CATEGORIAL GRAMMAR WITH FEATURES AND THE PARSER ON WEB PAGES

Hiroyuki Futo

Faculty of Liberal Arts
Tohoku Gakuin University
2-1-1 Ichinazaka, Izumi-ku,
Sendai, 981-3193 JAPAN

E-mail futo@izcc.tohoku-gakuin.ac.jp
http://www.izavc.tohoku-gakuin.ac.jp/~futo/futo.html

ABSTRACT

In this paper, I present a version of Categorial Grammar reinforced with subcategorizing and operational features. Employing the features allows the further specification of combinatory restrictions in natural languages. I show also that by assigning higher-order categories to words, such irregular expressions as "ago" and "last" and controversial constructions such as the formal subject "it," "tough" construction and subject raising can be analyzed only through the rule of functional application. Flat categories are assigned to Japanese verbs for their "scrambling" nature. These are demonstrated on a parser working on Web pages.

1. Classical Categorial Grammar

In classical categorial grammar, categories are defined as follows.

(1)　　a.　Any primitive category is a category.
　　　　b.　If A and B are categories, then A/B and B\A are categories.

The rule (1b) is recursively applied. In categorial grammar, two or three primitive categories are usually employed. I show a list of categories using the three primitive categories, N, NP, and S.

category	traditional name	expression
NP\S　(=IV)	Intransitive Verb	John <u>walks</u>
NP/N	Determiner	<u>the</u> dog
(NP\S) /NP	Transitive Verb	John <u>loves</u> Mary
N/N　(=A)	Adjective	<u>big</u> dog
IV/IV	Auxiliary Verb	<u>can</u> walk
A/A	Adverb	<u>very</u> big
IV\IV　(=ADP)	Adverb	walk <u>slowly</u>

ADP/ADP	Adverb	<u>very</u> slowly
S/S	Adverb	<u>probably</u> he walks
ADP/NP	Preposition	a letter <u>from</u> John
IV/NP	Preposition	run <u>in</u> the park
(S\S)/S	Conjunction	Mary walks <u>if</u> John walks
S/S	Conjunction	Mary knows <u>if</u> John walks

Two expressions are concatenated by the functional application rule.

(2) Functional Application:
    a. If X is an expression of the category A/B, and Y is an expression of the category B, then XY is an expression of the category A.
    b. If X is an expression of the category B\A, and Y is an expression of the category B, then YX is an expression of the category A.

All the expressions in the grammar are generated by the rule of functional application.

2. Categorial Grammar reinforced with features

Here, I present a version of Categorial Grammar reinforced with two kinds of features including features subcategorizing nouns and noun phrases, features subcategorizing sentences, directional features of "←" and "→" and other operational features for negation, copying, adding and deleting. These features will be shown and explained along with extended categories.
Now, the basic categories are defined as sets of subcategorizing features, and there are two groups of basic categories, noun phrase categories and sentential categories.

2.1 nouns and noun phrases

Unlike many versions of Categorial Grammar, the categories N and NP are basically the same and merely differentiated by the presence of the features "p" and "d". All nouns and noun phrases has the feature "N." ("N" hereafter refers to a feature, not a category.) The features are noted as follows.

feature	expressions having the feature
N	all nouns and noun phrases
p	all nouns and noun phrases that can be a subject or an object
d	all nouns and noun phrases with determiners, pronouns and proper nouns
sg	(third person) singular nouns

The following are examples of noun phrase categories and expressions.

(N sg)	dog
(N p)	dogs

(N p sg)	water, air
(N p d)	they, the dogs, his dogs
(N p d sg)	a dog , the dog, he

2.2 sentences

Sentential categories are sets of the following features and features of tense and aspect. All sentential categories have the feature "S" as one of its members.

S	all sentences
q	question

The following are expressions of categories of sentence. No lexical items have a sentential category.

(S present)	He walks, They love dogs
(S q present)	whether he walks, who she loves

2.3 verbs

All categories but noun phrases and sentences are derived from basic categories and they are categories of functional type. They take an expression of some category and give an expression of another category. They are functions from a set of categories to another set of categories. Derived categories can contain operational features.

Intransitive verbs or verb phrases take a noun phrase on its left side and give a category of sentence. Therefore, they are functions from categories of noun phrase to categories of sentence. The intransitive verb *runs* has a category that combines with a singular noun phrase to give a category of sentence with present tense.

runs   (← N p sg)(S present)

The first parenthesized part in the category, "(← N sg p)" expresses the condition of the category *runs* takes as its argument. It requires that the argument come on the left side, and that it have the features "N," "p," and "sg." Note that more than one category can satisfy the condition. The category "(N sg p d)" as well as the category "(N sg p)" satisfies the requirement. When the condition is satisfied, *runs* gives the sentential category "(S present)".

<pre>
       he   runs
   (N sg p d)   (← N p sg)(S present)
            _____/
                  (S present)
</pre>

Contrastively, *walk* has a slightly different category. The operational feature "~" expresses the negative requirement that the input category should not have the specified subcategorizing feature which follows "~." The rule of functional application is extended as categories are. It consists of two processes, the checking or matching process for argument or input categories, and the yielding process for output categories. This operational feature operates in the checking process. "(← N p ~sg)" can be satisfied by "(N p)" and "(N p d)" as its input category.

walk   (← N p ~sg)(S present)

A transitive verb of English takes a noun phrase or a sentence or other expressions on the right side and gives a category of intransitive verb.

loves   (→ N p) ((← N p sg)(S present))
know   (→ S) ((← N p ~sg) (S present))

he loves dogs is given a category of sentence as (3) shows.

(3)   he   loves   dogs
      (N p d sg)   (→ N p) ((← N p sg)(S present))   (N p)

                        (← N p sg)(S present)

                  (S present)
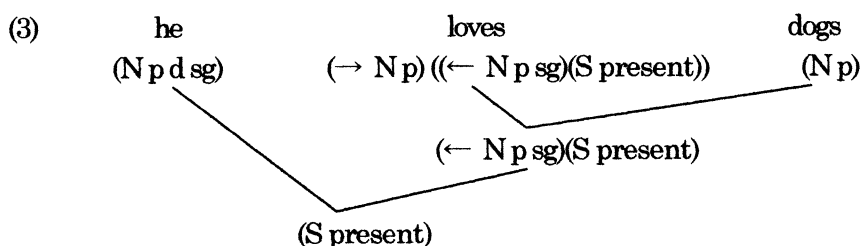
2.4 adjectives and determiners

Adjectives look for nouns on the right side and give the same category as the nouns. The category of an adjective is expressed as follows.

(→ @ N ~d)   good

The operational feature "@" means that the input category should copied and pasted as the output category. Determiners are like adjectives but they add some subcategorizing features to the output category. The operational feature "+" is ignored and does not wok in the process of checking the input category, but it adds the subcategorizing feature following it to the output category.

(→ @ N ~d ~sg)   many
(→ @ N ~p +p +d)   a
(→ @ N ~d +p +d)   the
(→ @ N ~d ~sg +p +d)   these

These categories above can be expressed by the other notation without using the operational feature

"@." The two occurrences of the index number in (4b) specify the ranges of copying and pasting. These are two notational variants of the same category.

(4) a. (→ @ N ~d +p +d)   the
     b. (→ N ~d 1) (1 p d)   the

The parsing of *a good dog runs* is shown below.

(5)

a	good	dog	runs
(→ @ N ~p +p +d)	(→ @ N ~d)	(N sg)	(← N p sg) (S present)

(N sg)

(N p d sg)

(S present)
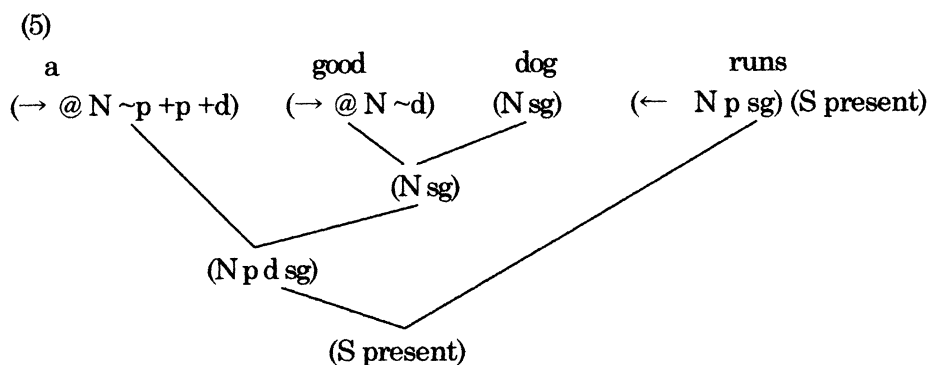
2.5 auxiliaries

Verb phrases concatenated with auxiliary verbs lose the restriction about the number of the subject noun phrase they seek. The operational feature "-" deletes the feature (here, the combination of the features "~" and "sg") that follows it. The adding and deleting operational features do not operate in the process of checking but work in giving output categories.

| can | (→ @ (← N p ~sg -~sg)(S present)) |
| could | (→ @ (← N p ~sg -~sg)(S present –present + past)) |

2.6 irregular expressions

Decomposing syntactic categories in terms of function and features might be compared to decomposing an atom into a nucleus and the electrons orbiting it. How something behaves can be explained by what internal structure it has. How and whether A and B combine can be explained by their internal structures. This "decomposing" tactic can successfully give proper categories to irregular expressions such as *ago* and *next*. The expression *ago* is usually classified as an adverb but behaves more like a preposition but it takes the noun phrase not on the right but left hand. The category of *ago* is like that of an ordinary preposition except that the directional feature is not "→" but "←" for the object noun phrase. Similarly, *next* as in *See you next week* is considered to have a category that takes a noun and gives an adverb.
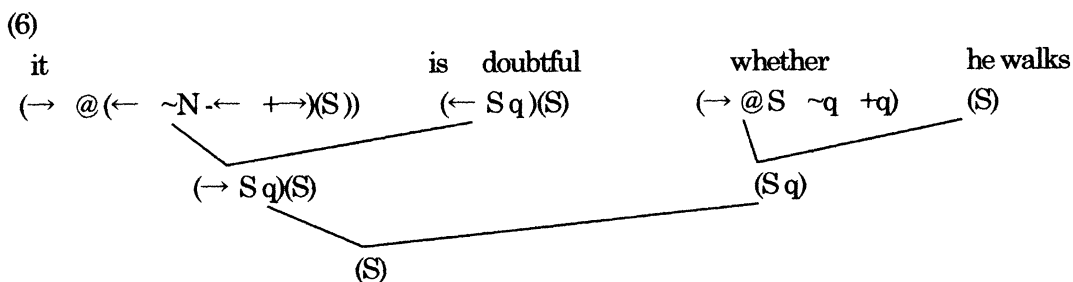
| ago | (← N p time)(← @ (← N p)(S past)) |

next   ($\leftarrow$ N ~p time)($\leftarrow$ @ ($\leftarrow$ N p)(S future))

2.7 the formal subject "it"

Assigning higher-order categories allows some syntactic rules to be dispensed with. The formal subject "it" can be given a category like that of an adverb which modifies an intransitive verb phrase. The difference is that the real subject can not be a usual noun phrase but must be a "that" clause or an indirect question or a "to" infinitive, and that the category of the verb phrase concatenated with the formal subject looks for the real subject not on the left hand but on the right hand. The phrase *is doubtful* below looks for the subject on its left hand, but *it is doubtful* takes it on the right hand. The expression *it* changes the direction. The categories are given as follows.

is doubtful	($\leftarrow$ S q)(S)
is believed	($\leftarrow$ S ~q)(S)
seems	($\leftarrow$ S ~q)(S)
it	($\rightarrow$ @ ($\leftarrow$ ~N -$\leftarrow$ +$\rightarrow$)(S))

Note the changing of the directional features in the two verb phrases in (6).

(6)

it	is doubtful	whether	he walks
($\rightarrow$ @($\leftarrow$ ~N-$\leftarrow$ +$\rightarrow$)(S))	($\leftarrow$ S q)(S)	($\rightarrow$ @S ~q +q)	(S)

($\rightarrow$ S q)(S)   (S q)

(S)

2.8 subject raising

A sentence containing subject raising as in (7b) can be parsed by assigning a higher order category to the expression *to*. The word *to* has a disjunction of multiple categories, and one of them requires on the right hand an intransitive verb like "walk" and then requires an intransitive like "seems" or "is believed" on the left and lastly seeks as the subject such a noun phrase that the first intransitive verb phrase required. (I leave the question open as to how the information about the number of the subject from the main verb survives after merging with that of the infinitive.)

seems	($\leftarrow$ S ~q sg)(S)
is believed	($\leftarrow$ S ~q sg)(S)
to	($\rightarrow$ ($\leftarrow$ N p ~sg 1)(S))(($\leftarrow$ ($\leftarrow$ S ~q 2)(S 3)) ((merge(1 2))(3)))

(7)   a.   It seems that he walks.
         b.   He seems to walk.

(8)

He   seems   to   walk

(N p d sg)   (← S ~q sg)(S)   (→ (← N p ~sg 1)(S))   (← N p ~sg)(S)

((← (← S ~q 2)(S 3))

((merge (1 2))(3)))

(← (← S ~q 2)(S 3))

((merge (← N p ~sg 2))(3))

(← N p sg)(S)

(S)

The expression "to" has another similar category causing "tough" construction. It takes a transitive verb phrase or a verb phrase with a noun phrase gap and then seeks on the left hand a verb phrase like "is easy" which requires "to" infinitive as its subject, and lastly takes as the subject a noun phrase like "a book" which the first transitive verb has been looking for as its object.

3. Categories in Japanese

The Japanese language is a non-configurational language, and has free word orders among the complements and modifiers of a verb. We assume flat categories for Japanese verbs. Then, the transitive verb *yomu* (read) is not given the category (9a) but (9b).

(9)   a.   (← N wo)((← N ga)(S))

         b.   (← N wo)(← N ga)(S)

Here, "wo" and "ga" are the features added to the output noun phrase by the case particles *wo* and *ga*, respectively. (9b) can be regarded as a disjunction of " (← N wo)((← N ga)(S))" and "(← N ga)((← N wo)(S))"Many Japanese auxiliary verbs change case patterns of verbs they modify. *Tearu* (be left) combines with a transitive verb and yields an intransitive verb. The category is as follows.

     tearu   (← @ (← N wo -wo +ga)   -(← N ga) (S +con +sta))

Tearu combines with a verb having a category like "(← N wo)(N ga)(S)" and gives a category like "(← N ga)(S con sta)."Other auxiliary verbs have the following categories, where "sta," and "con" are aspectual features.

     sase   (← @ (← N ga -ga +ni)   +(← N ga) (S))

     rare   (← @ (← N ga -ga +ni)   +(← N ga) (S))

                  (← @ (← N wo –wo +ga)   (← N ga –ga +ni) (S))

tai	(← @ (← N wo -wo +ga) (S +con +sta))
garu	(← @ (← N ga –ga -wo) (← N ga) (S sta -sta))

Aspectual features allow finer specification of selectional restrictions among verbs, auxiliaries, conjunctions and adverbials.

Treating long distance dependency requires another extension of categories. A slash category must be added, which is usually inactive and is inherited and lastly deleted when faced with the binding expression. In (10), a slash category is created by the interrogative pronoun *dono* and deleted by the adverbial particle *mo.*

(10)

dono	gakusei	no	ronbun	mo	kare-ga-yonda
any/which	student	of	paper	too	he read
(→ @ N)(/ wh)	(N)	(← N)(→ @ N)	(N)	(← @ (N +ga +wo) –(/ wh))	(← N wo)(S)

(N)(/ wh)

(→ @ N) (/ wh)

(N)(/ wh)

(N ga wo)

(S)
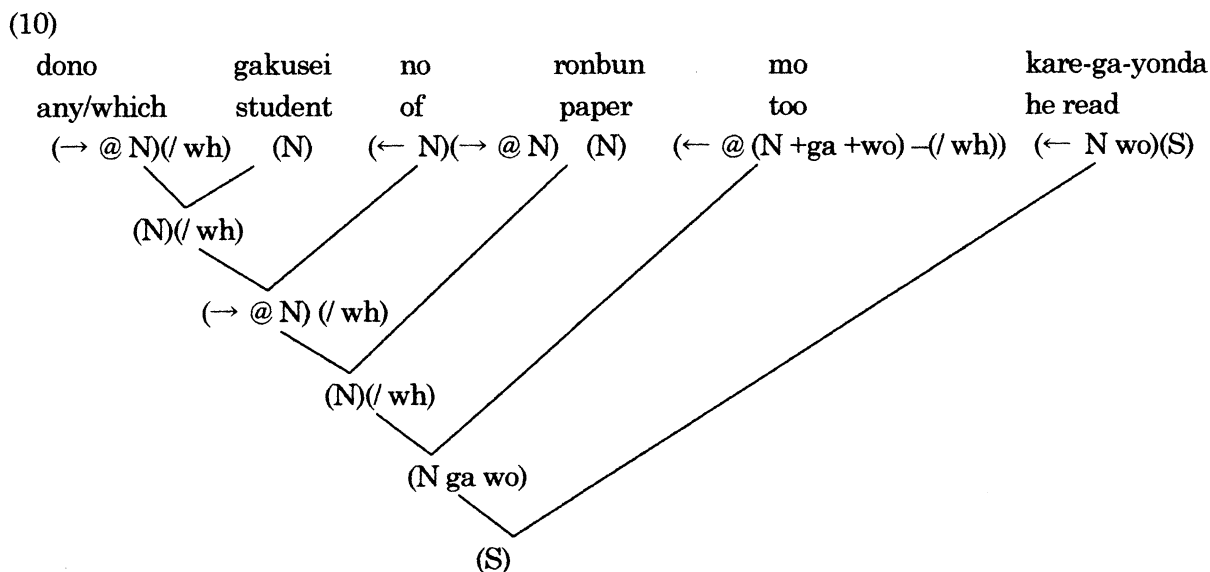
4. Web-based CG Parser

I developed a web-based parser employing the Categorial Grammar discussed above using DHTML and VBScript. It is accessible at http://www.izavc.tohoku-gakuin.ac.jp/~futo/futo.html.

The parser on the web page, after your selecting words by double-click and clicking the "parse" button, displays the parsing tree. Unlike many other grammars, Categorial Grammar, with its simple but firm combination principle called functional application, can be easily coded. Checking process of the input category and the calculation of the output process are the main parts of the scripting. The scripting of the parsing process itself is the simplest that can show the efficiency and effectiveness of the grammar. No hierarchic relations among subcategorizing features are assumed in the program. Anyone can add words and categories and test the parsing.

REFERENCES
Montague, Richard, 1973, "The Proper Treatment of Quantification in Ordinary English," *Approaches to Natural Language* pp.221-242, ed. by J. Hintikka, J. Moravcsic, and P. Suppes, D. Reidel Publishing Co., Dordrecht, Holland