

2B-8

GENERATION OF KOREAN FROM CONCEPTUAL REPRESENTATION

Cheol Ho KIM^{*} Byung Hong CHOE^{*} Gil Chang KIM^{*} Key-Sun CHOI^{**} Shunji ICHIYAMA^{**}

(*Korea Advanced Institute of Science and Technology, **NEC Corp.)

1. INTRODUCTION

This paper describes a current state of the Korean generation system which has been developed within a NEC/KAIST MT project. The generator synthesizes Korean sentences from conceptual representations produced from Japanese or English texts by PIVOT system [1,3,4,5]. Although Korean and Japanese have many similarities on language typology, there are various grammatical and stylistic differences between them. To cope with these differences, several techniques are introduced : sentence-structure planning, selection of surface case-marking particles, processing of irregular conjugation, allomorphs, hiatus, and so on.

2. SENTENCE-STRUCTURE PLANNING

A major difficulty in generating good styled expressions from conceptual representations is an absence of surface-level information on the target sentence-structure, so the sentence-structure planning is important to synthesis of Korean-specific natural expressions.

- Pragmatic and stylistic decision on target voice:

The voice generation is significant to generation of a natural target sentence retaining information structure of a source sentence. But there are several difficulties on the voice generation due to complex passive-verb system and heavy passive constraints in Korean. To cope with the difficulties, scope of passives is extended to notional passives like activo-passives which

are morpho-syntactically active but semantically passive, and information on exceptions to passivization is encoded into each lexical item. Autonomous passives are generated directly from dictionary, not through passivization. The decision on the target voice depends on both pragmatic factor (topicality) and stylistic factor.

- Stylistic changes of sentence-structures:

Each language has its own style of expressions, so the sentence-structure planning for good styled expressions has to be performed, which results in a structural/lexical changes of sentence-structures,

- for activo-passivization,
- for predicative use of quantitative adjectives,
- for denominal verbs, and so on.

- Replacement of function word with content word:

In order to eliminate meaning ambiguities of function word, the planning procedure replaces the function word with a content word having supplementary meaning.

3. SYNTACTIC GENERATION

From the Korean-specific meaning structure produced by previous phase, the syntactic generator obtains a grammatical structure which contains syntactic and word-ordering information for later morphological synthesis.

GENERATION OF KOREAN FROM CONCEPTUAL REPRESENTATION

Cheol Ho KIM¹ Byung Hong CHOE¹ Gil Chang KIM¹ Key-Sun CHOI² Shunji ICHIYAMA²¹Computer Science Dept., Korea Advanced Institute of Science and Technology²C&C Information Research Laboratories, NEC Corp.

- Processing Korean case-marking particles:

One of primary functions is the decision of case-marking particles for lexical realization of deep cases. There are seven representative case-marking particles in Korean, with which obligatory deep cases are associated in case frames [2]. So the surface case-marking particles for the obligatory deep cases are determined by accessing the case frame, but those for optional deep cases are determined by a optional case table. An use of Korean genitive case-marking particle "uy", which corresponds to Japanese "no", is slightly different from that of Japanese ones. That is, the use of the genitive case-marker may be obligatory, optional, or prohibitive. When optional or prohibitive, the syntactic generator determines null case-marker for the deep cases.

4. MORPHOLOGICAL GENERATION

From the grammatical node produced by the syntactic generation, the Korean morphological generation obtains the ordering and syntactic/semantic information for morpheme generation, and then performs the following steps.

- Picking out surface morphemes:

The morphological generator makes use of rules and tables in order to pick out proper morphemes for each grammatical node. While checking the connectivity matrix table, the selected morphemes are individuated and augmented with the connectivity property for correct word/phrase formation. Since Korean is an agglutinative language with diverse affixes/word-endings, words are formed by agglutinating word-endings to word-stems. Thus the detail information is required for selecting appropriate affixes during morphological generation. In case of auxiliary declinable-words, the clues to morpheme selection are provided from the syntactic generation phase.

Therefore the morphological generator merely decides on the order of auxiliary-morphemes using the auxiliary features.

- Word/phrase formation and word boundaries:

After picking out suitable morphemes, the generator handles irregular conjugations and phonological phenomena to form words/phrases, and makes an explicit word boundaries. The phonological problems such as vowel harmony, allomorph, and hiatus are handled by the generator which checks the connectability using connectivity matrix table and examines directly neighboring morphemes.

5. CONCLUSION

The Korean generation system is based on the PIVOT framework [4,5], but Korean has many grammatical and expressional differences from Japanese. So this paper describes some techniques for coping with the gaps between Korean and Japanese. Especially Korean has plenty of inflectional suffixes for predicates, so further research for selecting the most suitable suffix among them is to be proceeded.

REFERENCES

- [1] Muraki, K.(1986) "VENUS: Two-phase Machine Translation System", In Future Generation Computer Systems, 5.
- [2] Kim, J. et al.(1987) "The Construction of Korean Case Structure and Generation of Postposition from Internal Meaning Representation in PIVOT Machine Translation", In '87 Fall Conference of Korean Information Science Society, Seoul.
- [3] 村木、市山:「PIVOT:多言語間翻訳システム概要」、63年秋季全大。
- [4] 末松、福持:「PIVOT J-E:英語生成の概要」、63年秋季全大。
- [5] 壁谷、波多野、依田:「PIVOT E-J:日本語生成の概要」、63年秋季全大。