# Individual Constraints for Information Structure*

Sanghoun Song and Emily M. Bender

Department of Linguistics, University of Washington
Box 354340 Seattle, WA 98195-4340, USA
{sanghoun, ebender}@uw.edu

**Summary.** This paper, in the context of multilingual MT, proposes the use of ICONS (individual constraints) to add a representation of information structure to MRS. The value of ICONS a list of objects of type *info-str*, with the features CLAUSE and TARGET. The subtypes of *info-str* indicate which information structural role is played by the TARGET with respect to the CLAUSE. This proposal is designed to support both the calculation of focus projection from underspecified representations and the handling of multiclausal sentences.

**Keywords:** HPSG, MRS, information structure, ICONS, MT

## 1 Introduction

This paper presents an HPSG (Pollard and Sag, 1994) analysis of information structure marking, with an eye towards practical applications such as machine translation (MT), adding constraints on information structure to MRS (Copestake et al., 2005) representations. In particular, we aim to improve on our previous analysis presented in Song and Bender (2011), to overcome two difficulties facing that work: First, we did not specify how the analysis could handle the spreading of focus beyond the lexical item directly marked for focus. Second, by encoding information structure as constraints on features of semantic variables ('variable properties'), we predicted that all occurrences of an index could share the same information structural properties. This is not necessarily the case, especially in constructions where semantic indices are shared across multiple clauses. This paper suggests the use of individual constraints (henceforth, ICONS), which (i) leave the information structural values of some constituents preferentially underspecified, facilitating an analysis of focus projection, and (ii) allow us to anchor the constraints on information structure with respect to the clause they belong to.

## 2 Information Structure

### 2.1 Components of Information Structure

Information structure consists of three components: focus, topic, and contrast. Focus refers to what is informatively new and/or important in the sentence (Lambrecht, 1996). *Wh*-questions have been employed as a tool to probe the focus meaning and marking: For instance, if the question is *What barks?*, the constituent corresponding to the *wh*-word in the answer bears the A-accent (H*) in English, such as *The* DOG *barks.*[1] Topic is what an utterance is about. Choi (1999) suggests the tell-me-about test for identifying topic: e.g. a reply to *Tell me about the dog.* will contain a word with the B-accent (L+H*) in English: *The* ***dog*** BARKS. Contrast (realized as either contrastive

* We are grateful to Ann Copestake and Dan Flickinger for the suggestion to explore ICONS as a means of representing information structure.

HPSG 2012 Conference/Ellipsis Workshop, July 18-21, 2012, Chungnam National University, Daejeon, Korea

[1] In this paper, SMALL CAPS stands for an A-accented phrase, **boldface** for a B-accented one, and [$_f$ ] for focus projection.

topics or contrastive foci) always entails an alternative set, which can be lexically or syntactically expressed in some languages. Several tests to vet contrast, such as the correction test (Gryllia, 2009), have also been suggested.

## 2.2 Languages

While the analysis we develop is intended to be flexible enough to work cross-linguistically, we will use English, Japanese and Russian to exemplify three common types of information structure marking. English primarily uses prosody for this function (e.g. A/B-accents (Jackendoff, 1972)). Japanese employs morphological markers: For instance, if the topic marker *wa* is attached to an NP, the NP involves either topic or contrast, or both (i.e. contrastive topic). On the other hand, if the case markers (e.g. *ga* for nominatives) are used instead of *wa*, the NP cannot fill the role of topic (Heycock, 1994). In contrast to English and Japanese, Russian takes advantage of its relatively free word order to assign a specific position to signal focus: Non-contrastive focus appears clause-finally and contrastive focus is preposed (Neeleman and Titov, 2009).

## 2.3 Differences in Felicity

Information structure affects the felicity of a sentence in different discourse contexts. Sets of allosentences (Lambrecht, 1996) differing only in information structure will differ in felicity in a given context. Multilingual NLP systems (e.g. MT) can be improved by making them sensitive to such constraints. For example, *The dog barks.* can be translated into at least two sentences in Japanese and Russian respectively. If *dog* bears the B-accent in English, the corresponding Japanese word *inu* should be combined with the topic marker *wa*, and the corresponding Russian word *sobaka* cannot occur clause-finally, as given in the first column of (1). On the other hand, if *dog* bears the A-accent, the nominative maker *ga* has to be used in Japanese, and the corresponding word can show up clause-finally in Russian, as shown in the second column of (1).

(1) a. The **dog** BARKS. | The DOG barks.

b. inu-wa hoeru | inu-ga hoeru
dog-TOP bark dog-NOM bark (jpn)

c. sobaka laet | laet sobaka
dog bark bark dog (rus)

# 3 Individual Constraints

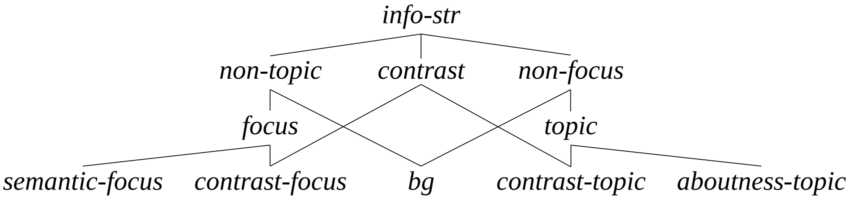
We propose to represent information structure via a feature ICONS (Individual CONstraintS) added to structures of type *mrs* (i.e. under CONT) in (2).

(2)
$$\left[\begin{array}{ll} \textit{mrs} & \\ \text{HOOK} & \left[\begin{array}{ll} \textit{hook} & \\ \text{LTOP} & \textit{handle} \\ \text{INDEX} & \textit{individual} \\ \text{XARG} & \textit{individual} \\ --\text{ICONS} & \textit{info-str} \\ --\text{CLAUSE} & \textit{event} \end{array}\right] \\ \text{RELS} & \textit{diff-list} \\ \text{HCONS} & \textit{diff-list} \\ \text{ICONS} & \left\langle !\ \ldots, \left[\begin{array}{ll} \textit{info-str} & \\ \text{CLAUSE} & \textit{individual} \\ \text{TARGET} & \textit{individual} \end{array}\right], \ldots\ ! \right\rangle \end{array}\right]$$

ICONS represents information structure as a binary relation between individuals and events. The items on the ICONS list are feature structures of type *info-str* which indicate which index (the value of TARGET) has an information structural property and with respect to which clause (the value of CLAUSE). ICONS behaves analogously to HCONS and RELS in that values of *info-str* are gathered up from daughters to mother up the tree.

In a particular ICONS element, the type will typically be resolved from *info-str* to a more specific type, drawn from the hierarchy in (3), to indicate the particular information structural role played by the TARGET in the CLAUSE. The *info-str* hierarchy is inspired by the analogous hierarchy from Song and Bender (2011), but is extended with three additional nodes: *non-topic*, *non-focus*, and *bg*: (i) *non-topic* means the target cannot be read as topic (e.g. case-marked NPs in Japanese); (ii) *non-focus* similarly indicates that the target cannot be the focus, and would be appropriate for e.g. dropped elements in pro-drop languages; (iii) finally, *bg* (background, a.k.a. tail) means the constituent is neither *focus* nor *topic*, which typically does not involve additional marking but may be forced by particular positions in a sentence.

(3)



The value of ICONS is constrained by both lexical and phrasal types. First, every lexical entry that introduces an index which can participate in information structure inherits from *icons-lex-item* (4a). This type bears the constraints which introduce an ICONS element as well as providing a pointer to the ICONS element inside the HOOK (−−ICONS), for further composition. *Icons-lex-item* also links the HOOK.INDEX to the TARGET value. On the other hand, lexical entries that cannot play a role in the information structure (e.g. semantically void lexical entries, such as case marking adpositions) inherit from *no-icons-lex-item* (4b), which provides an empty ICONS list.

(4) a.
$$\left[\begin{array}{l} \textit{icons-lex-item} \\ \text{HOOK} \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{−−ICONS} & \boxed{2} \end{array}\right] \\ \text{ICONS} \left\langle !\ \boxed{2}\left[\text{TARGET}\quad \boxed{1}\right] ! \right\rangle \end{array}\right]$$

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

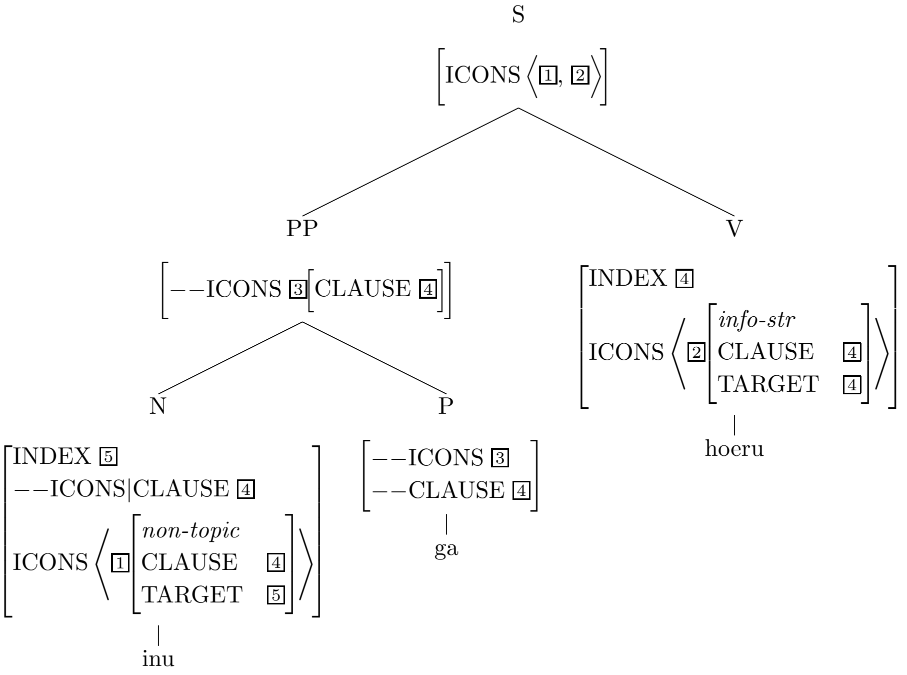
c.
$$\left[\begin{array}{l} \textit{verb-lex} \\ \text{HOOK} \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{−−CLAUSE} & \boxed{1} \\ \text{−−ICONS|CLAUSE} & \boxed{1} \end{array}\right] \end{array}\right]$$

d.
$$\left[\begin{array}{ll} \textit{head-icons-phrase} & \\ \text{HD-DTR|...|HOOK|−−CLAUSE} & \boxed{1} \\ \text{NON-HD-DTR|...|HOOK|−−ICONS|CLAUSE} & \boxed{1} \end{array}\right]$$

Because the CLAUSE value needs to reflect the position in which a constituent is realized overtly, it is constrained via the phrase structure rules. Verbs which head their own clauses (i.e., finite verbs, plus certain uses of non-finite verbs) identify their CLAUSE value with their own INDEX (and thus their own TARGET).[2] For elements that do not head clauses, the CLAUSE value is constrained to be the INDEX of the verbal projection they attach to by *head-icons-phrase* (4d). This type is supertype to headed rules which can constrain information structure: e.g. *head-subj-phrase*, *head-comp-phrase*, and *head-mod-phrase*. The type of the −−ICONS value of a constituent (which, recall, points to an element of the ICONS list) can also be constrained by lexical rules attaching information structure marking morphemes, phrase structure rules corresponding to distinguished positions, or particles like Japanese *wa* combining as heads or modifiers with NPs. The headed rules can have subtypes which handle information structure differently, resolving the type of an ICONS element or leaving it underspecified. For example, the Russian allosentences (1c) are instances of *head-subj-phrase*, but the first one (*sobaka laet*), in which the subject is in-situ, is licensed by a subtype that does not resolve the ICONS value, while the second one (*laet sobaka*), in which the subject is marked through being postposed, is licensed by the one which does. Hence, as shown in (6), the in-situ subject in Russian is specified as *info-str* (i.e. underspecified), whereas the overtly postposed subject is specified as *focus*.

The strategy of having phrase structure rules constrain the CLAUSE value of ICONS elements runs into a potential problem with *head-comp-phrase* because this rule is used in many different ways in our grammars. In particular, the problem arises with elements like Japanese case-marking adpositions: *inu-ga* ‘dog-NOM’ is an instance of *head-comp-phrase*, but *inu* has no informational structural relation with its head *ga*. On the other hand, when *head-comp* joins a verb with its object, we want to connect the object’s CLAUSE to the verb’s INDEX. Rather than creating subtypes of *head-comp* to handle this differing behavior, we add the feature −−CLAUSE to mediate between the INDEX of the head and the CLAUSE value of the dependent. The phrase structure rules identify the head’s −−CLAUSE with the non-head’s −−ICONS|CLAUSE. Clause-heading verbs identify their INDEX and −−CLAUSE values. Case marking adpositions, on the other hand, inherit from *no-icons-lex-item*, which identifies −−CLAUSE with −−ICONS|CLAUSE.[3]

(5) a.



[2] The restriction to clause-heading verbs is meant to allow for examples like *The dog sitting on the mat barks.* where we believe that all elements of the VP *sitting on the mat* should take the INDEX of *barks* as their CLAUSE, not that of *sitting*.

[3] Note, however, that the value of −−ICONS is not identified with anything on the actual ICONS list for these elements, allowing −−ICONS|CLAUSE to function as sort of a scratch slot.

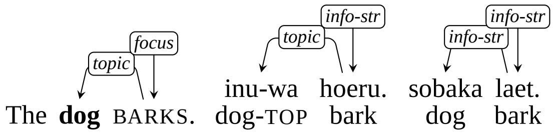
b.

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

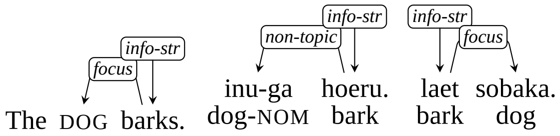
Building upon the constraints presented so far, a sample derivation for a Japanese sentence is illustrated in (5a): First, −−CLAUSE of the nominative marker *ga* is identified with its own −−ICONS|CLAUSE. Second, when the *head-comp-phrase* combines *inu* and *ga*, the −−ICONS|CLAUSE of *inu* is identified with the −−CLAUSE of *ga*, in accordance with *head-icons-phrase*. The −−ICONS of *ga* is passed up to the mother (Semantic Inheritance Principle). When the *head-subj-phrase* combines *inu-ga* and *hoeru*, the −−ICONS|CLAUSE of the subject *inu-ga* (and thus of both *inu* and *ga*) is identified with the INDEX of *hoeru*. The corresponding MRS representation is given in (5b).

In the remainder of the paper, we will present information structural constraints in the style of dependency graphs of DMRS (Dependency MRS, (Copestake, 2009)), for ease of exposition. The graphs of the translations given in (1) are sketched in (6). Unless there is a specific clue to identify information structure such as A/B-accents in English, the topic marker *wa* in Japanese, and the clause-final position in Russian, the ICONS value remains just *info-str*.

(6) a.



b.



Our approach has both similarities and differences to earlier work representing information structure in MRS. Wilcock (2005) models the scope of focus similarly to quantifier scope (i.e. HCONS), which is close to the idea that we take as our departure point for discussion. The difference between Wilcock's proposal and ours is that information structure in his model is represented as variables over handles, but ICONS captures the clause that an individual informatively belongs to as a binary relation, which facilitates scaling to multiclausal constructions. Paggio (2009) also models information structure within the MRS formalism, but information structural components in her proposal are represented as a part of the context, not the semantics. Though each component under CTXT|INFOSTR involes co-indexation with individuals in MRS, her approach cannot be directly applied to the LOGON MT infrastructure that requires all transfer-related ingredients accessible in MRS (Oepen et al., 2007). Bildhauer and Cook (2010) offer an MRS-based architecture, too: Information structure in their proposal is represented directly under SYNSEM (i.e. SYNSEM|IS) and each component (e.g. TOPIC, FOCUS) has a list of indices identified with ones that appear in EPs in SYNSEM|LOC|CONT|RELS, which is not applicable to the LOGON infrastructure for the same reason.

In the context of implementing NLP systems, using ICONS has two merits; (i) underspecifiability, and (ii) a binary relation between individuals. The former facilitates flexible, partial representations and the latter enables us to capture the various types of sentences. The following sections cover each of these points in turn.

## 4 Underspecifiability

The A-accent on DOG in (6b) can project focus to two constituents as shown in (7), which correspond to questions like *What barks?* (i.e. *focus-bg*) and *What happens?* (i.e. *all-focus*), respectively.[4]

(7) a. [$_f$ The DOG] barks.

b. [$_f$ The DOG barks.]

Regarding the interpretation of (7), we can assume that (i) the two readings correspond to two distinct structures (parse trees), or (ii) the two readings are further specializations of one MRS, which is associated with one syntactic structure and includes some underspecified values. Here, as our goal is a computational model, we take the second approach for practical reasons and underspecify the type of the ICONS element for unmarked constituents such as *barks* in (6b). Some previous work (Engdahl and Vallduví, 1996; De Kuthy, 2000; Chung et al., 2003), in contrast, takes the first approach: All sentences, within these frameworks, have as many trees as the potential interpretations, as given in (7). This approach however does not work productively in NLP systems, because a large number of trees eventually has an adverse effect on performance as well as accuracy.[5] Since it is important for transfer-based MT to reduce the number of potential analyses in each step, it is necessary to use a more effective and flexible method to represent information structure. We believe that underspecified representations can be further constrained to represent different focus spreading interpretations (consistent with the given ICONS list) in the same way that scope-underspecified MRSs can be further constrained with handle identities consistent with the given HCONS list. In the similar way that a sentence which has a scopal ambiguity (e.g. *Every dog chases some white cat.*) has a single MRS partially constrained via *qeq*, the current work assumes sets of allosentences such as (7) share the same MRS partially constrained via ICONS.

We leave the development of the algorithm that calculates focus projection over MRS+ICONS to future work. We are particularly interested to investigate whether the MRS structure augmented with ICONS is sufficient, or if the focus projection algorithm would require access to syntactic structure. We note that previous work on focus projection (De Kuthy, 2000; Chung et al., 2003) highlights the importance of grammatical functions. However, the relevant distinctions (argument *vs.* adjunct status, peripheral *vs.* non-peripheral arguments) can be reconstructed on the basis of the MRS alone. Therefore, we consider it at least plausible that MRS+ICONS will contain enough information to calculate the range of fully-specified information structures for each sentence.

[4] Heycock (1994) and Chung et al. (2003) claim whether the focus on subjects can be projected to the whole sentence or not depends on the aspectual property of the predicates (i.e. individual-level *vs.* stage-level). Exploring naturally occurring texts, however, presents quite a number of examples which the distinction between individual-level and stage-level cannot be straightforwardly applied to. Thus, it would be more feasible to leave formally unmarked constituents (e.g. *barked* in (6b)) informatively underspecified.

[5] In early work on information structure in HPSG, Kuhn (1996) also suggests an underspecified representation for information structure from the viewpoint that prosodic marking of information structure often yields ambiguous meanings, which cannot in general be resolved in computational sentence-based processing.

## 5 Multiclausal Utterances

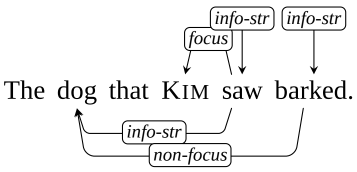
Though underspecifiability makes the obvious difference between previous work and our proposal, using underspecification itself has been already tried in our previous proposal (Song and Bender, 2011). The difference between the previous one and the current one is in the representation of the constraints: Where the previous one used features on semantic variables, the current work introduces binary relations on ICONS in order to handle information structure in multiclausal sentences within the MRS representation.

(8-9) show how the binary relation helps represent an individual that has different information structural relations to the matrix and subordinate clauses. The answer in (8), which assigns the main stress (i.e. A-accent) on a constituent inside a relative clause, can be a proper answer to only Q1. Q2 is not a contextually appropriate question from the fact that the daughters in a non-headed phrase cannot project focus to the head (Chung et al., 2003). In other words, [$_f$ *The dog that* KIM *saw*] is not a possible focus projection result because the head noun *dog* without an accent cannot inherit focus from KIM on the relative clause.[6] For the same reason, the answer sounds infelicitous in the *all-focus* context such as Q3, too. From these facts, we assume two types of focus projection as (9a), which can be illustrated as (9b) (with unmarked elements left underspecified).

(8) Q1: Which dog barked?
Q2: #What barked?
Q3: #What happened?
A: The dog that KIM saw barked.

(9) a. The dog that [$_f$ [$_f$ KIM] saw] barked.

b.



c.

$$\begin{bmatrix} \textit{mrs} & \\ \text{LTOP} & \boxed{h1} \\ \text{INDEX} & \boxed{e2} \\ \text{RELS} & \left\langle \begin{array}{l} \begin{bmatrix} \textit{\_the\_q\_rel} & \\ \text{LBL} & \boxed{h3} \\ \text{ARG0} & \boxed{x5} \\ \text{RSTR} & \boxed{h6} \\ \text{BODY} & \boxed{h4} \end{bmatrix}, \begin{bmatrix} \textit{\_dog\_n\_rel} & \\ \text{LBL} & \boxed{h7} \\ \text{ARG0} & \boxed{x5} \end{bmatrix}, \begin{bmatrix} \textit{proper\_q\_rel} & \\ \text{LBL} & \boxed{h8} \\ \text{ARG0} & \boxed{x10} \\ \text{RSTR} & \boxed{h9} \\ \text{BODY} & \boxed{h11} \end{bmatrix}, \\ \begin{bmatrix} \textit{named\_rel} & \\ \text{LBL} & \boxed{h12} \\ \text{ARG0} & \boxed{x10} \\ \text{CARG} & \textit{kim} \end{bmatrix}, \begin{bmatrix} \textit{\_see\_v\_rel} & \\ \text{LBL} & \boxed{h7} \\ \text{ARG0} & \boxed{e13} \\ \text{ARG1} & \boxed{x10} \\ \text{ARG2} & \boxed{x5} \end{bmatrix}, \begin{bmatrix} \textit{\_bark\_v\_rel} & \\ \text{LBL} & \boxed{h14} \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{x5} \end{bmatrix} \end{array} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{h6} \\ \text{LARG} & \boxed{h7} \end{bmatrix}, \begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{h9} \\ \text{LARG} & \boxed{h12} \end{bmatrix} \right\rangle \\ \text{ICONS} & \left\langle \begin{array}{l} \begin{bmatrix} \textit{focus} & \\ \text{CLAUSE} & \boxed{e13} \\ \text{TARGET} & \boxed{x10} \end{bmatrix}, \begin{bmatrix} \textit{info-str} & \\ \text{CLAUSE} & \boxed{e13} \\ \text{TARGET} & \boxed{e13} \end{bmatrix}, \begin{bmatrix} \textit{info-str} & \\ \text{CLAUSE} & \boxed{e13} \\ \text{TARGET} & \boxed{x5} \end{bmatrix}, \\ \begin{bmatrix} \textit{non-focus} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{x5} \end{bmatrix}, \begin{bmatrix} \textit{info-str} & \\ \text{CLAUSE} & \boxed{e2} \\ \text{TARGET} & \boxed{e2} \end{bmatrix} \end{array} \right\rangle \end{bmatrix}$$

[6] If *dog* also bears the A-accent, it can get focus (i.e. multiple foci: *The* DOG *that* KIM *saw barked.*), but it cannot be focused through focus projection from the adjunct (Chung et al., 2003).

What is important in (9b) is that one element *dog* has different relations with two verbs; one is *barked* in the matrix clause, and the other is *saw* in the relative clause. On the one hand, *dog* has the *non-focus* relation (i.e. either *topic* or *bg*) with the main verb *barked*, because it cannot inherit focus from the A-accent in the relative clause. On the other hand, since there is no specific clue to identify the relation between *dog* and *saw*, *dog* is specified as just *info-str* in relation to *saw*.

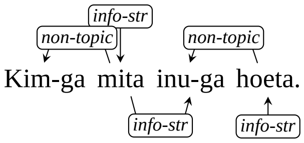
Other than the two relations, we can see three relations as well: On the one hand, KIM with the A-accent (i.e. overtly marked) has the *focus* relation with *saw* in the relative clause. On the other hand, *saw* and *barked* without any specific markings are underspecified.

## 6 A Sample Translation

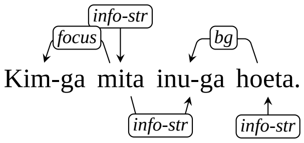
The MRS representation of (9b) is given in (9c), which is the input MRS in translating the English sentence into other languages. One of the potential translations of (9) in Japanese is given in (10a), and the information structure can be monolingually analyzed as (10b). (10a) can be generated as the translation, if (10b) is not inconsistent with (9c). The intersection between (10b) and the output MRS transferred from (9c) is sketched out in (10c). The *focus* relation between *Kim* and *mita* 'saw', which is a more specific type of *non-topic*, is taken from (9c). *Non-focus* between *dog* and *barked* in (9c) and *non-topic* between *inu* 'dog' and *hoeta* 'barked' is consistent with each other, and unified as *bg*. The others are the same with those in (10b).

(10) a. Kim-ga mita inu-ga hoeta
dog-NOM saw dog-NOM barked (jpn)

b.



c.



## 7 Summary and Outlook

This paper, in the context of multilingual MT, shows that information structure can be effectively represented within MRS via ICONS. ICONS takes as its value a list of *info-str* objects with CLAUSE and TARGET properties; the subtypes of *info-str* indicate which information structural role is played by the TARGET with respect to the CLAUSE.

Our future work includes two directions: Theoretically, it is important to understand how information structure works in more various types of embedded clauses (e.g. clefts, control constructions) as well as what kinds of embedded constituents create their own information structural domains (e.g. relative clauses *vs.* progressive participles used as modifiers). Distributionally, we plan to exploit multilingual parallel texts to learn whether ICONS can be straightforwardly applied to other languages from a cross-linguistic viewpoint.

## References

Bildhauer, Felix and Cook, Philippa. 2010. German Multiple Fronting and Expected Topichood. In Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*, pages 68–79, Stanford: CSLI Publications.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI Publications.

Chung, Chan, Kim, Jong-Bok and Sells, Peter. 2003. On the Role of Argument Structure in Focus Projections. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 39, pages 386–404, Chicago Linguistic Society.

Copestake, Ann. 2009. Slacker Semantics: Why Superficiality, Dependency and Avoidance of Commitment can be the Right Way to Go. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 1–9, Athens, Greece: Association for Computational Linguistics.

Copestake, Ann., Flickinger, Dan., Pollard, Carl. and Sag, Ivan A. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4), 281–332.

De Kuthy, Kordula. 2000. *Discontinuous NPs in German – A Case Study of the Interaction of Syntax, Semantics and Pragmatics*. CSLI publications.

Engdahl, E. and Vallduví, E. 1996. Information Packaging in HPSG. *Edinburgh Working Papers in Cognitive Science* 12, 1–32.

Gryllia, Styliani. 2009. *On the Nature of Preverbal Focus in Greek: a Theoretical and Experimental Approach*. Ph. D.thesis, Leiden University.

Heycock, Caroline. 1994. Focus Projection in Japanese. In *Proceedings North East Linguistic Society*, volume 24.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar.*. Cambridge, MA.: The MIT Press.

Kuhn, Jonas. 1996. An Underspecified HPSG Representation for Information Structure. In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, pages 670–675, Association for Computational Linguistics.

Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

Neeleman, Ad and Titov, Elena. 2009. Focus, Contrast, and Stress in Russian. *Linguistic Inquiry* 40(3), 514–524.

Oepen, Stephan, Velldal, Erik, Lønning, Jan T., Meurer, Paul, Rosén, Victoria and Flickinger, Dan. 2007. Towards Hybrid Quality-Oriented Machine Translation – On linguistics and probabilities in MT –. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation*, Skövde, Sweden.

Paggio, Patrizia. 2009. The Information Structure of Danish Grammar Constructions. *Nordic Journal of Linguistics* 32(01), 137–164.

Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: The University of Chicago Press.

Song, Sanghoun and Bender, Emily M. 2011. Using Information Structure to Improve Transfer-based MT. In Stefan Müller (ed.), *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, pages 348–368, Stanford: CSLI Publications.

Wilcock, Graham. 2005. Information Structure and Minimal Recursion Semantics. *Inquiries into Words, Constraints and Contexts: Festschrift for Kimmo Koskenniemi on his 60th Birthday* pages 268–277.