Introduction to Natural Language Processing

"*Ku-ro-i me no o-o-ki-na o-n-na no ko*"

Hiroshi Nakagawa

(Information Technology Center; Mathematical Informatics, Graduate School of Information Science and Technology; Graduate School of Interdisciplinary Information Studies, The University of Tokyo)

nakagawa@dl.itc.u-tokyo.ac.jp

http: //www.r.dl.itc.u-tokyo.ac.jp/~nakagawa/

“*Ku-ro-i me no o-o-ki-na o-n-na no ko*”
'A girl with black eyes'

- The relations between morphemes should be clarified after morphologization.
- Syntax analysis is the study of syntax, or the relations that govern the way morphemes combine to form sentences.
- Approach 1: to seek to construct phrase structures.
- Approach 2: to seek to analyze dependency.
  - The dependency parsing in Japanese is one approach to syntax analysis.
- How many dependency structures can be found?

  e.g. *Ku-ro-i me no o-o-ki-na o-n-na no ko* ‘a girl with black eyes’

  - The non-crossing constraint is embodied in the dependency structures in Japanese.

Phrase Structure Grammar & Rewrite Rules

- Grammar consists of four elements:
  - Lexicon (terminal symbols),
  - Grammatical categories (non-terminal symbols, parts of speech),
  - Rewrite rules, and
  - Initial symbol (sentence).
- Sentence generation starts from the initial symbol from which a rewrite rules begins. According to the rule, a right word is selected from lexicon for every grammatical category.
- Sentence structure analysis is a computational process by which a matching sentence element (morpheme) is selected from the lexicon to be rewritten into grammatical categories by the application of the rewrite rules. This process ends with terminal strings.

Grammar: (examples)

- Lexicon: *Taro, Hanako, Jiro, ga, wo, to, na-gu-ru* 'hit'
- Grammatical category: noun, verb, particle, noun phrase, postposition phrase, verb phrase, and sentence
- Initial symbol: sentence
- Rewrite rule:
  - noun ->*Taro*, noun ->*Hanako*, noun ->*Jiro*, particle->*ga*,

    particle ->*wo*, particle ->*to*, particle ->*ha*, verb ->*na-gu-ru* 'hit'
  - postposition phrase ->noun and particle, postposition phrase -> noun phrase and particle

    noun phrase ->noun, verb phrase ->verb,

    verb phrase ->postposition phrase and verb phrase, verb phrase ->sentence,

    noun phrase ->verb phrase and noun phrase

1. sentence ->postposition phrase + verb phrase
2. sentence ->noun + particle + verb phrase
3. sentence ->*Ta-ro ga* + verb phrase
4. sentence ->*Ta-ro ga* + postposition phrase + verb phrase
5. sentence ->*Ta-ro ga* + noun + particle + verb phrase
6. sentence ->*Ta-ro ga Jiro wo* + verb
7. sentence ->*Ta-ro ga Ji-ro wo na-gu-ru* 'hit'

(Addendum)
Formal Language Theory (FLT) & Automata Theory

- The types of grammars we have discussed are “context-free grammars (CFG)”.
- Three other types of grammars are:
  - Regular grammars (RG): X->aY, X->a
  - Context-free grammars (CFG): X->YZ, X->a
  - Context-sensitive grammars (CSG): The grammars allow the form aXb->aYb besides those of CFG as long as *Y* is longer than *X*.
  - Type 0 grammars: A production rule such as XY->Z is allowed as a shorter version besides that of CSG.

  where *X, Y*, and *Z* each denotes a grammatical category and both *a* and *b* represent terms.
- The hierarchy of these grammars was first described by Noam Chomsky (Chomsky hierarchy).

(cont.)

- The analysis of these grammars are performed on certain types of hardware.
- Regular: Finite state automata. Similar with circuits without a memory device.
- Context-free: Pushdown automata. Circuits and FIFO memory.
- Context-sensitive: Linear-bounded non-deterministic automata. Circuits permitting to write/delete on a finite length of tape. Close to modern computers.
- Type 0: Turning machine. Circuits with an infinite length of tape. Modern computers in real terms.

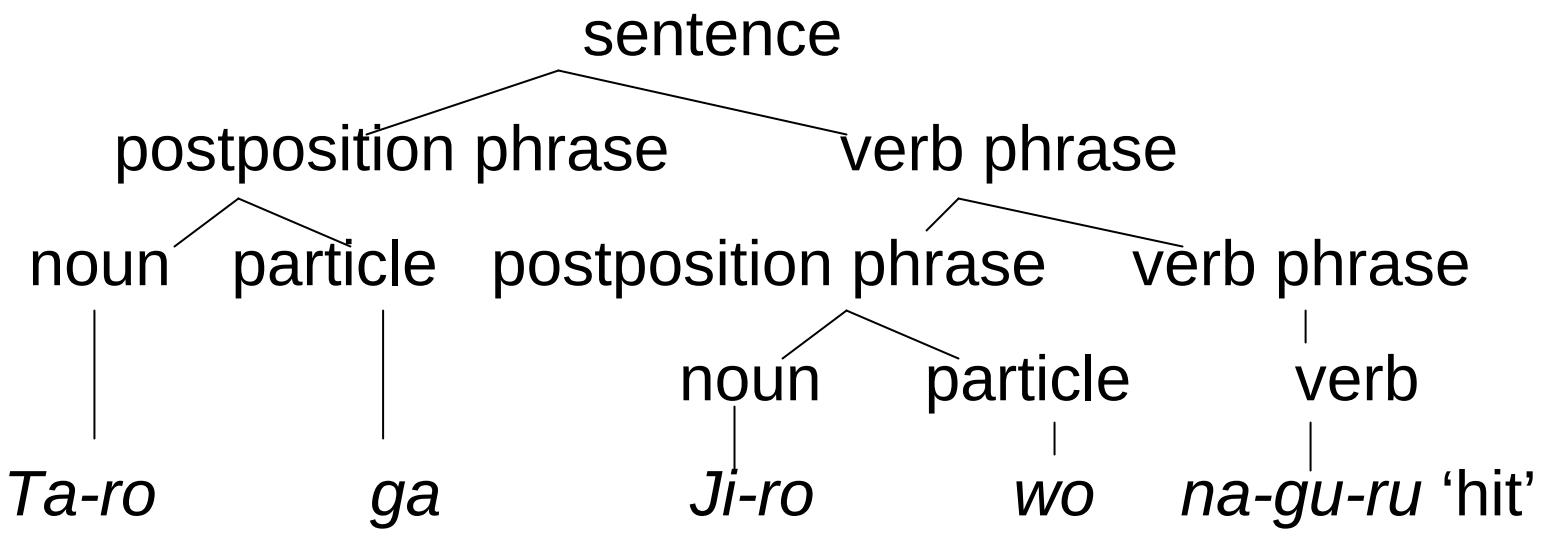
Addendum ends here.

A type of sentences which are hard to process:

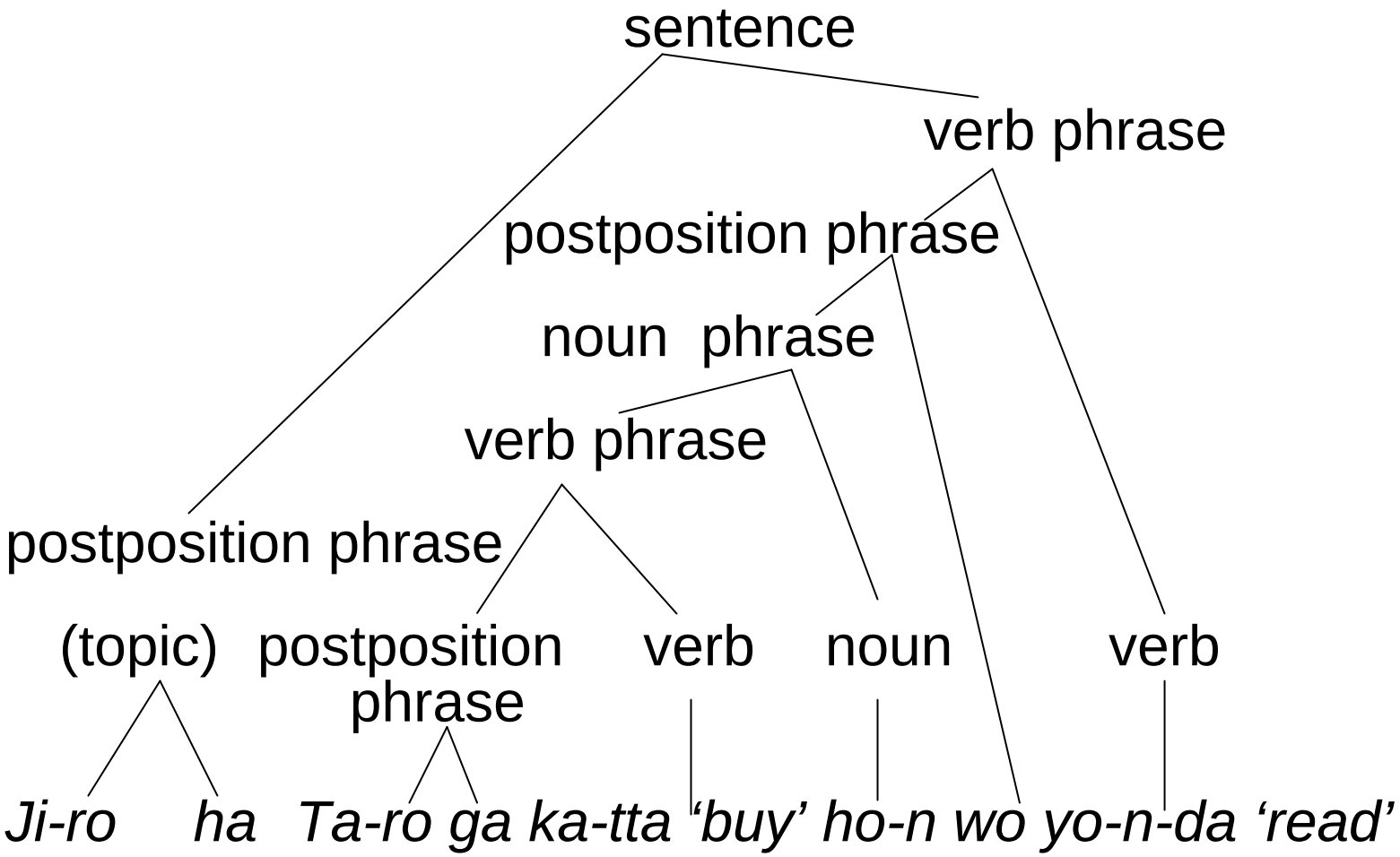
- *Ji-ro ha Ta-ro ga ka-tta ho-n wo yo-n-da* 'Jiro read the book that Taro bought.'
- *Ta-ro ga Ji-ro to Ha-na-ko wo na-gu-ru* 'Taro hit Jiro and Hanako.'
- I saw a girl with a telescope.
- Since Jay always walks a mile seems like a short distance to him.

Representation of Syntax Structure

- Brackets:
  - (*Ta-ro ga* ((*Ji-ro wo*) (*na-gu-ru* ‘hit’)))
- Syntax Tree:

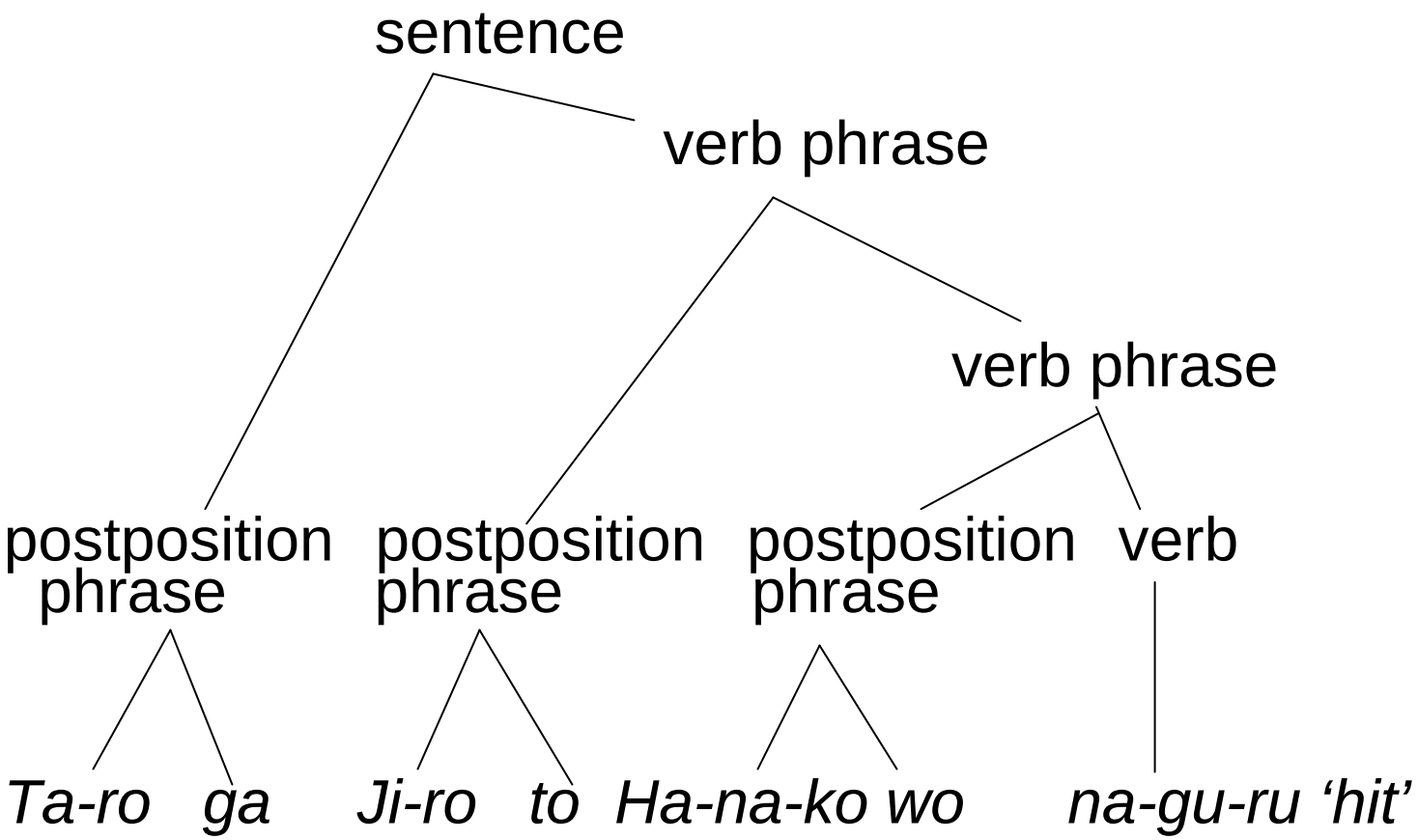


- *Ji-ro ha Ta-ro ga ka-tta ho-n wo yo-n-da* ‘Jiro read the book that Taro bought.’



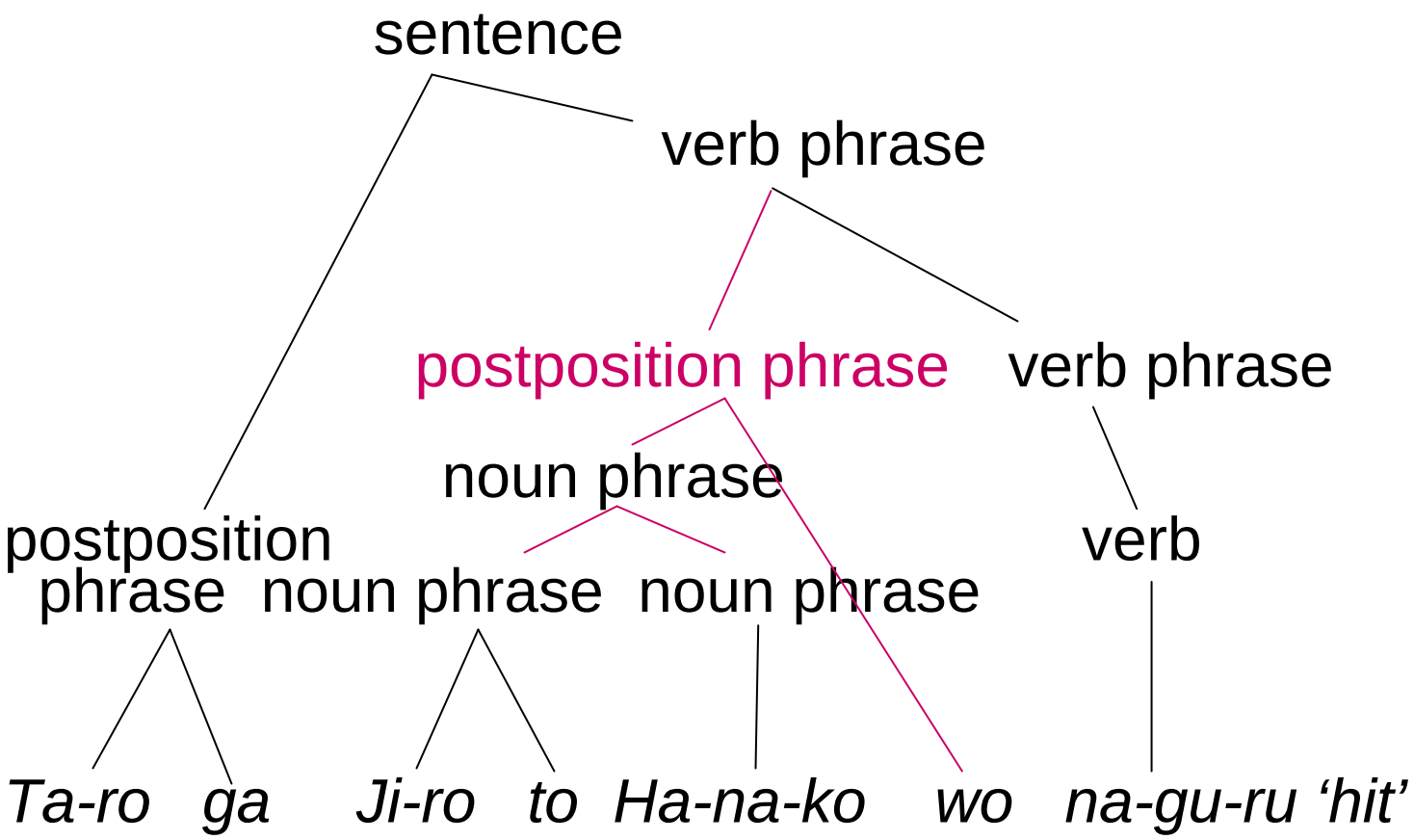
- *Ta-ro ga Ji-ro to Ha-na-ko wo na-gu-ru* ‘Taro hit Jiro and Hanako.’

noun phrase ->noun + noun



■ *Ta-ro ga Ji-ro to Ha-na-ko wo na-gu-ru* ‘Taro hit Jiro and Hanako.’

noun phrase ->noun + noun



A type of sentences which are hard to process:

- *Ji-ro ha Ta-ro ga ka-tta ho-n wo yo-n-da* ‘Jiro read the book that Taro bought.’
- *Ta-ro ga Ji-ro to Ha-na-ko wo na-gu-ru* ‘Taro hit Jiro and Hanako.’
- I saw a girl with a telescope.
- Since Jay always walks a mile seems like a short distance to him. -> Garden path sentence

Algorithm for Syntax Analysis

- Syntax analysis is an integral part of natural language processing, but not a whole.
- Syntax analysis was once a dominant study area of natural language processing.
- Algorithms for syntax analysis are as numerous as stars. (Too many to introduce. No sense to do so in this course.)
- Categories:
  - Top-down, bottom-up, and left-corner
- In the following slides, a powerful algorithm will be described.

Algorithm for Top-down Syntax Analysis

Sample grammar:

✧S->VP, VP->*de-ki-ru* ‘can’, VP->PP V, VP->Adv VP

✧Adv->*su-gu* ‘quickly’, V->*ho-me-ru* ‘praise’, PP->NP *wo*, NP->*Taro*, NP->VP NP

✧0 *su-gu* ‘quickly’ 1 *de-ki-ru* ‘can’ 2 *Taro* 3 *wo* 4 *ho-me-ru* ‘praise’ 5

✧From 0 to 5 are called “locations”.

✧Grammatical categories:

noun: N, verb: V, noun phrase: NP, verb phrase: VP, adverb: Adv, sentence: S, etc.

Algorithm for Top-down Syntax Analysis

Sample grammar :

✧S->VP, VP->*de-ki-ru* ‘can’, VP->PP V, VP->Adv VP

✧Adv->*su-gu* ‘quickly’, V->*ho-me-ru* ‘praise’, PP->NP *wo*, NP->*Taro*, NP->VP NP

✧0 *su-gu* ‘quickly’ 1 *de-ki-ru* ‘can’ 2 *Taro* 3 *wo* 4 *ho-me-ru* ‘praise’ 5

✧From 0 to 5 are called “locations”.

✧Analyzing a sentence from location *i* to *j* to determine grammatical category *X* (POS, noun phrase, verb phrase, clause, etc.) is a task of “instantiation” defined as *X (i,j).* When *X* represents an analysis from location *i* and thereafter, it is written as *X (i)*.

✧The right-hand side of the rewrite rules are instantiated in the same way:

✧Expressed, for example, VP->1PP V, VP->1PP2V, and VP->1PP2V3.

Overview of Algorithm for Analysis

1. Define a location of an input sentence *i = 0, and* a sentence *X (i).*
2. Push X (i) on a stack.
3. Let's find a rule for *X->a*. When word *a* matches the input sentence *i a i+1,* it is instantiated as *X (i,i+1)* and popped out from the stack. Now, it is *i = i+1.*
4. If no rule is found in process 3, then turn to find a rule for *X->YZ* and represent its result as *R = {X->YZ}* (= a set of rules).
5. foreach (R)

   {Push *Y (i)Z* onto the stack.

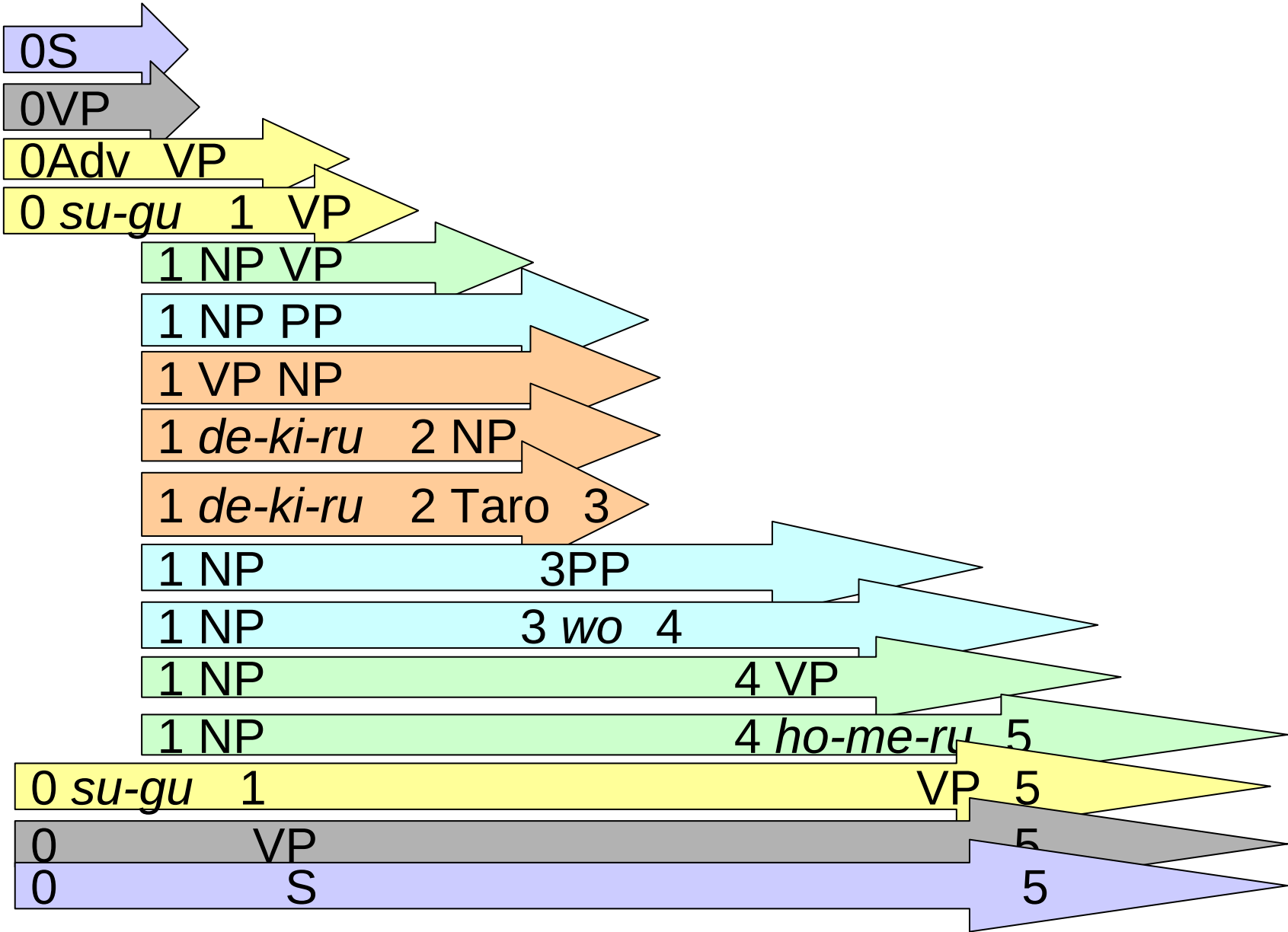
   Run process 3 and thereafter recursively.}
6. Read to the end of the input sentence.

   If the stack is empty, then the analysis went successful. Else unsuccessful.

Generate all possible analyses by applying every applicable rules in the recursive process of 'foreach' (process 5).

Analysis Sample:

- 0 *su-gu* ‘quickly’ 1 *de-ki-ru* ‘can’ 2 *Taro* 3 *wo* 4 *ho-me-ru* ‘praise’ 5



- Input sentence: “(I) praise Taro who can quickly act for what to do”.

	Input word	Stack	Rules for Instantiation
1.		S (0)	
2.		VP (0)S (0)	S->0VP
3.		Adv (0)VPVP (0)S (0)	VP->0AdvVP
4.	*su-gu* ‘quickly’ (0,1)	Adv (0,1)VPVP (0)S (0)	Adv->0 *su-gu* ‘quickly’ 1
5.		VP (1)VP (0)S (0)	VP->0Adv1VP
6.		PP (1)V VP (1)VP (0)S (0)	VP->1PP V
7.		NP (1)PP PP (1)VVP (1)VP (0)S (0)	PP->1NP PP
8.		VP (1)NPNP (1)PP PP (1)VVP (1)VP (0)S (0)	NP->1VP NP
9.	*de-ki-ru ‘can’* (1,2)	VP (1,2)NPNP (1)PP PP (1)VVP (1)VP (0)S (0)	VP->1 *de-ki-ru ‘can’* 2

1.		NP (2)NP (1)PP PP (1)VVP (1)VP (0)S (0)	NP->1VP2NP
2.	*Taro* (2,3)	NP (2,3)NP (1)PP PP (1)VVP (1)VP (0)S (0)	NP->2Taro3
3.		NP (1,3)PP PP (1)VVP (1)VP (0)S (0)	NP->1VP2NP3
4.		PP (3)PP (1)VVP (1)VP (0)S (0)	PP->1NP3 *wo*
5.	*wo* (3,4)	PP (1,4)VVP (1)VP (0)S (0)	PP->1NP3 *wo* 4
6.		V (4)VP (1)VP (0)S (0)	VP->1PP4V
7.	*ho-me-ru ‘praise’* (4,5)	V (4,5)VP (1)VP (0)S (0)	V->4 *ho-me-ru ‘praise’* 5
8.		VP (1,5)VP (0)S (0)	VP->1PP4VP5
9.		VP (0,5)S (0)	VP->0Adv1VP5
10.		S (0,5)	S->0VP5

- In the end, a rewriting is occurring on the stack:

X (i)->X (i,j).

- If there are multiple types of applicable rewrite rules, a separate process is performed accordingly.

- In the last sample, *su-gu* ‘quickly’ is dependent on *ho-me-ru ‘praise’.*

- In the process 3 on the first page of the analysis sample, *su-gu* ‘quickly’ is found to be dependent on *ho-me-ru ‘praise’* by applying VP->PP VP instead of VP->Adv VP.

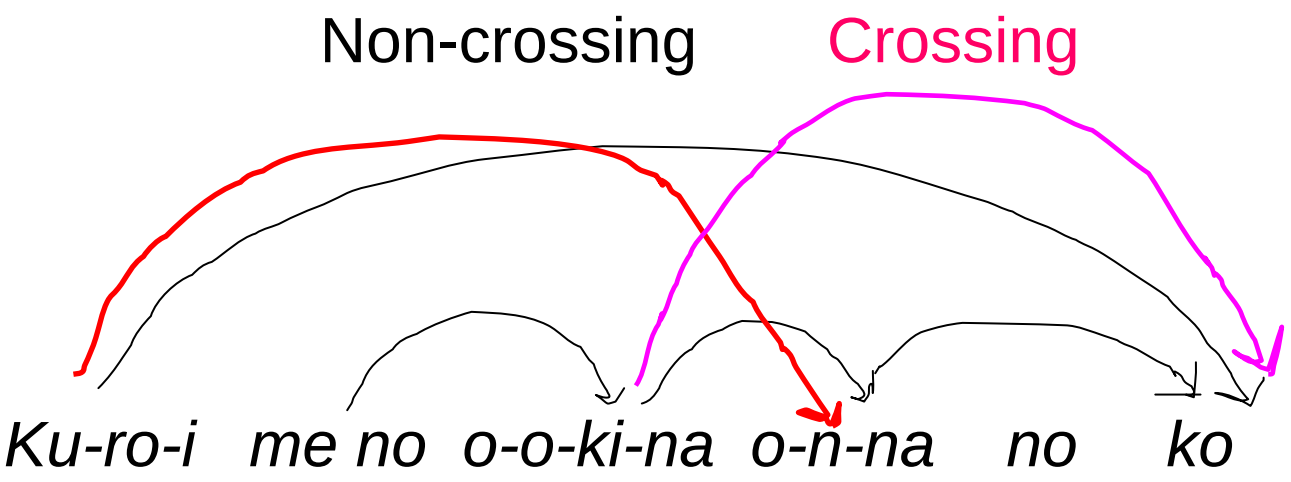
- This operation is performed simultaneously on the generation of structure trees. The structure trees clarify dependency.

System of Dependency Analysis

- Japanese dependency analysis receives the result of morphological analysis (morpheme sequences). These sequences are parsed into clauses to determine the dependency between clauses.
- Due to existence and non-existence of dependency between clauses,
- the analysis begins with the parsing of a string into clauses.
- Non-crossing rule constrains dependency.

Constraint in Dependency

- The non-crossing constraint is embodied in the dependency structures in Japanese.



Possible Dependency Relationships: Sample rules

- If a clause satisfies any condition below:
  - (Dependency: *de* case particle tōten ‘reading mark’)
    (Dependency: *ka-ra* case particle tōten ‘reading mark’)
    (Dependency: *ma-de* case particle tōten ‘reading mark’)
- AND if the latter word meets the constraint below:
  - (word dependency: strong)

Then,

- this relationship is a regular dependency structure.

➢ Such rules are to be defined for each and every type of clause.

Behavior of Dependency Analysis

1. Analyze homographs to convert into a morphological sequence which bears a single meaning.
2. Attach certain marks to morphemes, which will show their behaviors. These marks are dictionary data as well as irregulars such as self-sufficient words and ancillary words which help find clauses.
3. Parse the morphological sequence based on the marks attached.
4. Attach certain marks to clauses, which will show their behaviors. These marks are such as words, indeclinable words, *ga* case, *wo* case, and the possibility of coordinate structure. These marks are technically called “features”.

Behavior of Dependency Analysis

5. If the sentence includes an expression indicating the possibility of coordinate clause, find a similar clause sequence in its vicinity. Group a coordinate clause.
   - E.x. ((((*ri-n-go* ‘apple’ (noun)) *to* (particle (coordinate)))
     ((*ba-na-na* ‘banana’ (noun)) coordinate structure)
     ga (case particle))
6. Follow the rule of possible dependency to find all possible dependency structures of the whole sentence.
   - The coordinate structure should be agreed with headword grammatical characteristics.
7. Evaluate the generated possible structures to output the most preferred.
   - Evaluation is performed based on basic criteria: the aggregate of the distance between each dependency relation and the degree of satisfaction of the surface case of each selected word.

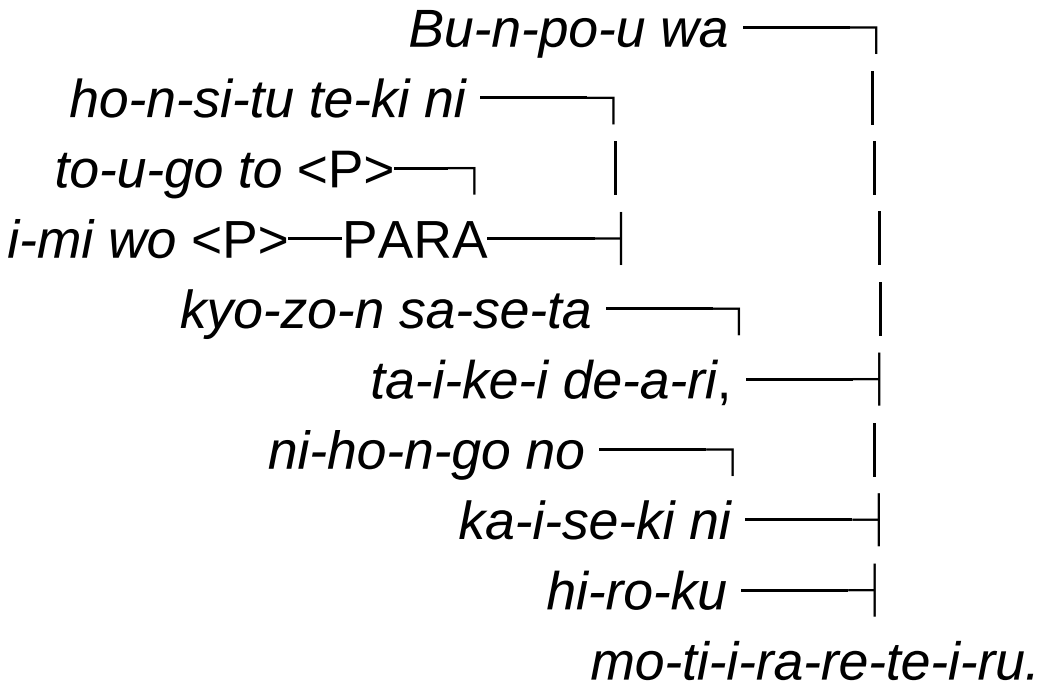
KNP: Sample Rules

- (((Dependency: *de* case particle tōten ‘reading mark’) (Dependency: *ka-ra* case particle tōten ‘reading mark’) (Dependency: *ma-de* case particle tōten ‘reading mark’))
- ([((word: strong)) D])
- ((level: C))
- 2)

◆ Meaning: The clause featured with (Dependency: *de* case particle tōten ‘reading mark’) and (Dependency: *ka-ra* case particle tōten ‘reading mark’) has in a general dependency relationship *D* with the other clause featured with (word: strong). If the sentence consists of a clause featured with (level: C), the second closest clause is to be selected and any other clause located at a farther distance is not taken into considerations.

Example

- *Bu-n-po-u wa ho-n-si-tu te-ki ni to-u-go to i-mi wo kyo-zo-n sa-se-ta ta-i-ke-i de-a-ri, ni-ho-n-go no ka-i-se-ki ni hi-ro-ku mo-ti-i-ra-re-te-i-ru.* ‘Grammar is a system in which syntax and semantics coexist in nature, and widely utilized in the analysis of Japanese.’



Lexicon-based Unification Grammar

- English has subject-predicate agreement in gender and number.
  - He stops -> O, He stop -> X
- Using rewrite rules to process such an agreement would require a tremendous volume of the description of rules.
- Rewrite rules have been decreasing since the 1980‘s. The overall trend shifted toward the use of grammatical characteristics (called features) embedded in each word.
- Rewrite rules are applied only if the features of words agree with each other. For instance, the combination of “he = noun phrase (sing.)” and “stops (sing.)” tells that “he” is the right subject of “stops”. Because “stop (pl.)” does not satisfy the condition, “he” is not the subject of “stop”.
- The matching process is called unification. These grammars are called “unification grammar”.

Explosion of Rewrite Rules

- Start from a basic rule. To analyze *Ta-ro ga ha-shi-ru* 'Taro runs.'…
- Verb phrase -> postposition phrase + verb
- Separately identify intransitive and transitive verbs to analyze a sentence: *Ta-ro ga wa-i-n wo no-mu* 'Taro drinks wine.'
  - verb phrase ->*ga* postposition phrase + intransitive verb
  - verb phrase ->*ga* postposition phrase + *wo* postposition phrase + transitive verb
- Analyze a sentence including two objects: *Ta-ro ga Ha-na-ko ni wa-i-n wo o-ku-ru* 'Taro sends Hanako wine.'
  - verb phrase ->*ga* postposition phrase + intransitive verb
  - verb phrase ->*ga* postposition phrase + *wo* postposition phrase + transitive verb
  - verb phrase ->*ga* postposition phrase + *ni* postposition phrase + *wo* postposition phrase + transitive verb

Explosion of Rewrite Rules

- *Ta-ro ga wa-i-n wo o-ku-ru* 'Taro sends wine.'
  - verb phrase ->ga postposition phrase + intransitive verb
  - verb phrase ->ga postposition phrase + *wo* postposition phrase + transitive verb
  - verb phrase ->ga postposition phrase + *ni* postposition phrase + *wo* postposition phrase + transitive verb
  - verb phrase ->ga postposition phrase + *wo* postposition phrase + *ni* postposition phrase + transitive verb
- Cf. *Ta-ro ga ϕ Ha-na-ko ni wa-i-n wo o-ku-ru* 'Taro sends ϕ Hanako wine.'
- ϕ is an omitted zero noun in *ga* case. *Ta-ro ha* is a topic.
  - verb phrase -> *ha* postposition phrase + *wo* postposition phrase + *ni* postposition phrase + transitive verb
- The line above also must be described. However, *ha* postposition phrase may appear at various locations. Cannot handle by producing a number of rewrite rules…

Then,

- Researchers has sought for a new approach in which complex linguistic features are embedded in each word rather than more rewrite rules are created.
  - Generalized Phrase Structure Grammar: GPSG
  - Lexical Functional Grammar: LFG
  - Head-driven Phrase Structure Grammar: HPSG
- These grammars have been proposed from 1980’s. They have become a basic theory for computational syntax analysis.

Head-driven Phrase Structure Grammar (HPSG) as an Unification Grammar

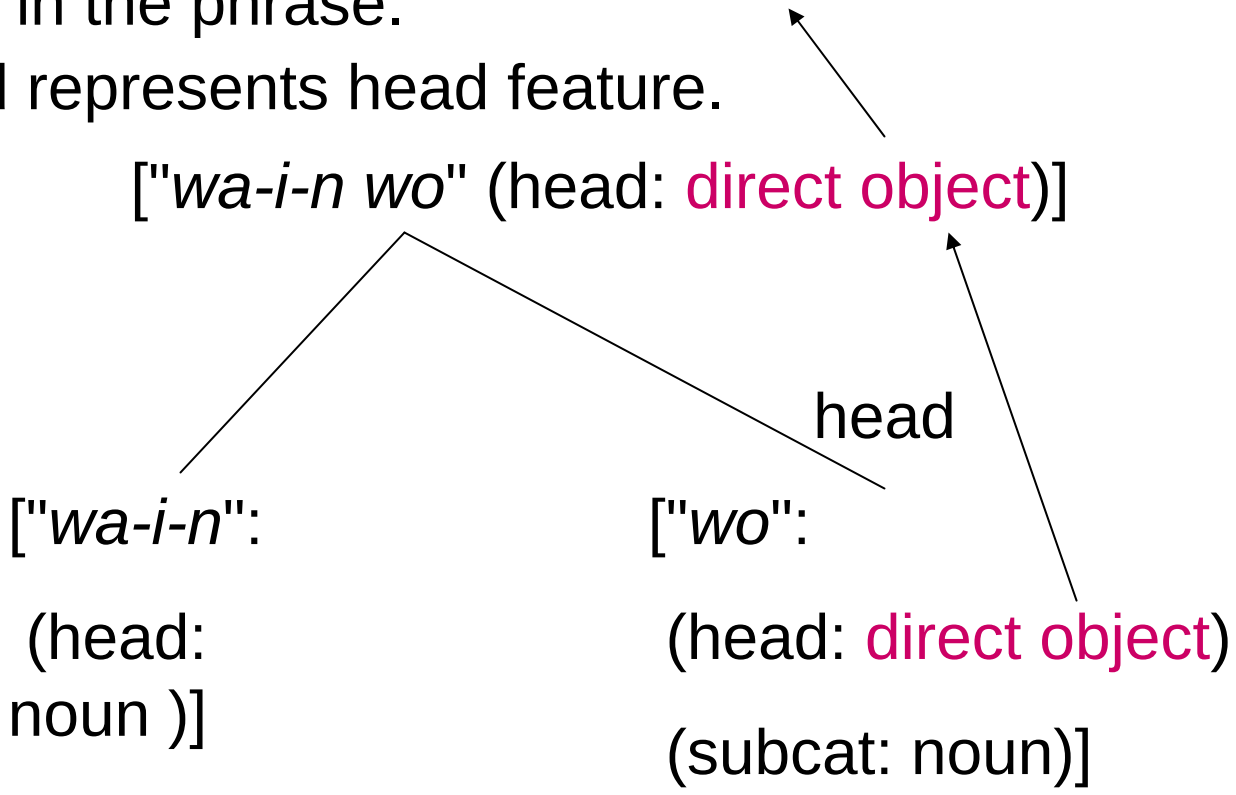
- “Head” (head word) is:
- Which is more grammatically important, *wa-i-n* ‘wine’ or *wo* ‘case particle’ in *wa-i-n wo*?
  - OK: *wa-i-n wo no-mu*
  - X: *wa-i-n ni no-mu*
  - X: *wa-i-n ga no-mu*
- The grammatical relation with the following verb is carried by *wo,* not by *wa-i-n.* Thus, the grammatical entity of the postposition phrase *wa-i-n wo* is *wo.* Such a grammatical entity is called “**head**”.
- Grammar system describes various grammatical features of the head word. The head word conveys grammatical information to adjacent nodes in the syntax tree.

Feature & Subcategorization

- Grammatical characteristics are expressed in several elements.
- Each element is called “feature”.
- For instance, personal pronoun “he” has features as follows:
  - “Pronoun"
  - “Reference (referring to another element)"
- But,
  - “he” is not “recursive”.
  - "himself“ is recursive.
- The system that governs the relation between lexicon and complement is called subcategorization. The grammatical roles of the lexicon are categorized into smaller groups (or “sub” groups) based on this information.
- For instance, verb “drink” has the following subcategorized features:
  - *ga* case subject (postposition phrase with case particle *ga*) and
  - *wo* case object (postposition phrase with case particle *wo*)

Head Feature Principle (PSG)

- The grammatical characteristics pertaining to the head of any phrase, or the value of head feature, is equal to the value of head feature pertaining to the head in the phrase.
- Head represents head feature.



Sample Description of Lexicon after Subcategorization

- intransitive verb:

  [head : verb , subcat {postposition phrase (*ga*)}]

- transitive verb 1:

  [head : verb ,

  subcat {postposition phrase (*ga*),postposition phrase (*wo*)}]

- transitive verb 2:

  [head : verb ,

  subcat {postposition phrase (*ga*),postposition phrase (*wo*),

  postposition phrase (*ni*)}]

- postposition:

  [head : postposition phrase (case particle),subcat {noun phrase }]

The subcategorization principle describes the transition of features in subcategories during the process in which words with subcategorized features are connected to other words and phrases to form a phrase or sentence in an upper category.

- The Subcategorization Principle
  - The value of subcategorized features in the whole phrase is equal to the value pertaining to the head less complement.
- Let‘s examine a case in which a transitive verb is complemented with an object (a postposition phrase) to form a sentence according to the principle.

Application of The Subcategorization Principle

- The process of structuring the postposition phrase with a noun and case particle: (For the sake of easiness of understanding, the case particle takes noun based on the subcategorized feature.)

1. *wo*: [head: postposition phrase (*wo*), subcat {noun}]
2. *wa-i-n wo*: [head: postposition phrase (*wo*), subcat { }]

- Next is the process of combining the postposition phrase and verb:

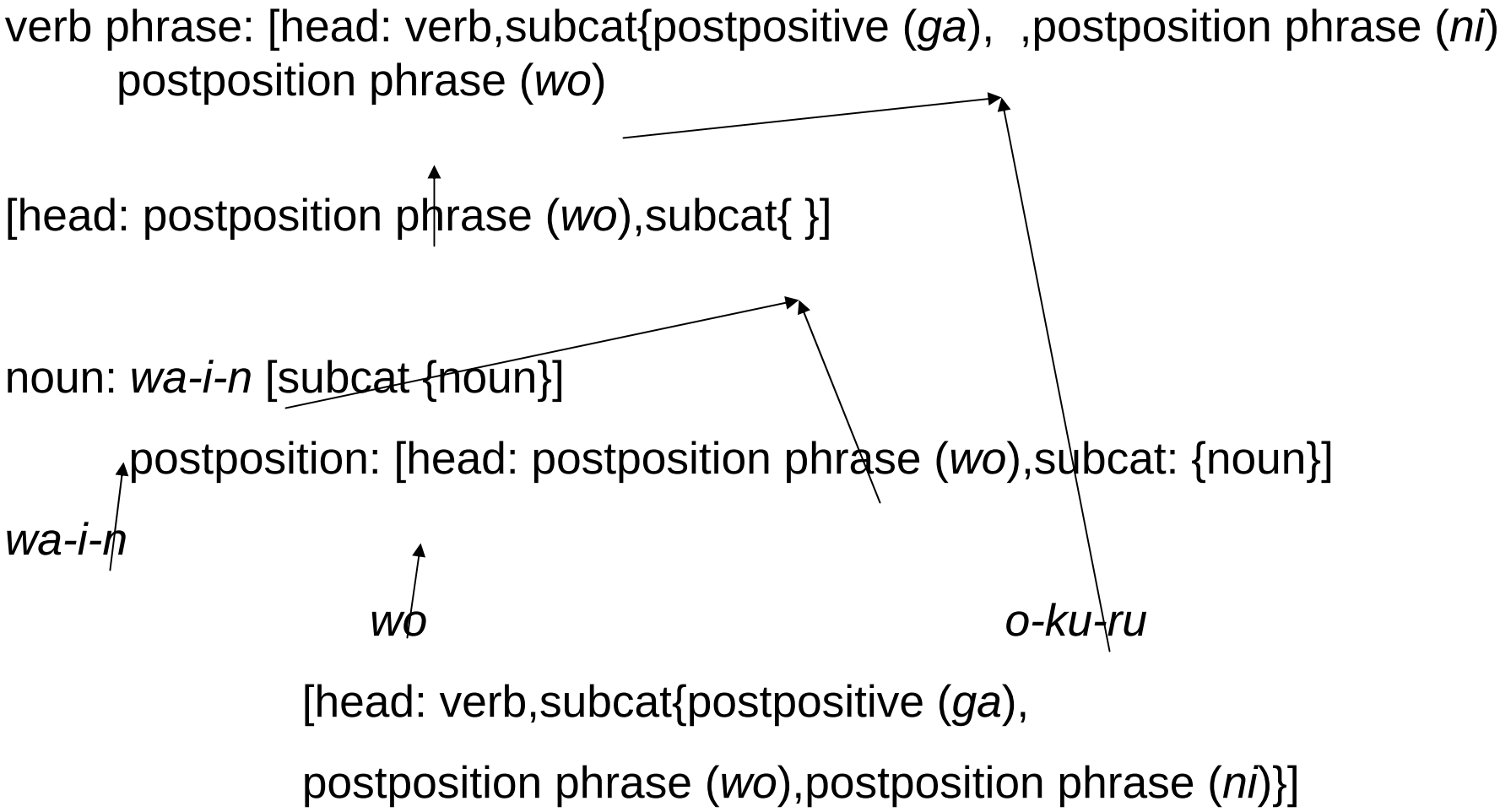
3. *o-ku-ru* 'send': [head: verb, subcat {postposition phrase (*ga*),postposition phrase (*wo*),postposition phrase (*ni*)}]
4. *wa-i-n wo o-ku-ru* 'send wine': [head: verb, subcat {postposition phrase (*ga*), postposition phrase (*ni*)}]
5. *Ha-na-ko ni wa-i-n wo o-ku-ru* 'send wine to Hanako': [head: verb, subcat {postposition phrase (*ga*)}]
6. *Ta-ro ga Ha-na-ko ni wa-i-n wo o-ku-ru* 'Taro sends Hanako wine': [head: verb, subcat { }]

Semantics of words and the abovementioned principle have taken most of these jobs. The number of rewrite rules was contained to a minimal number to achieve the initial purpose.

For instance, there are only three rules for a verb phrase structured with noun, case particle and verb.

- Postposition phrase ->noun postposition
- Verb phrase ->postposition phrase + verb
- Verb phrase ->postposition phrase + verb phrase

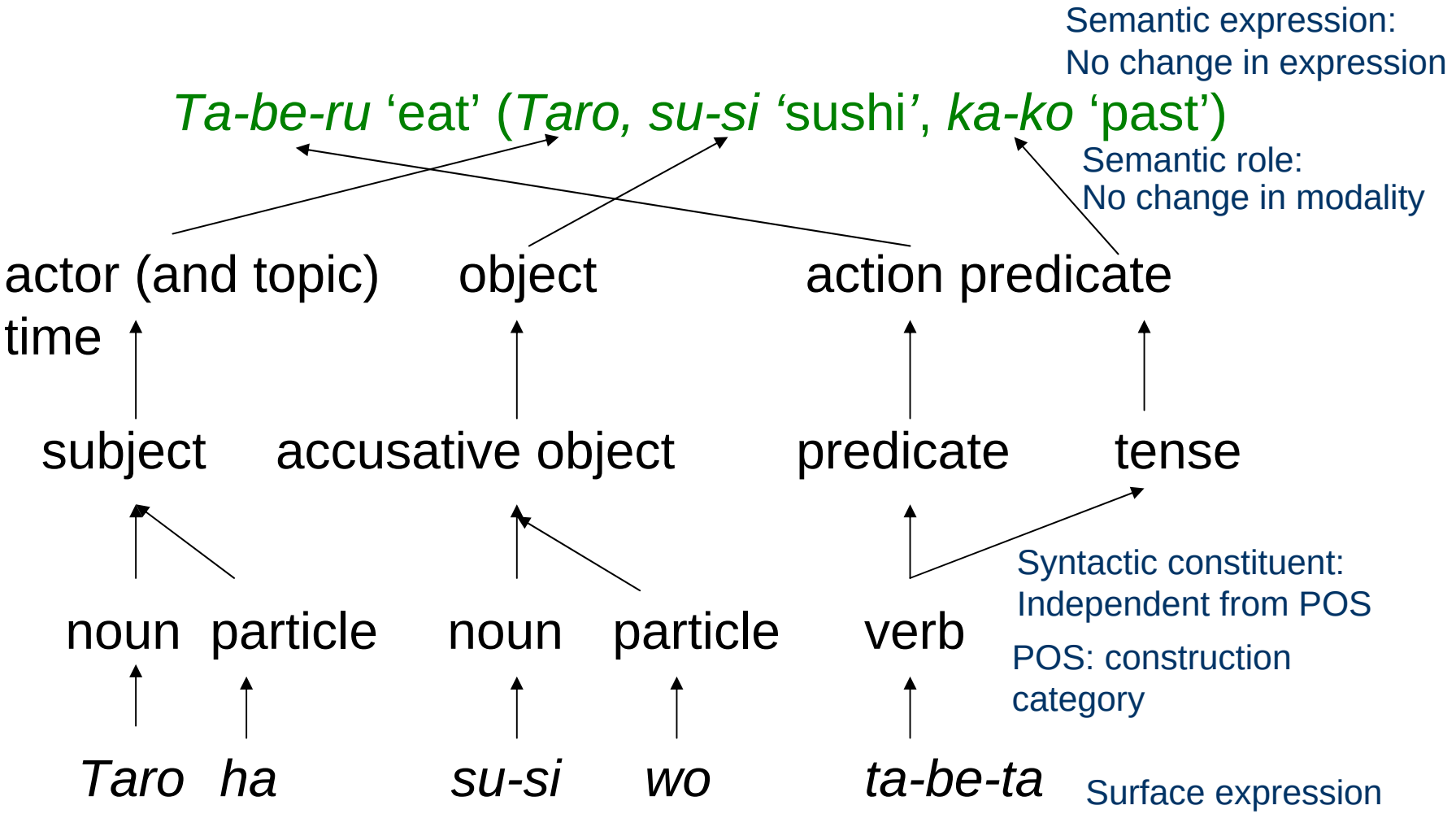
- A part of the verb phrase *wa-i-n wo o-ku-ru* in the previous example is parsed according to the rewrite rules and the result is presented in the following syntax tree.


verb phrase: [head: verb,subcat{postpositive (ga), ,postposition phrase (ni) }]
postposition phrase (wo)
[head: postposition phrase (wo),subcat{ }]
noun: wa-i-n [subcat {noun}]
postposition: [head: postposition phrase (wo),subcat: {noun}]
wa-i-n
wo
o-ku-ru
[head: verb,subcat{postpositive (ga),
postposition phrase (wo),postposition phrase (ni)}]

Semantic Expression

- Part of speech such as verb and noun (construction category)
- Grammatical role such as subject and predicate (syntactic constituent)
- Semantic role such as actor and object
  - Semantic roles in linguistic expressions correspond to goods and things of the real world.
  - Ambiguity roles are the subject and the victim (adversity passive): pragmatic roles.
- A semantic expression of proposition such as *ka-su* ‘lend’ (Taro, Hanako, *ho-n* ‘book’)
- One of Grammar’s tasks is to define these relations.
- Essential in computational linguistic modeling.

Connection to Linguistic Expression to Semantics

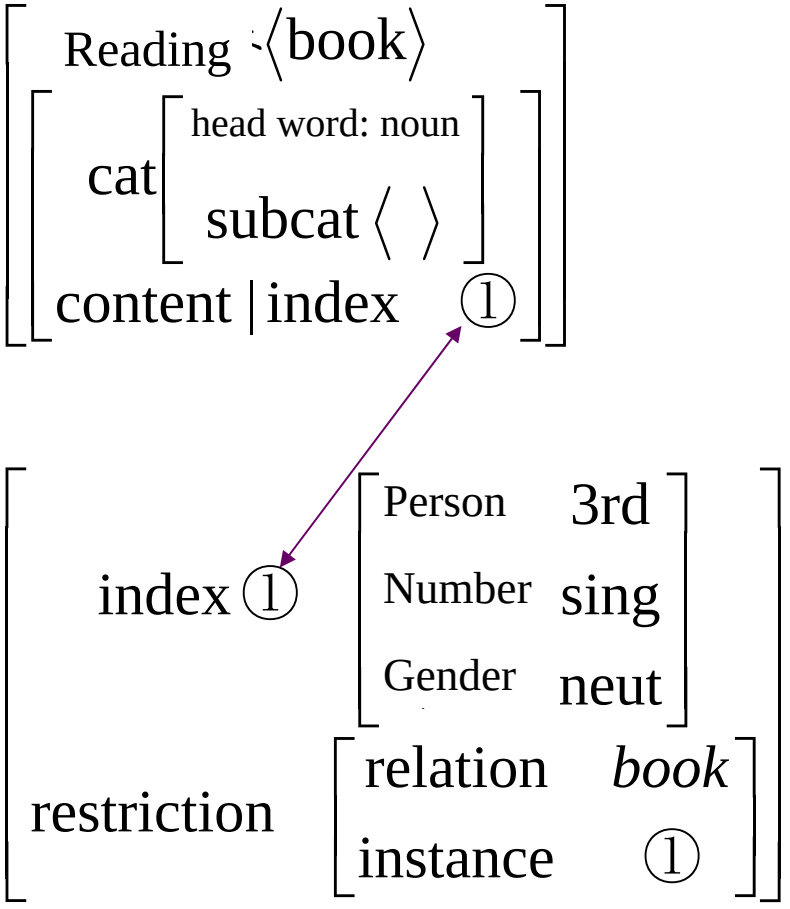


Relation between Level

- The subcategorization principle mediates subcat information between POS.
- One of the grand themes in linguistics is how grammatical roles (constituent: cases (nominative, accusative, dative, etc.)) mediate semantic roles.
- θ role (Chomsky Theory) is semantic though slightly grammatical.
  - Topic (theme) is a θ role; however it is a grammatical concept rather than a semantic one.
  - Linguistics as a refined theory in which the relation between θ roles and grammatical cases are defined.

Common Description Method of Grammar & Semantics by HPSG

- Sample structural expression of noun's characteristics



Common Description Method of Grammar & Semantics by HPSG: E.x. verb gives

$$\left[\begin{array}{l} \text{Reading} \quad \langle \text{gives} \rangle \\ \left[\begin{array}{l} \text{cat}\left[\begin{array}{l} \text{head word: verb} \quad [\text{finite}] \\ \text{subcat} \left\langle \text{NP [nominative case]}_{1[\text{3rd sing}]}, \text{NP [dative case]}_{2}, \text{NP [accusative case]}_{3} \right\rangle \end{array}\right] \\ \text{content}\left[\begin{array}{cc} \text{relation} & \textit{give} \\ \text{giver} & ① \\ \text{given} & ② \\ \text{gift} & ③ \end{array}\right] \\ \text{context}\left[\text{speaker} \quad ①\right] \end{array}\right] \end{array}\right]$$

Meaning of Goods

- Goods play various semantic roles as au augment of predicate.
  - The meaning of goods in real world is related to that of predicate and plays a role as an augment of the predicate. (Details further explained.)
- First comes the meaning of goods itself.
- Thesaurus describes classification schemes to define the semantic hierarchy of goods (and things).
- A detailed description of meaning would require to break up words into smaller units. Therefore, they are segmented into independent semantic constituents (semantic characteristics).

Semantic Characteristics Constituting the Meaning of Goods

IPA Nominal Dictionary

- ANI (Animal)
- HUM (Human), AML (Animal other than human)
- CON (Concrete)
  - AUT (Autonomy e.x. computer), EDI (Edible), LIQ (Liquid), PAS (Viscous), SOL (Solid)
  - Rice: EDI&PAS&SOL, Beer: EDI&LIQ
- SPA (Place)
  - LOCUS (Initial point, End point), INT (Internal e.x. room), ORG (Organization), NET (Network e.x. transportation network)

Semantic Characteristics (cont.)

- Event, action, and effect: PRC
  - ACT (action)
  - EVE (event)
  - APO (appointment: e.x. the bank opens at 9 am.)
  - RES (result: e.x. natural disaster)
  - PRO (product: e.x. to bake bread)
  - PHE (natural phenomenon: e.x. to ice over)
  - NAT (natural objects and phenomenon: e.x. typhoon, the sun)
  - PLA (planet)
  - GAS (gas: e.x. smog, breath)
  - ELM (elements that the five senses cannot perceive: e.x. protein, nerve system)
  - POT (potency: e.x. foot, shoulders, lung, intestines)

Semantic Characteristics (cont.)

- Abstract: ABS
  - Price (income, price)
  - Measure (height, weight)
  - Information (information, height, novel, music, critique, address)
  - Quantity (weight, area)
  - Social bonds (disparity, relation)
  - Grade (status, evaluation, scale)
  - Form (attribute to be evaluated: e.x. taste, form)
  - Attribute (measurable by degree: e.x. no common sense, progress, salt)
  - Reciprocity (compatibility)
  - Personality (pride, personality)
  - Mind (feeling, nerve)
  - Manner (ability, nature, etc.: e.x. cooking, end, presentation, driving, color design, people management)

Semantic Characteristics (cont.)

- Abstract: ABS
  - Method (manner, method)
  - Objective-value (e.x. read, square)
  - Sensational-value (sweet, hot)
  - Evaluation (e.x. financial difficulties, financing, taste)
  - Currency (price: e.x. $100, ¥1000)
  - Duration (period: e.x. 3 years)
  - Distance (e.x. 3 km)
  - Item (numbers: e.x. 3 persons, 1 piece)
  - Ratio (e.x. 30%)
  - Quantity (e.x. 30kg)
  - State (e.x. stable, happy, unhappy, quiet, possible, stubborn)

Semantic Characteristics (cont.)

- Abstract: ABS
  - Role
  - Relational-term (relative, friend)
  - Direction (east, west, south, north, left, right, up, down, front, back)
  - Phase (time, location order)
  - Reference-point (relation to reference: e.x. opposite, more than)
  - Norm (rule, principle, law, equation)
  - Subfield (academic principle, art, sports, etc.)
  - Inclination (psychological inclination: e.x. Interested, accustomed)
  - Appearance (e.x. Impression, attitude, trace)
  - Unit
  - Time-point
  - Time (order of events, abstract time: e.x. future)

Semantic Characteristics (cont.)

- Abstract: ABS
  - Ordinal (order)
  - Name
  - Entity
  - Congregation (e.x. crowd, society, volunteer)
  - Kind (e.x. human kind)
  - Abstract (the other abstract concepts)

Semantic Roles of Verb (Human)

- Agent
  - Initiator in control of the action. Can be a causer.
- Actor
  - Performer of the action who cannot animate, but may act in some intuitive way. Cannot be a causer.
- Patient
  - The entity undergoing the effect of the action.
- Experiencer
  - The entity that has already underwent the effect and reached a certain state

Example

- *Ta-ro ha ha-si-tta* ‘Taro ran.’
  - *ha-si-ru* ‘run’ (actor = *Taro*, time <now)
- *Ha-na-ko ha ri-n-go wo ta-be-ta.* ‘Hanako ate an apple.’
  - *ta-be-ru* ‘eat’ (agent = *Hanako*,patient = *ri-n-go* ‘apple’, time <now)
- *Ji-ro ha o-do-ro-i-ta* ‘Jiro was surprised.’
  - *O-do-ro-ku* ‘surprised’ (experiencer = *Jiro*, time <now)
- *Ta-ro ha Ji-ro wo o-do-ro-ka-se-ta* ‘Taro surprised Jiro.’
  - cause (agent = *Taro*, patient = *Jiro*,
    O-do-ro-ku ‘surprised’ (experiencer = *Jiro*), time <now)

Semantic Roles of Verb

- Object or Thing
  - Object of the action.
- Place
  - Place where the action was taken place.
- Resultant
- State
  - State refers to a neutral situation. Resultant corresponds to the stative situation that follows the action expressed with a verb. E.x. bear->alive. Other examples that are not the resultant state are, such as sleepy and hungry.

Subcategorized Semantic Roles of Verb
(consisted of such semantic elements as below)

- Object & Thing
  - Object of the action. Distinction between physical object vs. abstract object.
- Location
  - Place where the action takes place.
- Source
  - Place where the action starts.
- Goal
  - Place where the action completes.
- Direction
- Path
- Instrument
  - Taro goes to school from his house by bike on route101.
  - go (actor = Taro, source = his house, goal = school, path = route101, instrument = bike, time = now)

Subcategorized Semantic Elements of Verb

- Subcategorize into basic semantic elements:
  - affect — affect (actor, patient)
  - effect — effect (actor, resultant)
  - act — act (actor, X), X = patient, object, …
  - experience — experience (experiencer, state)
  - order — order (agent, action (actor)),
  - …subcategorize into smaller elements…
- be, move, cause, alive, die, see, hear, have,eat, sleep, sell, buy…
- Meta elements such as intuition: volitional
- E.x.:
  - *Ko-ro-su*: kill (actor, patient) = $\text{cause}_{\text{volitional}}$ (actor (not (alive (patient)))
  - *U-mu*: kill (actor, patient) = $\text{cause}_{\text{volitional}}$ (actor (alive (patient))
- E.x.: Ta-ro ni ga-kko ni I-ke to me-i-zu-ru. '…told Taro to go to school.'
  - order (agent = X,act: *I-ku* 'go' (actor = Taro, goal = *ga-kko* 'school'))

From Thing to Statement Vol.1

- A “thing” bears meaning when the meaning of predicate (verb) is combined with its augment ’s meaning.
- The thing is called proposition when the subjectivity of speaker is eliminated.
  - Thing (proposition): e.x. *Ta-ro ga Ji-ro wo na-gu-ru* ‘Taro beats Jiro.'
  - beat (agent = Taro, patient = Jiro)
- Speaker’s attitudes in the proposition and modality are statement.
  - The relation between the times when the thing begins to happen and when the statement is made is expressed with tense (past, present, and future).
  - *Ta-ro ga Ji-ro wo na-gu-tta* ‘Taro beat Jiro.'
  - beat (agent = Taro, patient = Jiro, time < ST)

ST: speech time

From Thing to Statement Vol.1

- Speakers attitude towards the proposition and modality is called statement.
- The modality toward the proposition is presumption (*da-ro-u* ‘assume’: subjective, *ra-su-i* ‘suppose’: hearsay), hearsay (so-*u-da* ‘hear’)
  - *Ta-ro ga Ji-ro wo na-gu-tta* so-u-da. ‘I heard Taro beat Jiro.’
  - hearsay (beat (agent = Taro, patient = Jiro, time < ST) news-source = X)
- The meaning of the sentence is expressed in an embedded structure in terms of the context in which the sentence is spoken: (statement (proposition)).

Pragmatic Roles of Statement

- Topic & Theme
  - Things are critical elements in a sentence.
- Speaker
  - Speakers can be those who hear. E.x. *Ka-re ni yo-re-ba mo-u o-wa-tta-ra-si-I.* ‘He says it is over.’
- Hearer
  - Hearers are implicitly introduced with ending particles in Japanese such as *yo* and *ne* ‘isn’t it’.

Modality of Hearer

- Topic is speaker's evaluation of the proposition. It can also be the ordering and prioritization of what the speaker wants to convey to the hearer.
  - *Ta-ro ga Ji-ro wo na-gu-tta* so-u-da. 'I heard Taro beat Jiro.'
  - (topic = Taro,hearsay (beat (agent = topic, patient = Jiro, time < ST)))
- Modality of hearer: How is "proposition + modality" communicated to the hearer? Interrogative *ka,* reassertion *yo*, or recognition *ne.*
  - *Ta-ro ga Ji-ro wo na-gu-tta* so-u-da ne. 'I heard Taro beat Jiro, didn't he?'

  ->The speaker tries to reassure the proposition with the hearer. The proposition may be what the speaker has heard from the hearer.
  - *Ne* (speaker, hearer, hearsay (beat (agent = Taro, patient = Jiro, time < ST) news-source = ?hearer))
  - *Ta-ro ga Ji-ro wo na-gu-tta* so-u-da yo. 'I heard Taro beat Jiro, didn't he?'

  ->The speaker tries to assert the hearer again (reassertion).
  - *Yo* (speaker, hearer, hearsay (beat (agent = Taro, patient = Jiro, time < ST) news-source = X)) X is not the hearer.

Time (Reihenbach)

- ST (speech time)
- ET (event time)
- RT (reference time): speaker's empathy
  - Past perfect: ET <RT <ST = now
- Tense: Co-relation between ST, RT, and ET.
- Aspect: specifies as a time span in relation to the continuous time frame of a phenomenon. Expressed as -ing and -ed in English.
  - Continuation, repetition, completion, residual effects, etc.

Japanese Tense & Aspect

- First aspect: Tense: imperfective form *su-ru v.s.* past *si-ta*
  - *su-ru* is future tense. *Bo-ku ha be-n-kyo-u su-ru.* 'I am going to study.' *su-ru* also expresses speaker's will. 'I will study.'
- Aspect: *te-i-ru* is time aspect expressing the ambiguity of perfect and continuative.
  - cf. (*su-ru* vs. *si-ta*) vs. (*si-te-i-ru* vs. *si-te-i-ta*)
- Second aspect: "verb + *te* + *" such as *si-te-I-ru carries information other than time aspect, for example, actor's intention and object's characteristics.*
  - This form is frequently seen in Japanese.
- Third aspect: Certain types of verbs antecedent other verbs to generate complex meaning.
  - *E.x. Si-ha-ji-me-ru 'get started'*
- Mood and modality are to be expressed with auxiliary verbs.

Second Aspect

- *te-i-ru*
  - Continual verb + *te-i-ru*: on-going
    - *Yo-n-de-I-ru, ta-be-te-I-ru*
  - Punctual verb + *te-i-ru*: on-going effects
    - *Si-n-de-I-ru* ‘be dead’*, ki-ma-tte-I-ru* ‘be fixed’*, o-wa-tte-I-ru* ‘be completed’
    - When *te-i-ru* is used with polysemous verbs which are both punctual and continuative, *te-i-ru* also bears polysemous meanings.
  - State verb + *te-i-ru*: emphasized in a state
    - *So-bi-e-te-I-ru* ‘be high up’, *ba-ka-ge-te-I-ru ‘be foolish’*
- *te-si-ma-u* emphasized perfect
- *te-i-ku and te-i-ru*
  - They represent (spatial and temporal) directions for the actor.
  - *Mi-ni-I-ku* ‘go to see’*, ka-tte-ku-ru* ‘go to buy (something)’

Second Aspect

- *te-a-ru:* Result of a past intended action
  - X: *Byo-ki ni na-tte a-ru. ‘remain sick.’*
  - Can be used when the actor and his/her intention are unknown. E.x. “A CD is left.”
- *te-o-ku, te-mi-ru, te-mi-se-ru:* An intended ction of the actor.
  - Focusing on action rather than result.
  - *te-mi-se-ru:* Focusing on the opponent of the action.
  - X: *ka-yu-ku na-tte mi-se-ru* Cf. O: *ka-yu-ga-tte mi-se-ru*
- *te-ya-ru/te-a-ge-ru, te-mo-ra-u/te-i-ta-da-ku, te-ku-re-ru/te-ku-da-sa-ru:*
  - Focusing on the relations of the actor and the beneficiary.

Third Aspect

- Certain verbs (main verbs) are succeeded with a group of verbs (sub classes of verbs), which can also be used by themselves, to form combined meanings.
- *ha-ji-me-ru* ‘get started’, *ko-mu* ‘get into’, *da-su* ‘start out’, *a-u* ‘meet’, *tu-zu-ke-ru* ‘continue’, *ka-ke-ru* ‘get to it’, *a-ge-ru* ‘give’, *ki-ru* ‘complete’, *tu-ke-ru* ‘attach’, *tu-ku* ‘reach’
- These words proceed the second and first aspects.
- Semantically, they combine the meaning of main and sub verbs.
  - Most semantic structures are (meaning of verb in sub classes (meaning of main verb)).