

Using Information Structure to Improve Transfer-based MT

Sanghoun Song and Emily M. Bender

Department of Linguistics, University of Washington

1 Introduction

This paper hypothesizes that transfer-based machine translation (MT) systems can be improved by encoding Information Structure (henceforth, IS) in both the source and target grammars, and preserving IS in the transfer stage. This paper delves into how IS can be represented within the HPSG/MRS formalism (Pollard and Sag, 1994; Copestake et al., 2005), and how it can help refine multilingual MT. Building upon that framework, this paper also provides a sample translation between English and Japanese, and checks out the feasibility in small-scale translation systems built with the HPSG/MRS-based LOGON MT infrastructure (Oepen et al., 2007).

2 Puzzle

In the context of MT, we find that allosentences – close paraphrases which share truth conditions (Lambrecht, 1996) – are not always felicitous as translations of the same inputs. For example, a simple English sentence (1a) can be translated into at least two Japanese allosentences such as (1b) (i.e. with the nominative marker *ga* or with the topic marker *wa*).

- (1) a. I am Kim. (English)
b. *watashi-ga/wa Kim desu.*
I-NOM/TOP Kim COP (Japanese)

However, the selection between them is conditioned by the given context; the NP marking hinges on whether *watashi* ‘I’ functions as the topic or not. If the sentence is an answer to a question like ‘Who are you?’, the topic marker *wa* is much more preferred. In contrast, if the sentence replies to a question like ‘Who is Kim?’, the topic marker seldom appears.

Another example which is deeply concerned with translational equivalents can be found in active/passive pairs. Consider the Japanese sentences in (3), which are translations of the English sentences in (2). The active sentence (3a) is just fine, but the passive sentence (3b) sounds like a clumsy translation, as inanimate nouns tend not to appear in subject position of passive clauses in Japanese. That is, passives in one language cannot always be translated into passives in another; the active sentence (3a) is one of the potentially legitimate translation of the English passive one (2b), while the passive one (3b) is not.

- (2) a. Kim tore the book.
b. The book was torn by Kim. (English)
- (3) a. *Kim-ga sono hon-o yabut-ta.*
Kim-NOM DET book-ACC tear-PST
‘Kim tore the book.’
b. *?sono hon-ga Kim-ni yabu-rare-ta.*
DET book-NOM Kim-DAT tear-PASS-PST
‘The book was torn by Kim.’ (Japanese)

As presented in (1), case makers (e.g. *ga* for nominatives and *o* for accusatives) in Japanese are in complementary distribution with the topic marker *wa*. In addition, scrambling (OSV order) is highly productive in Japanese (Ishihara, 2001). Hence, (3a) has at least eight allosentences ($2 \times 2 \times 2$) as given in (4).¹

Though these eight sentences presumably share the same truth conditions, they are not felicitous in the same contexts.

- (4) a. *Kim-ga/wa sono hon-o/wa*
Kim-NOM/TOP DET book-ACC/TOP
yabut-ta.
tear-PST
b. *sono hon-o/wa Kim-ga/wa*
DET book-ACC/TOP Kim-NOM/TOP
yabut-ta.
tear-PST

The difference in felicity conditions between allosentences is the subject of study of IS. Thus, in order to solve the mismatch and refine the translations, this paper turns to IS. IS is hypothesized to be

¹An anonymous reviewer took notice of two facts regarding these allosentences. First, the so-called double *wa* construction, in which the topic marker *wa* attaches to both the subject and the object, rarely occur in Japanese. But, it is also true the double *wa* construction is not illegitimate in Japanese, though the productivity is rather low. We assume that the first *wa*-marked NP in a sentence is the topic of the sentence, and the second *wa*-marked NP conveys the meaning of contrastive-focus. Second, since Japanese belongs to the so-called pro-drop languages, we can consider one more option; namely topic-drop. That is, *Kim* and *sono hon* ‘the book can be freely dropped, depending on the given context. Moreover, since NP markers (e.g. *ga* and *wa*) are optionally used in Japanese, we have at least 32 allosentences, in total. But, this paper, which basically aims to verify whether or not IS can improve performance of transfer-based MT with a small-scale experiment, tentatively ignores the last two options.

universal: All languages have some way to mark topics and foci, such as with pitch accent, word order, morphological marking or some combination of these (Gundel, 1999), though the marking is not necessarily unambiguous. The universality of IS suggests that it should transfer well, and that it in turn can help facilitate transfer when the syntactic structures in the source and target languages diverge.

In order to check out the computational feasibility of this proposal, this paper looks at the particular case of translating English passive sentences into Japanese and Korean, such as (2–3), in that active/passive pairs can yield relatively larger numbers of allosentences.

3 IS in HPSG/MRS

Because assignment of IS categories to referents can be constrained by both lexical marking (morphological or prosodic) and phrase-structural configurations, we analyze IS in terms of three levels of structure: a semantic feature INFO-STR reflecting the information in the MRS (§3.1), a syntactic feature MKD encoding the lexical marking (§3.2), and a set of constraints on phrase structure rules relating the two (§3.3).

Our analysis builds on the following assumptions: First, while sentences always have at least one focus, they do not always have a topic (Gundel, 1999); further, constituents may be ‘background’ (i.e. neither topic nor focus). Second, we treat ‘contrast’ as one of the cross-cutting IS categories, which contributes the entailment of an alternative set (Molnár, 2002).² Third, semantically empty categories (e.g. complementizers, expletives) and syncategorematic items (e.g. relative pronouns) are informatively empty as well (i.e. assigned no IS category).

- (5) a. It is Kim who tore the book.
 b. The book was torn by Kim.

For example, in (5a), an expletive ‘it’ and a copula ‘is’ are semantically empty and a relative pronoun ‘who’ is syncategorematic; thus, they are informatively vacuous. Likewise, since the preposition ‘by’ in English passive sentences is assumed to be semantically void, it cannot take part in IS, as shown in

²That means there are four components that participate in IS; topic, focus, contrast, and background. Lambrecht (1996) regards ‘contrastiveness’ as a merely cognitive concept, yet there are several cross-linguistic counterexamples to his claim; some languages employ specific markers or syntactic ways to express contrastiveness. For example, Vietnamese uses a contrastive-topic marker *thì* which is distinctive from the regular topic (i.e. *aboutness-topic* in this paper) (Lee, 2007). In Portuguese, similarly, contrastive focus precedes the verb, whereas non-contrastive focus (i.e. *semantic-focus* in this paper) follows the verb (Ambar, 1999).

(5b). Finally, we assume the canonical position of topics is sentence-initial in our sample of languages (English, Japanese, and Korean), though this generalization does not hold for all languages (Erteschik-Shir, 2007).

3.1 MRS: *info-str*

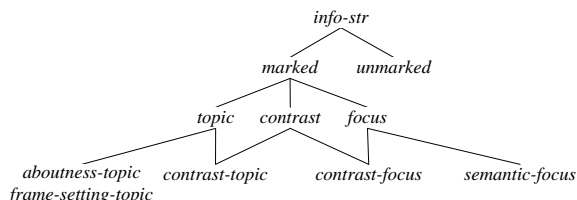


Figure 1: Type Hierarchy of *info-str*

Although IS is strictly speaking pragmatic rather than semantic, we represent it in our MRS semantic representations. Our motivation for doing so is primarily practical: The MT infrastructure we are using (Oepen et al., 2007) does MRS-based transfer. Therefore, any information that we wish to have access to in transfer must be encoded in the MRS. Because we want to consolidate information accessible in transfer in the MRS, we encode the IS in the semantics (MRS), rather than in the context, contra (Engdahl and Vallduví, 1996; Paggio, 2009). Like Paggio, we associate IS with semantic indices; however, while Paggio has IS-related lists in the CONTEXT structure taking indices as their elements, we represent IS with a feature on indices directly in the MRS. This feature (INFO-STR) draws its values from the hierarchy in Figure 1.

3.2 Markedness: *mkd*

The lexical marking itself is recorded via a syntactic feature MKD, inside of CAT. MKD has two subfeatures, TP and FC, which can be constrained independently.³

$$(6) \left[\text{MKD} \begin{bmatrix} \text{TP} & \text{bool} \\ \text{FC} & \text{bool} \end{bmatrix} \right]$$

The value of MKD is always a subtype of *mkd*, drawn from the hierarchy in Figure 2 (*tp* is constrained to be [TP +], *non-tp* [TP –], *fc* [FC +], and *non-fc* [FC –]).⁴

³What we expect in our future work is that *mkd* could in principle be used in modeling focus projection, in the sense that foci can be classified into narrow focus and wide focus (a.k.a. argument focus vs. predicate focus).

⁴For example, the Vietnamese contrastive-topic marker *thì* (Lee, 2007) has the *tp-only* because it cannot be used to

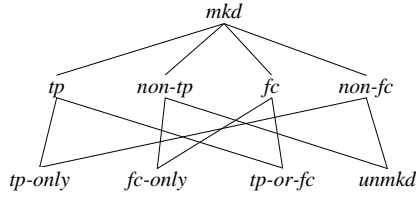


Figure 2: Type Hierarchy of *mkd*

The MKD value reflects the morphological marking but does not necessarily match with the actual INFO-STR value because in some languages the constructions assign the INFO-STR, taking into account both the MKD value of the daughters and construction-specific constraints on their order. For instance, the topic markers *wa* in Japanese and *(n)un* in Korean can involve a focus reading if the topic-marked NP is scrambled as shown in (4b), which will be explained in detail in (§4.2).

3.3 Sentential Forms: *sform*

Building on (Lambrecht, 1996; Engdahl and Vallduví, 1996; Paggio, 2009), we propose the classification of phrase types in Figure 3. *topicality* is mainly concerned with how the topic is realized in a sentence. In *topic-comment* constructions, topics are followed by other constituents. As noted, not all sentences have topics. We provide for this with the type *topicless* (e.g. cleft sentences in English such as (5a)). *focality* is divided into *narrow-focus* and *wide-focus*, because some languages have different ways to express these (e.g. Somali (Lecarme, 1999)).

Several of these sentence types are illustrated in English allosentences (7), where we have added annotations disambiguating the IS: SMALL CAPS indicate prosodically marked foci, **boldface** indicates prosodically marked topics, and [_f] indicates the extent of focus projection.

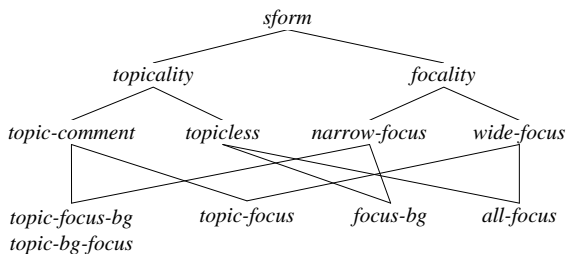


Figure 3: Type Hierarchy of *sform*

- (7) a. **The book** was torn by [_f KIM].

express foci.

- b. **The book** [_f was torn] by Kim.
 c. **The book** [_f was torn by Kim].
 d. [_f THE BOOK] was torn by Kim.
 e. [_f The book was torn by Kim].

(7a) is encoded as *topic-bg-focus*. (7b-c), with predicate foci (i.e. *wide-focus*), are *topic-focus-bg* and *topic-focus*, respectively. (7d-e) are *topicless*; *focus-bg* and *all-focus*, respectively.

If INFO-STR is lexically determined like (7a) in which INFO-STR is prosodically or lexically conditioned, SFORM can be easily detected as well. For example, the ‘as for’ construction in English, such as (8), belongs to *topic-comment* because the (near) lexical expression ‘as for’ which has the *tp-only* (i.e. [TP +, FC -]) forms (*contrastive*)-*topic*, and the NP precedes the comment; (8) is encoded as *topic-focus-bg*.

- (8) As for **the book**, KIM tore it.

However, since the Japanese marker *wa* itself is informationally ambiguous, the syntactic configuration is required to determine SFORM as well as INFO-STR of each sentence in (4), as discussed in the next section.

4 IS in English and Japanese/Korean

4.1 English

In English, IS is normally constrained by pitch accents (Jackendoff, 1972)⁵; thus, an A-accented phrase (H*) takes *focus*, while a B-accented phrase (L+H*) takes *topic*, as presented in (9). As for contrast in English, its prosodic marking is partially similar to both A/B-accent (Hedberg and Sosa, 2007). As a result, both accents can be interpreted as contrast, in an appropriate context. Therefore, we assign the INFO-STR values *topic* and *focus*, which are compatible with the more specific *contrast-topic* and *contrast-focus*.

$$(9) \begin{array}{l} \text{fp-lex-rule} \rightarrow \\ \left[\begin{array}{l} \text{PROSODY A-accent} \\ \text{INFO-STR focus} \end{array} \right] \end{array} \quad \begin{array}{l} \text{tp-lex-rule} \rightarrow \\ \left[\begin{array}{l} \text{PROSODY B-accent} \\ \text{INFO-STR topic} \end{array} \right] \end{array}$$

In the context of our text-based MT, this property might be problematic, because written English does not explicitly mark IS. However, IS categories presumably could be added to an English input sentence as a preprocessing step, either on the basis of prosodic analysis in a speech-based system or on the basis of a classifier which takes extra- as well as intra-sentential

⁵We are not considering the pitch accents directly in this study.

context into account. For present purposes, we represent these patterns with typeface variations in this paper. In the evaluation process of this study, since English does not orthographically mark prosody, we tentatively made use of hypothetical suffixes ‘-TP’, ‘-FP’, which represent B-accent for topics, and A-accent for foci respectively. For instance, (7a) is entered into our system as ‘The book-TP was torn by Kim-FP’.⁶

4.2 Japanese/Korean

Japanese and Korean employ topic markers (*wa* and (*n*)*un*, respectively), and often allow scrambling (e.g. (4b)) which contributes to IS in these languages (Choi, 1999; Ishihara, 2001). From these facts, (4) illustrates the range of possible translations in Japanese corresponding to the English passive sentence (2b).

The topic markers in Japanese and Korean can also be used to denote contrastiveness. Choi differentiates contrasts from non-contrastive foci and topics in Korean. First, contrasts can freely scramble, while non-contrastive foci (a.k.a. *semantic-focus* (Gundel, 1999)) cannot. Second, when (*n*)*un* attaches to the in situ (i.e. non-scrambled) subject, the subject can be either topic or contrast. On the other hand, when (*n*)*un* attaches in situ non-subjects (e.g. objects), such constituents have only the contrastive reading. Since scrambling and IS interact the same way in Japanese as in Korean, we can create generalizations about the interaction between scrambling and IS in Korean and Japanese as follows: First, we mark the constituents with *wa* and (*n*)*un* themselves as *tp* (i.e. [TP +, FC bool]). Second, we note the three possible interpretations of a *tp*-marked NP, depending on its syntactic function and position, shown in Table 1. Third, *wa* or (*n*)*un* cannot occur in *all-focus* constructions, which allow only *semantic-focus* lacking contrastive meanings. Finally, the constituents marked with *wa* or (*n*)*un* are necessarily marked for some IS category, i.e. cannot simply be ‘background’.

Table 1: IS of topic-marked NP

	in-situ	scrambling
subject	<i>topic</i>	<i>contrast-focus</i>
non-subject	<i>contrast-focus</i>	<i>contrast-topic</i>

In short, as presented in Table 1, *contrast-topic* and *contrast-focus* play a role in Japanese and Korean, be-

⁶English also uses syntactic patterns to mark information structure, notably clefts, English focus movement, and *as for*. As these are much less pervasive than prosodic marking of information structure in English (and morphosyntactic marking in Japanese and Korean), we leave the integration of these into our English grammar fragment for future work.

cause NPs with topic markers in these languages can have more than one reading by itself. Thus, the topic markers mark only *tp* on MKD as represented in (10).

$$(10) \quad \begin{array}{c} \text{nom-marker} \rightarrow \\ \left[\begin{array}{l} \text{ORTH} \langle \text{ga} \rangle \\ \text{MKD } \textit{unmkd} \\ \text{CASE } \textit{nom} \end{array} \right] \end{array} \quad \begin{array}{c} \text{topic-marker} \rightarrow \\ \left[\begin{array}{l} \text{ORTH} \langle \textit{wa} \rangle \\ \text{MKD } \textit{tp} \\ \text{CASE } \textit{case} \end{array} \right] \end{array}$$

The value of MKD is mapped to values of INFO-STR via the constraints on the various *sform* types. *topic-comment* requires *tp* of non-head-daughter such that only NPs with topic markers can participate in *topic-comment*.

$$(11) \quad \left[\begin{array}{l} \textit{topic-comment} \\ \text{MKD} \quad \textit{tp} \\ \text{HD} | \text{MKD} \quad \textit{fc} \\ \text{NON-HD} | \text{MKD} \quad \textit{tp} \end{array} \right]$$

In this way, INFO-STR in Japanese and Korean is specified at the phrasal level (i.e. each grammatical rule, such as *subj-head* and *comp-head*) unlike English. For example, the phrase structure rule building the node combining the subject and the verb for (4b) (attaching *Kim-wa* ‘Kim-TOP’ to the rest of the sentence) is an instance of *top-scr-subj-head*, which inherits from both *subj-head* and *topic-comment*. Since the subject *Kim-wa* in (4b) is not in-situ, the construction constrains the INFO-STR as *contrast-focus*. On the other hand, NPs with nominative markers (e.g., *Kim-ga* in (4b)) can be interpreted as neither topic nor contrast, because the non-head-daughter of *topic-comment* is incompatible with [TP –] as given in (11). In a similar vein, *sono hon-wa* ‘DET book-TOP’ in (4b) is a scrambled complement; it is licensed by (13) which inherits from both *comp-head* and *topic-comment*. Its INFO-STR is *contrast-topic*, because it is interpreted as both contrast and topic.

$$(12) \quad \left[\begin{array}{l} \textit{top-scr-subj-head} \\ \text{HD} | \text{VAL} | \text{COMPS} \langle \square \rangle \\ \text{NON-HD} | \text{INFO-STR } \textit{contrast-focus} \end{array} \right]$$

5 Translation

5.1 A Sample Translation

The sample sentence we try to translate is (7a), which has at least eight potential translations in Japanese as given in (4). Since ‘Kim’ is focused in the input, if the Japanese translation of this NP includes *wa*, it

$$(13) \left[\begin{array}{l} \text{top-scr-comp-head} \\ \text{HD | VAL | COMPS } \langle \rangle \\ \text{NON-HD | INFO-STR } \textit{contrast-topic} \end{array} \right]$$

must be the *contrast-focus* use of *wa*, and not *topic* or *contrast-topic*. However, as ‘Kim’ in (7a) is semantically focused in the given context (i.e. *semantic-focus* lacking a contrastive meaning), sentences in which *wa* attaches to ‘Kim’ are naturally excluded in the target language, Japanese. In addition, since ‘the book’ is the topic, and topics in Japanese must occur sentence-initially, only the scrambled variant (14) is the felicitous translation directly corresponding to (7a).

(14) sono hon-wa Kim-ga yabut-ta.

sono hon-wa in (14) has two interpretations *per se*; either regular topic (e.g. *aboutness-topic* (Krifka, 2008)) or contrast. But when it combines with the VP *Kim-ga yabut-ta* at the phrasal level, its INFO-STR is specified as *contrast-topic* by the constraint on (13). The input MRS in Figure 4 specifies [INFO-STR *focus*] on the index of *named_rel* for ‘Kim’, and this is preserved in the mapping to the MRS in Figure 5. Accordingly, the structure generated by the Japanese grammar with the MRS in Fig 5 as input is all compatible with this constraint. That is, our grammar only gives one output for the input MRS given in Figure 4. In particular, we get the *wa*-marked NP in the position where it receives *contrast-topic* as their INFO-STR (i.e. non-subject and scrambling in Table 1) and *ga*-marked NPs as the realization of *named_rel* for ‘Kim’. One such derivation is shown in Figure 6.

5.2 Evaluation: Translating Passives

In our system, IS serves to help filter unsuitable strings from the output; consequently, we can obtain more felicitous translations corresponding to the input sentences.

This study has implemented the proposals in tdl (type description language), the high-level language interpreted by the LKB (Copestake, 2002). The first step is to construct toy grammars for English, Japanese, and Korean, using the Grammar Matrix customization system (Bender et al., 2010). As a second step, other rules to produce allosentences (e.g. actives/passives) are added to each starter grammar. The third step is to implement IS into each grammar, as given earlier. Finally, we create the mapping between internal and external features of indices (*semi.vpm*), in accordance with the MT infrastructure of LOGON (Oepen et al., 2007).⁷

⁷In this study, we avoided the need for transfer rules by

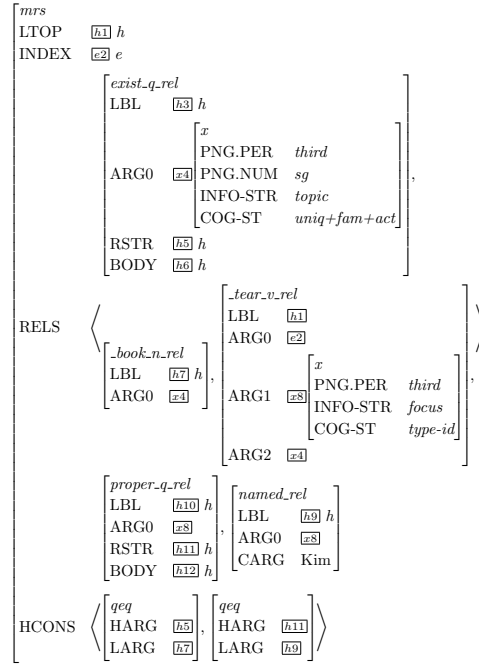


Figure 4: Input MRS (English)

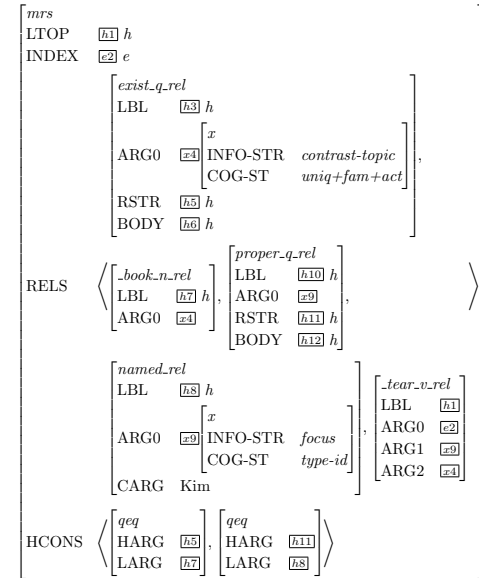


Figure 5: Output MRS (Japanese)

For our experiment, we made use of 24 input sentences in English; eight types of allosentences for each of the three verbal types: ‘tear’, ‘chase’, and ‘hit’. The first one takes inanimate nouns as complements, and thus resists passivization in Japanese and Korean. The second one tends to be freely passivized. The third one does not have passive forms in Korean, whereas it is often passivized in Japanese.

Our experiment shows our IS-based system, com- using pseudo-interlingual predicate names.

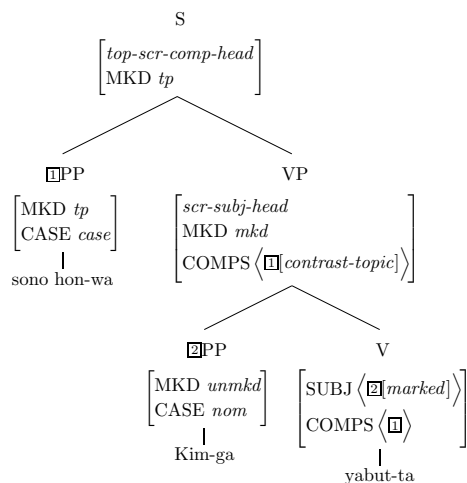


Figure 6: A Sample Derivation in Japanese

paring to the baseline that lets all of potential translations through (without filtering for IS), filters out 288 unsuitable outputs in Japanese and 295 unsuitable outputs in Korean. Consequently, we can reduce the numbers of outputs from 350 to 62 (82.29% reduction) for Japanese, from 344 to 49 (85.76% reduction) for Korean, leaving only more felicitous translations. Thus our IS-based MT system has reduced the number of translations dramatically, which has two obvious effects on the performance of transfer-based MT: First, the processing burden of MT component which ranks the translations and select only suitable results can be greatly lightened, which should improve translation speed. Second, though it is still necessary to harness a re-ranking model for choosing translations, we can start from once-refined sets of translations, which should improve translation accuracy.

6 Conclusion

So far, we have covered how we represent IS within the HPSG/MRS framework and how it is used to refine translations, especially focusing on translating English passives. The implications of this study are as follows: One the one hand, since the type hierarchies for IS that this paper proposes are constructed almost language-independently, we are optimistic that they will apply to other languages pairs as well. On the other hand, by enriching our semantic representations with IS, we effectively move further up the MT pyramid (Vauquois, 1968), reducing the burden on the transfer component. We suggest that IS can play a similar role in other cases of structural mismatch in transfer-based MT.

Our future work includes the following: First, we

plan to evaluate our IS-based system with various types of sentences, such as clefting, topicalized sentences, and topic-drop sentences. Second, other language pairs also need to be covered in order to check out the feasibility of this proposal. In particular, MT from Japanese/Korean to English has to be examined in the sense that Japanese/Korean employ more specific IS than English in our proposal. Finally, we plan to build up a library of IS analyses for the Grammar Matrix customization system (Bender et al., 2010), which contains the main proposal of this paper.

References

- Ambar, Manuela. 1999. Aspects of the Syntax of Focus in Portuguese. In Rebuschi, Georges and Laurice Tuller, editors, *The Grammar of Focus*, pages 23–54. John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation*, 8(1):23–72. 10.1007/s11168-010-9070-1.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. CSLI Publications, Stanford, CA.
- Copestake, Ann., Dan. Flickinger, Carl. Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation*, 3(4):281–332.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA.
- Engdahl, Elisabet and Enric Vallduví. 1996. Information Packaging in HPSG. *Edinburgh Working Papers in Cognitive Science*, 12:1–32.
- Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford University Press, USA.
- Gundel, Jeanette K. 1999. On Different Kinds of Focus. In Bosch, Peter and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 293–305. Cambridge University Press, Cambridge.
- Hedberg, Nancy and Juan M. Sosa. 2007. The Prosody of Topic and Focus in Spontaneous English Dialogue. In Lee, Chungmin, Matthew Gordon, and Daniel Bring, editors, *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, pages 101–120. Kluwer Academic Publishers, Dordrecht.

- Ishihara, Shinichiro. 2001. Stress, Focus, and Scrambling in Japanese. *MIT Working Papers in Linguistics*, 39:142–175.
- Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. The MIT Press, Cambridge, MA.
- Krifka, Manfred. 2008. Basic Notions of Information Structure. *Acta Linguistica Hungarica*, 55(3):243–276.
- Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, UK.
- Lecarme, Jacqueline. 1999. Focus in Somali. In Rebuschi, Georges and Laurice Tuller, editors, *The Grammar of Focus*, pages 1–22. John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Lee, Chungmin. 2007. Contrastive (Predicate) Topic, Intonation, and Scalar Meanings. In Lee, Chungmin, Matthew Gordon, and Daniel Bring, editors, *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, pages 151–175. Kluwer Academic Publishers, Dordrecht.
- Molnár, Valéria. 2002. Contrast – from a Contrastive Perspective. In Hasselgrd, H., S. Johansson, B. Behrens, and C. Fabricius-Hansen, editors, *Information Structure in a Cross-Linguistic Perspective*, pages 147–162. Rodopi, Amsterdam, Netherland.
- Open, Stephan, Erik Velldal, Jan T. Lønning, Paul Meurer, Victoria Rosén, and Dan Flickinger. 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*, Skvde, Sweden.
- Paggio, Patrizia. 2009. The Information Structure of Danish Grammar Constructions. *Nordic Journal of Linguistics*, 32(01):137–164.
- Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago, IL.
- Vauquois, Bernard. 1968. A survey of formal grammars and algorithms for recognition and transformation in mechanical translation. In *IFIP Congress (2)*, pages 1114–1122.