Towards Implementing Quantification and Other Constructions <Outline>

Kiyong Lee
Korea University
klee@korea.ac.kr

1 Introduction

I believe syntax must be kept as simple as possible, although the task of understanding natural language communication is complex. Fancy structural descriptions would not be of any use unless they can provide explicit information to the interpretation of natural language expressions thus analyzed. The purpose of this presentation is thus to show in computational ways how a simple syntax can be constructed to produce complex semantic reprsentations. For demonstration, a rule-based system will be built to analyze some basic quantified sentences in English (and some other more complex constructions in Korean) and then to produce their appropriate semantic representations in automatic computational ways.

2 Basic Assumptions

As part of Hausser's (2002) SLIM theory, I make two basic assumptions for this work. One is the principle of linearity or left-associative combinatorial operations and the other, the principle of syntactic minimality that provides the minimal amount of data structure for semantic interpretation subsuming the notion of under-specification.

A corollary to the first principle is the notion of continuation, opposed to the notion of substitution or that of phrase structure. Accordingly, the left-associative analysis (((John may) have)left) is acceptable under the notion of continuation without requiring the constituent analysis (John (may (have left)).[1]

Syntactic minimality prohibits unnecessarily complex syntactic analyses. This is especially the case, since most of the linguistic information can be encoded into each lexical item or morpho-syntactic unit. The information about complement or argument structures is, for instance, part of such lexical encoding and, on its basis, the well-formedness of each sentence is checked. Hence, syntax requires a checking mechanism or constraint dealing with each piece of relevant information.

Under these assumptions, a deconstructional approach is adopted for designing semantic representations that ultimately result from syntactic analysis. By this approach, lexical bases and syntactic rules are formulated in view of pre-designed semantic representations. Consider the scope ambiguity of a sentence "every student didn't pass" that results from the two possible overscope relations between the universal quantifier and the negation, OVERSCOPEUQ, neg. The content of each lexical entry, for instance, is filled in by examining how it contributes to the

[1] This analysis may be acceptable if "may have left" is treated as a complex word form or some other type of a linguistic unit at some level, say a morpho-syntactic level of analysis. In a non-analytic langauge like Korea, an equivalent expresssion "tten-ass-ul kes-i-ta" may be treated as a single morpho-syntactic unit.

ultimate construction of an adequate semantic representation of an input string that undergoes various combinatorial steps of syntactic analysis.

Semantic structures as well as each lexical entry are represented in feature structure. Feature structures and their combinatorial rules are described by an implementation language called MALAGA. The interpretation rules for semantic representations are left untreated for future work.

3 Theoretical Framework

The two assumptions made here are part of the metalinguistic theory called the SLIM theory proposed by Hausser (2002). The present work adopts Left-Associative Grammar as a basic grammatical framework.

The lexicon consists of two parts, (1) a base component that lists lexemes with minimally necessary information represented in feature structure format and (2) a derived component that lists allomorphic forms as well as derivational word forms with fuller information again in feature structure format.

The syntactic rules consist of an initial rule, combinatorial rules, subrules, and a final rule. The initial rule provides the basic frame for each input feature structure by specifying the types of each feature or attribute. The attribute ARG-ST, for instance, is specified as taking as value a list of feature structures, while the attributes PHON and CAT take atomic values. Combinatorial rules combine two input feature structures constrained by subrules. The number agreement between a Subject and its verb in English should be treated by a subrule for agreement. The final rule is a cleaning rule that retains only those pieces of information necessary, say for semantic interpretation.

The option of rule application is constrained by a specified rule package. The subject of a sentence may simply consist of one proper name or begin with a determiner or quantifier. Hence to deal with the introduction of a subject, there is a choice of either taking a rule that introduces names or a rule that introduces a determiner. Accordingly, the preceding rule, say the initial rule, is marked with a rule package consisting at least of a Name Rule and a Det Rule. The Name Rule may then immediately proceed to a Verb Rule, while the Det Rule may have to undergo a Noun Rule that introduces (common) nouns or an ADJ rule that introduces adjectives.

4 Building Systems

Two rule systems, QuiE.pro and KorSyn6.7.pro, are introduced here. They are all implemented in a C-based programming language with debugging for developing natural language grammars, called MALAGA.

4.1 The System QuiE.pro for English

This system is a simple set of syntatic rules with a tiny lexicon quickly improvised for today's demonstration. No morphology has been incorporated in it.

4.1.1 Lexicon

Each lexical entry is represented in feature structure. Here are a few examples:

```
     #names
 [PHON: "Mia",
  CAT: name,
  SEM: [REF: <named_as,"Mia">,
        TyObj: human],
  Weight: 1];
    #quantifiers
 [PHON: "every",
  CAT: quantifier,
  SEM: [REL: <"every">,
        QType: universal,
        VAR: <>,
        RESTR: <>],
  Weight: 1];
    #common nouns
 [PHON: "man",
  CAT: noun,
  SEM: [REL: <property,"man">,
        TyObj: human,
    ARGstr:<[par_index: 1,
             Role: characterized_as]>],
  Weight: 1];
    #intransitive verbs
 [PHON: "walks",
  CAT: verb,
  SEM: [REL: <"walking">,
    TyObj: action,
        ARGstr:<[par_index: 1,
                 Role: agent]>],
  Weight: 1];
```

4.1.2 Syntax

The syntactic rules consist of an intial rule and the rules that introduce nouns, quantifiers, and verbs. These rules operate on the following initial structure and their derived structures:

(1) **Initial Propositional Frame**

```
[PHON: ""
 CAT: nil
 SEMstr: /\
 COMP-stack: ⟨⟩
 NP-list: /\    ]
```

Their operations are, however, constrained by subrules like np_indexing, qp_indexing, concat_string and valency_check.

4.2 The System KorSyn6.7.pro for Korean

Unlike the previous English system, this Korean system has a set of morphological rules that handle inflectional phenomena as well as some derivations in Korean morphology. The syntactic rules also handle semi-free word order, coordination, and adnominal relative constructions. (No quantification has been dealt with here.)

5 Demonstration for English

Here are sample sentences:

5.1 Argument Structure

(2) Mia walks

(3) Mia loves Bill

(4) Mia buys Bill Majuang

5.2 Quantification

(5) every man walks

(6) every man loves a woman

(7) every man buys a woman Majuang

6 Demonstration for Korean

6.1 semi-free word order

(8) 미아가 사과를 먹었다
mia-ka sakwa-lul mek-ess-ta

(9) 사과를 미아가 먹었다
sakwa-lul mia-ka mek-ess-ta

6.2 Conjoined sentences

(10) 미아가 사과를 먹고 잤다
mia-ka sakwa-lul mek-ko ca-ss-ta

6.3 Adnominal relative constructions

(11) 미아가 먹은 사과를 용이 먹었다
mia-ka mek-un sakwa-lul yong-i mek-ess-ta

(12) 용이 미아가 먹은 사과를 먹었다
yong-i [[mia-ka mek-un]$_{REL}$ sakwa]$_{NP-lul}$ mek-ess-ta

7 Concluding Remarks

Syntax generates semantic representations and then semantics interpretes them. Semantic representations yield no meaning for themselves. Only when interpreted properly, they make some contributions to the understanding of communications in natural language. The present work has just demonstrated how semantic representations can be derived through syntactic combinatorial operations on lexical items each in a feature structure format. Besides going beyond the treatment of a small frament, the future work should be able to present a set of interpretation rules that also operate in automatic computational ways, thus laying a real basis for computational semantics.