ll}
\textit{mrs} & \\
\text{LTOP} & \boxed{h1} \\
\text{INDEX} & \boxed{e2} \\
\text{RELS} & \left\langle \left[\begin{array}{ll} \textit{exist_q_rel} & \\ \text{LBL} & \boxed{h3} \\ \text{ARG0} & \boxed{x4} \\ \text{RSTR} & \boxed{h5} \\ \text{BODY} & \boxed{h6} \end{array}\right], \left[\begin{array}{ll} \textit{_dog_n_rel} & \\ \text{LBL} & \boxed{h7} \\ \text{ARG0} & \boxed{x4} \end{array}\right], \left[\begin{array}{ll} \textit{_bark_v_rel} & \\ \text{LBL} & \boxed{h8} \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{x4} \end{array}\right] \right\rangle \\
\text{HCONS} & \left\langle \left[\begin{array}{ll} \textit{qeq} & \\ \text{HARG} & \boxed{h5} \\ \text{LARG} & \boxed{h7} \end{array}\right] \right\rangle \\
\text{ICONS} & \left\langle \left[\begin{array}{ll} \textit{semantic-focus} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{x4} \end{array}\right], \left[\begin{array}{ll} \textit{info-str} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{e2} \end{array}\right] \right\rangle
\end{array}\right]
$$

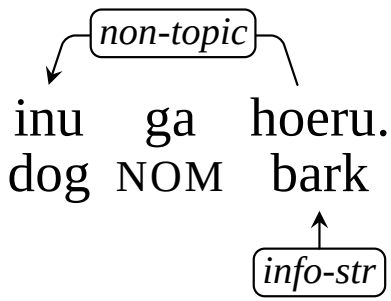


inu ga hoeru.


S
ICONS ⟨1, 2⟩
PP
V
I-KEY 3 CLAUSE 4
N
P
INDEX 4
ICONS ⟨2 info-str, CLAUSE 4, TARGET 4⟩
hoeru
INDEX 5
I-KEY|CLAUSE 4
ICONS ⟨1 non-topic, CLAUSE 4, TARGET 5⟩
inu
I-KEY 3
C-KEY 4
ga

MRS & Dependency Graphs

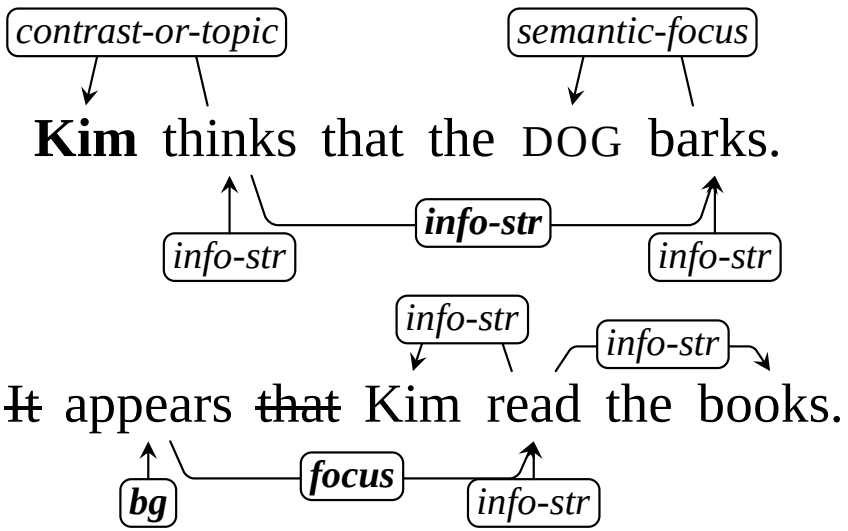
$$
\begin{bmatrix}
\textit{mrs} & \\
\text{LTOP} & \boxed{h1} \\
\text{INDEX} & \boxed{e2} \\
\text{RELS} & \left\langle \begin{bmatrix} \textit{_inu_n_rel} & \\ \text{LBL} & \boxed{h3} \\ \text{ARG0} & \boxed{x4} \end{bmatrix}, \begin{bmatrix} \textit{exist_q_rel} & \\ \text{LBL} & \boxed{h5} \\ \text{ARG0} & \boxed{x4} \\ \text{RSTR} & \boxed{h6} \\ \text{BODY} & \boxed{h7} \end{bmatrix}, \begin{bmatrix} \textit{_hoeru_v_rel} & \\ \text{LBL} & \boxed{h8} \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{x4} \end{bmatrix} \right\rangle \\
\text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{h6} \\ \text{LARG} & \boxed{h3} \end{bmatrix} \right\rangle \\
\text{ICONS} & \left\langle \begin{bmatrix} \textit{non-topic} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{x4} \end{bmatrix}, \begin{bmatrix} \textit{info-str} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{e2} \end{bmatrix} \right\rangle
\end{bmatrix}
$$



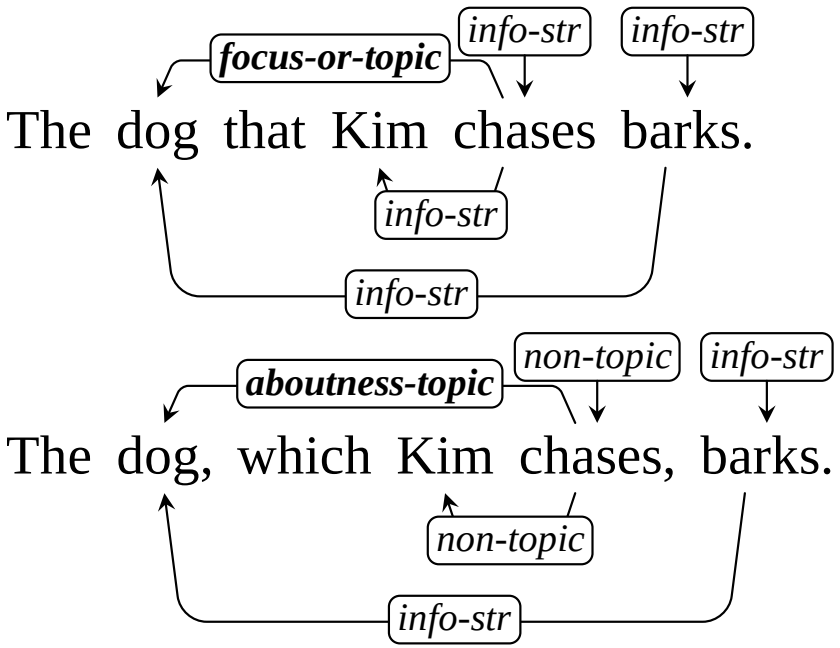
Multiclausal Constructions

- complement clauses: verbs of saying ("say"), semi-factive verbs ("realize"), and quasi-evidential verbs ("it appears")
- relative clauses: restrictive *vs.* non-restrictive
- advervial clauses: preposed *when*/*if* clauses

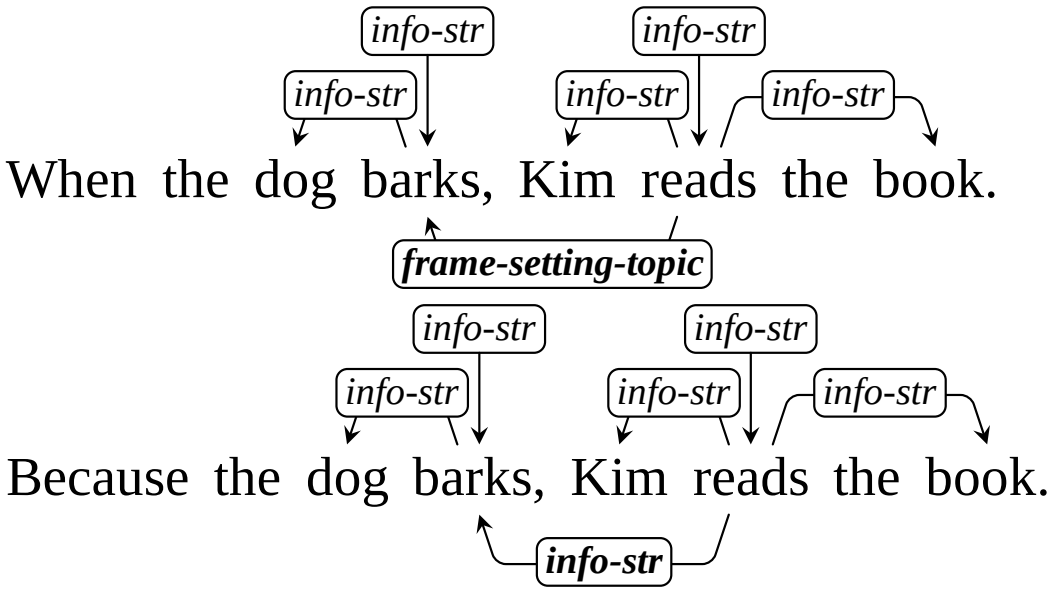
Complement Clauses



Relative Clauses



Advervial Clauses



1 Individual Constraints
- Motivation
- Type Hierarchies
- Sample Representations

2 MMT
- ACE
- Basic Machinery
- Grammars
- Evaluation

3 Progress & Plan

config.tdl

```
enable-icons := yes.
mrs-icons-list := ICONS LIST.
icons-left := CLAUSE.
icons-right := TARGET.
```

English

The dog-**a** barks [ICONS: < e2 **semantic-focus** x4, e2 info-str e2 >]
(i) The dog barks
(ii) The dog-a barks
(iii) The dog barks-a
(iv) The dog-a barks-a
(v) ~~The dog-b barks~~
(vi) ~~The dog-b barks-a~~

Japanese

inu **ga** hoeru [ICONS: < e2 **non-topic** x4, e2 info-str e2 >]
(i) inu ga hoeru
(ii) ~~inu wa hoeru~~
(iii) inu hoeru

Translation

a. The dog-**a** barks [ICONS: < e2 **semantic-focus** x4, e2 info-str e2 >]
   (i) inu ga hoeru
   (ii) ~~inu wa hoeru~~
   (iii) ~~inu hoeru~~

b. inu **ga** hoeru [ICONS: < e2 **non-topic** x4, e2 info-str e2 >]
   (i) The dog barks
   (ii) The dog-a barks
   (iii) The dog barks-a
   (iv) The dog-a barks-a
   (v) ~~The dog-b barks~~
   (vi) ~~The dog-b barks-a~~

Grammars: sobaka

- Pseudo Grammars: CN IV / PN TV CN
- Illustrative Grammars
  - English: prosody
  - Korean: affixes, scrambling
  - Japanese: adpositions, scrambling
  - Russian: modifiers, word order (cf. Varya Gracheva)
  - Spanish: *pro*-drop, clitic left dislocation
- Resource Grammars
  - ERG: discourse relations (focus_d_rel, parg_d_rel, itcleft_d_rel)
  - JaCY: adpositions, scrambling

MMT script

- using ACE
- running all tasks in an automatic way at once
  - parsing, transfer, generation
  - all directions (n×n)
- creating PDF files
  - ling567
- providing a report

MMT Evaluation

	eng	jpn	kor	rus	spa	ERG	JaCY
eng							
jpn							
kor							
rus							
spa							
ERG							
JaCY							

Items with end to end success

w ICONS

	eng	jpn	kor
eng	3	3	3
jpn	3	3	3
kor	3	3	3

w/o ICONS

	eng	jpn	kor
eng	3	3	3
jpn	3	3	3
kor	3	3	3

Items with exact match output

w ICONS

	eng	jpn	kor
eng	3	3	3
jpn	3	3	3
kor	3	3	3

w/o ICONS

	eng	jpn	kor
eng	3	3	3
jpn	3	3	3
kor	3	3	3

Total number of outputs

w ICONS

	eng	jpn	kor
eng	14	14	14
jpn	56	14	14
kor	56	14	14

w/o ICONS

	eng	jpn	kor
eng	18	18	18
jpn	72	18	18
kor	72	18	18

Average number of outputs

w ICONS

	eng	jpn	kor
eng	4.67	4.67	4.67
jpn	18.67	4.67	4.67
kor	18.67	4.67	4.67

w/o ICONS

	eng	jpn	kor
eng	6.0	6.0	6.0
jpn	24.0	6.0	6.0
kor	24.0	6.0	6.0

Max number of outputs

w ICONS

	eng	jpn	kor
eng	6	6	6
jpn	24	6	6
kor	24	6	6

w/o ICONS

	eng	jpn	kor
eng	6	6	6
jpn	24	6	6
kor	24	6	6

Test Suites

- Currently, 178 sentences + α
  - information structure-related phenomena
    - common: embedded clauses, argument structure, ...
    - language-specific: lexical markers, scrambling, *pro*-drop, left dislocation, ...
- In the future,
  - a bitext: the Tanaka corpus
  - a naturally occurring text: *The Adventure of the Speckled Band* out of the *Sherlock Holmes* stories

1 Individual Constraints
- Motivation
- Type Hierarchies
- Sample Representations

2 MMT
- ACE
- Basic Machinery
- Grammars
- Evaluation

3 Progress & Plan

What I have done so far

- a cross-linguistic survey
- a corpus study
- an information structure library
- grammars: pseudo / illustrative (eng, jpn, kor, rus) / ERG
- MMT settings

What I have to do soon

- grammars: illustrative (spa) / ERG / JaCY
- MMT evaluation
- a journal article

Issues

- focus projection
- other processors
- other grammars
- regression test
- unbound relations