1 Theoretical foundations

This book, encompassing both empirical and theoretical issues of the Korean grammar, develops a constraint- and construction-based HPSG (Head-driven Phrase Structure) analysis. The constraint view assumes that grammar is a system of constraints governing the relation between form and meaning. This view eventually is combined with the philosophy of Construction Grammar in which functional knowledge of language is being based on an individual's systematic collection of "form and function pairings." The book tries to merge these two perspectives of the grammar while seeking the greatest possible precision.

1.1 Derivational vs. constraint-based views

The central tasks of current linguistic theory are twofold.¹ The first task is to explain how children are able to acquire grammatical competence even though they do not receive explicit instruction or have limited access to the data. This task is thus to account for the gap between children's knowledge of language and the apparent lack of substantive input for language, also known as Plato's problem. For example, assume that Korean speakers may hear both of the following Korean sentences including a relative clause:

- a. [_ ppang-ul manhi sa-n] ai-ka wus-ess-ta bread-ACC much buy-MOD child-NOM smile-PST-DECL
 'The child who bought a lot of bread smiled.'
 - b. [[__ppang-ul manhi sa-ss-ta-ko] Mimi-ka malha-n] bread-ACC much buy-PST-DECL-COMP Mimi-NOM say-Mod ai-ka wus-ess-ta child-NOM smile-PST-DECL 'The child who Mimi said bought a lot of bread smiled.'

¹ The discussion in this chapter is based on Pollard and Sag (1994), Sag *et al.* (2003), Kim (2004), Culicover and Jackendoff (2005), Kim and Sells (2008a), and Sag (2013).

1

2 The syntactic structures of Korean

From these two examples, speakers may infer that a subject from a matrix clause as well as from an embedded clause can be relativized.² However, note that Korean speakers may not utter sentences like the following:

 (2) *[[_ ppang-ul manhi sa-n] kakey-lul] Mimi-ka malha-n bread-ACC much buy-Mod store-ACC Mimi-NOM say-Mod ai-ka wus-ess-ta child-NOM smile-PST-DECL '(int.) The child who bought a lot of bread from the store which Mimi mentioned smiled.'

Note that just like those in (1), the sentence in (2) also relativizes the subject of the embedded clause, but Korean speakers would not utter such a sentence. The question that follows is then how and why the language users, with no explicit instruction, can avoid generating such an unacceptable sentence.³

The next and related task is to explain syntactic competence in language. Native speakers have the ability to combine structures to create simple as well as complex sentences. That is, speakers also have no difficulties in producing or understanding novel sentences. For instance, Korean speakers will not have problems in understanding sentences like the following, excerpted from a newspaper article:

(3) mikwuk kwukmwu.cangkwan-i posuthen-uy cathayk aphey America secretary-of-state-NOM Boston-GEN house front ssahi-n nwun-ul an chiw-ess-taka ttakci-lul accumulate-Mod snow-ACC not clear-since ticket-ACC ttey-ess-ta receive-PST-DECL 'The United States Secretary of State got a ticket because he didn't clean up the accumulated snow in front of his Boston house.'

Korean speakers may have never heard or used such a sentence before, but would comprehend or generate such a new sentence without any difficulty. A plausible explanation for this kind of competence can be found from the assumption that each speaker has linguistic potential and creativity to process or generate grammatically new sentences. This implies that a feasible linguistic theory needs to account for such creativity of language speakers.⁴

² In romanizing Korean expressions, we follow the Yale Romanization System, but for personal names we capitalize the first letter.

³ For the discussion of Korean relative clauses, refer to Chapter 13.

⁴ See Hilpert (2014) for discussion of issues related to children's creativity from a Construction Grammar perspective.

CAMBRIDGE

Cambridge University Press 978-1-107-10375-7 — The Syntactic Structures of Korean Jong-Bok Kim Excerpt <u>More Information</u>

Theoretical foundations

In advancing the research into these two tasks (children's grammatical competence and creativity), there have been two main strategies: "derivational" (or minimalist) and "constraint-based" views. Both views assume that sentences are basic units of language, representing pairings of form and meaning. In both views, sentences are also taken to be composed of smaller expressions (e.g., words and morphemes) which are composed into units with hierarchical structures. Departing from these starting points, the two views follow different tracks in many respects. The derivational (minimalist) view minimizes what has to be learned by finding ways to minimize complexity in adult grammar and pack as much complexity as possible into an innate Faculty of Language so that children have less to learn (see, for instance, Chomsky 1977, 1981, 1995, 2000). Meanwhile, the constraint-based view tries to find ways to formulate complexity in adult grammar so that more of it can be learned (see, among others, Pollard and Sag 1994, Jackendoff 2002, Sag et al. 2003, Culicover and Jackendoff 2005, Goldberg 1995, 2006, Sag 2013). One important sub-strategy of the constraint-based view is to minimize elements of linguistic structure that children cannot infer from overt form. This is one main reason the constraint-based view avoids postulating abstract entities such as null elements (e.g., traces or PRO), covert syntactic structures, and movement operations.5

These differences between the two views in the theoretical assumptions also led to differences in modeling the grammar. In the derivational view, linguistic structures are constructed by applying a sequence of rules, each applying to the output of the previous step. In the traditional derivational view, syntax is the sole generative component in language: phonology and semantics are interpretive. Syntax is the source of all combinatorial complexity while phonology and semantics are interpretive, as illustrated in Figure 1.1.



Figure 1.1 Grammar architecture of the derivational view

As shown in Figure 1.1, there is an inherent "directionality" in the logic of sentence construction: certain rules and rule components necessarily apply after others. For instance, semantic components follow from syntactic

3

⁵ See Culicover and Jackendoff (2005) and Sag (2013) for further discussion of the differences between the derivational and constraint-based views.

4 The syntactic structures of Korean

structures, which are projected from the lexicon. Within the derivational view, the formal technology of grammar descriptions is thus derivational. In particular, within the Minimalist Program's view on language (Chomsky 1995, 2000), language is a "perfect" system with an optimal design in the sense that natural language grammars create structures which are designed to interface perfectly with other components of mind – more specifically with *speech* and *thought* systems.

Meanwhile, in the constraint-based view, each constraint determines or licenses a small piece of linguistic structure or a relation between two small pieces. A linguistic structure is acceptable only if it conforms to all applicable constraints. There is no logical ordering in the grammar.



Figure 1.2 Architecture of the constraint-based view

As seen from Figure 1.2, there is no level of representation derived by transforming (destructively operating upon) another level. Instead, all levels are parallel and mutually constrained by the grammar. In the constraint-based view, there are thus no "hidden levels" built of syntactic units. Combinatory complexity arises independently in phonology, syntax, and semantics. There is a continuum of grammatical phenomena from idiosyncratic to general rules of grammar. "Peripheral" phenomena are thus inextricably interwoven with the "core" phenomena. The constraint-based view of grammar has led to the development of LFG (Lexical Functional Grammar), HPSG (Head-driven Phrase Structure Grammar), CxG (Construction Grammar), and SBCG (Sign-Based Construction Grammar). The framework we adopt in this book in describing Korean syntactic phenomena also follows the constraintbased view, in particular, the "construction-based HPSG" which blends the pivotal ideas of HPSG with those of CxG (see Ginzburg and Sag 2000, Michaelis 2013, and Sag 2013). This construction-based HPSG is later further developed as SBCG (Sign-Based Construction Grammar). In this book, we use the terms construction-based HPSG and SBCG interchangeably.

1.2 Linguistic signs and feature structures

In both construction-based HPSG and SBCG, language is taken to be an infinite set of "signs" (whose notion is borrowed from Saussure (1916)). Consider an illustration in Figure 1.3. The form (sound) of *namwu* in Korean is a signifier and its associated meaning or denotation 'tree' is a signified. The linguistic sign is thus a link between a concept and a sound pattern.



Figure 1.3 An example of sign

Adopting Saussure's notion of sign, HPSG models signs as feature structures. All the linguistic objects are thus represented by feature structures notated by attributed-value matrices (AVM), such as:⁶

(4) *n-word* FORM (namwu) SYN *noun* SEM 'tree'

The expression *namwu* 'tree', being a type of *n-word*, has at least attributes about its morphological (FORM), syntactic (SYN), and semantic (SEM) information. All these attributes have their own values. Extending this feature system over all linguistic signs, words and phrases can be also modelled as feature structures. For example, we may represent the information of a verb like *cala-n-ta* 'grow-PRES-DECL' in Korean as follows:

(5) $\begin{bmatrix} v \cdot word \\ FORM & \langle cala - n - ta \rangle \\ SYN & \begin{bmatrix} HEAD & \begin{bmatrix} POS & verb \\ VFORM & fin \end{bmatrix} \\ VAL & \begin{bmatrix} SUBJ & \langle \square NP_i \rangle \\ COMPS & \langle \rangle \end{bmatrix} \end{bmatrix}$ ARG-ST $\langle \square NP \rangle$ SEM grow(i)

As specified in the feature structure, the lexical expression has information about its form (FORM), syntax (SYN), semantics (SEM), as well as

5

⁶ Each linguistic expression also has a phonological (PHON) value, which we ignore in this book. For the logic of feature structures, see Copestake (2002).

6 The syntactic structures of Korean

argument structure (ARG-ST). What we can observe here is that the value of the feature attributes can be either simple (atomic) or complex (feature structure). The value of the attributes POS (parts-of-speech) and VFORM (verbform) is atomic while that of the attribute SYN is another feature structure. These attributes indicate that the finite (VFORM) verb *cala-n-ta* 'grow-PRES-DECL' is a verb (POS) requiring only a subject NP (SUBJ). The ARG-ST information indicates that the verb selects only one argument, realized as the SUBJ of the VAL (valence) feature. The semantic information tells us that there is one individual (*i*) participating in the "grow" semantic relation and this individual is coindexed with the individual denoted by the SUBJ.

Note that in HPSG and SBCG, feature structures are "typed." That is, the feature structure is defined in a more sophisticated way with type information. For example, in (4) and (5), the type information of each expression *namwu* and *cala-n-ta* is marked on the left corner of the matrix that represents a feature structure.⁷ Feature structures are "well-typed" in the sense that every feature structure of some type includes only the features that are appropriate for that type. For example, the VFORM (finite or nonfinite) feature is appropriate only for *v-word* or its subtypes. This is why the following is not a possible feature structure for any *n-word*:

Each *type* thus indicates what kind of feature attributes are appropriate for the given type.

Another important property of the feature structures adopted in HPSG is that typed feature structures are hierarchically organized in terms of linguistic types. The hierarchical classification allows us to capture cross-cutting generalizations among types. As noted, for each linguistic type, certain constraints are stated (the constraints are declared in terms of constraints on feature structures) and the constraints each type carries correspond to grammatical properties shared by all members of that type. The technique of hierarchical inheritance ensures that a type inherits all the constraints of its supertypes. Consider the following examplar hierarchy:

⁷ Throughout this book, the *type* value of each feature structure is marked only when necessary.



The type *verb* is subclassified in accordance with its VFORM (verb form) and number of arguments (TRANSITIVITY), each of which has its own subtypes. Note that each of the sample verbs is cross-classified with respect to relevant supertypes. For instance, the word *cwu-ess-ta* 'give-PST-DECL' is assigned to two distinct subtypes, *ditran-v* (ditransitive) and *fin* (finite). Each of these types specifies a different subset of the information, as illustrated in (8).⁸

(8) a.
$$fin \rightarrow \begin{bmatrix} \text{HEAD}[\text{VFORM } fin] \\ \text{SUBJ} \langle \text{NP}[nom] \rangle \end{bmatrix}$$

b. $ditran-v \rightarrow \begin{bmatrix} \text{COMPS } \langle \text{NP}[acc], \text{PP} \rangle \end{bmatrix}$

cwu-ta 'give-DECL' is the one given in (9):

These constraints are inherited by their subtypes (instances here), like *wus-e* 'smile-CONN' and *cwu-ta* 'give-DECL'. Due to the organization of the lexical signs in this hierarchical fashion, the only information we need to encode for such a word type is its own properties (or constraints) not inherited from the supertypes. For example, all that needs to be stated for the lexical entry of

(9)	<i>ditran-v</i>	1
	FORM	(cwu-ess-ta)
	SEM	give(x,y,z)

As specified here, the lexical information the lexicon needs to specify for the verb is just its FORM value and semantic (SEM) relation. The multiple inheritance mechanism in the hierarchically organized lexicon then allows the lexical entry to inherit all the constraints of its two supertypes in (8), resulting in the more specified lexical entry:⁹

 $^9\,$ The boxed integer is a variable used to 'tag' certain feature values as being token-identical.

⁸ A feature path is a sequence of one or more feature names which is used to select a value from a feature structure. Throughout this book, in representing feature structures, we freely omit feature paths unless required. For example, the SUBJ attribute here omits the feature path SYN and VAL.

8 The syntactic structures of Korean

(10) $\begin{bmatrix} ditran-v & fin \\ FORM & \langle cwu-ess-ta \rangle \\ & & \begin{bmatrix} HEAD & [VFORM fin] \\ VAL & \begin{bmatrix} SUBJ & \langle \boxed{1}NP \rangle \\ COMPS & \langle \boxed{2}NP, \boxed{3}PP \rangle \end{bmatrix} \end{bmatrix}$ ARG-ST & $\langle \boxed{1}NP, \boxed{2}NP, \boxed{3}PP \rangle$ SEM give(x, y, z)

The notion of hierarchical classification of words and multiple inheritance, thus, enables us to eliminate the redundancy and further to capture crosscutting generalizations in a non-redundant, deductive fashion.

1.3 Constructions and multiple inheritance hierarchy

The "construction-based" HPSG (or SBCG) that this book adopts follows the philosophy of Construction Grammar (CxG), in which "constructions" are taken to be the basic units of language and central to all linguistic descriptions and theories of languages. Interpreted within the sign-based system, this means all linguistic signs are taken to be "constructions." A construction consists of a form and a meaning or a function connected with that form, which can be defined as follows (Goldberg 2006: 5):

(11) Definition of grammatical "constructions":

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.

To put it simply, a construction is a form-meaning pair whose meaning we cannot predict from syntactic combinations as well as a form-meaning pair with high frequency whose meaning we can predict. Within this definition, all levels of linguistic description, including morpheme, word, phrase, and clause, are understood to involve pairings of form with semantic or discourse functions, as long as the pairing of form and function is idiosyncratic or unpredictable. Any word-level expression in English (e.g., *smile, laugh, giggle, etc.*) is thus a construction since these parings of form and function cannot be derived from any general rules of the language. The paring is unpredictable and idiosyncratic. An idiom like *kick the bucket* is also a construction where the meaning of parts here (*kick, the, bucket*) does not correspond to the meaning of the whole. The pattern *What is X doing Y*?

Theoretical foundations

9

is also a construction. For example, the sentence *What is the fly doing in my soup?* has an additional, idiosyncratic implicature that the fly's being in my soup is inappropriate (see Kay and Fillmore 1999). In addition to these distinctive constructions, sentences like *John enjoys playing the piano* also involve general constructions like the Subject–Predicate (combination of a subject NP with a predicate VP) and Transitive (combination of a verb with its object) constructions. The meanings of these two constructions are quite predictable and compositional in the sense that the whole meaning can be inferred from the meaning of its parts. The meanings of such typical constructions are predictable and occur with sufficient frequency in daily usage of the language.

The same analogy applies to Korean, as shown in Table 1.1.

Constructions	Examples	
Morpheme	phwus- 'premature', sayng- 'living', -ess 'past'	
Word	salang 'love', ipyel 'separation'	
Complex word	<i>pwus-salang</i> 'puppy love', <i>sayng-ipyeol</i> 'separation while alive'	
Complex word	[V-ess-ta] (for regular past verbs)	
Idiom	<i>pay-ka aphu-ta</i> 'feel jealous' stomach-NOM sick-DECL	
V reduplication	mek-ki-nun mek-ess-ta 'ate it, but'	
'V-ki-nun V'	eat-NMLZ-TOP eat-PST-DECL	
Transitive	kong-ul cha-ss-ta 'kicked a ball'	
'Obj V'	ball-ACC kick-PST-DECL	
Subject-Predicate	namwu-ka cala-n-ta 'The tree grows.'	
'Subj-VP'	tree-NOM grow-PRES-DECL	

Table 1.1 Examples of constructions, varying in size and complexity

The examples given in Table 1.1 illustrate that constructions vary in size and complexity. For example, prefix constructions like *phwus-* 'premature' and *sayng-* 'living' are the smallest in size while the Subject–Predicate construction is the biggest in size here. The prefixes have no variables to fill in, but those like V-reduplication (or Echo), Transitive, and Subject–Predicate constructions have variables to be filled in. The V-reduplication construction involves a variable "V," indicating that any verb can be used in this construction to induce a negative implicature reading (as seen from the English translation) which cannot be composed from any part of the expressions involved. Meanwhile, the Subject–Predicate construction has two variables to be filled in: subject and predicate. Languages can be taken as being composed of constructions of varying sizes.

10 The syntactic structures of Korean

One point worth noting here is that there is no principled distinction between words and phrases: they are all constructions. A lexical entry is more word-like to the extent that it is fully specified, and more rule-like to the extent that it can also have variables that have to be filled by other items in the sentence. The other point to note is that constructions are organized into multiple inheritance hierarchies which enable us to capture language-specific generalizations across constructions (see Goldberg 2003, 2006, and Sag 2013). In what follows, we will see how this notion of an inheritance hierarchy of constructions plays an important role in capturing cross-cutting generalizations among constructions in Korean.

1.4 Korean Phrase Structure Grammar

As noted earlier, this book attempts to describe the major syntactic phenomena of Korean within the framework of a construction-based HPSG. We refer to the grammar of Korean developed from this framework as the KPSG (Korean Phrase Structure Grammar) (see Kim 2004). The KPSG starts with the type hierarchy system in (12) in which every linguistic sign is "typed" with appropriate constraints and hierarchically organized. As defined in (12), the linguistic sign (*sign*) is classified into syntactic (*syn-ex*) and lexical (*lex-ex*) expressions, each of which in turn has appropriate subtypes.¹⁰



The type *syn-ex* represents the expressions that appear at syntax while *lex-ex* indicates morphology-related ones. The former has *phrase* and *word* as its subtypes, implying that only phrasal and word level expressions can appear at syntax. Lexemic expressions (*lexeme*), roughly corresponding to the headwords of a dictionary, are abstract proto-word or root-like expressions.¹¹ Lexemes can be projected into stems and then into words through inflectional processes. For example, the lexeme *wus-* 'smile' will give rise to stems like

¹¹ A related concept is the lemma (or citation form), which is a particular form of a lexeme that is chosen by convention to represent a canonical form of a lexeme.

¹⁰ The HPSG literature defines the subtypes of *sign* in slightly different ways. The type hierarchy given here is a simplified version of Sag (2013). See Ginzburg and Sag (2000), Sag *et al.* (2003), Kim *et al.* (2011), Sag (2013).