

ÉTUDE DE LA PHRASE COMPLEXE

Dans ce chapitre, nous tenterons de cerner les notions connexes à la phrase complexe, notamment la proposition. Le chapitre, ouvert par la définition des deux problèmes centraux de la proposition japonaise (§ 7.1), sera composé de quatre parties. La première sera dédiée au premier problème, que nous commencerons par décrire (§ 7.2), avant de nous intéresser aux travaux antérieurs sur la définition de la proposition (§ 7.3), qui nous permettront de découvrir les critères concrets pour déterminer les syntagmes à mot variable non-propositionnels (§ 7.4), et de réaliser la première définition de la proposition et de quelques autres unités (§ 7.5). La seconde partie, consacrée au second problème, sera constituée d'une description de celui-ci (§ 7.6) suivie d'un état de l'art (§ 7.7) et se terminera par la présentation de notre catégorisation des mots susceptibles de réaliser la connexion des propositions (§ 7.8). La troisième partie comportera les études sur la typologie des subordonnées, composées d'un état de l'art (§ 7.9) et de la présentation de notre typologie (§ 7.10). Enfin, la dernière partie commencera, pour une récapitulation, par une définition formelle de la phrase (§ 7.11), et se terminera par un débat sur les principaux problèmes de la phrase complexe liés à l'analyse syntaxique : les questions des relations entre le syntagme thématisé et les subordonnées (§ 7.12) et les problèmes liés au phénomène d'ellipse (§ 7.13).

7.1 Deux questions centrales pour une définition de la proposition

Nous avons défini jusqu'ici la phrase comme constituée d'une proposition et éventuellement d'un thème. Mais la proposition peut comporter une sous-

structure phrastique. Nous appelons les phrases contenant une sous-structure phrastique des phrases complexes. Néanmoins, nous n'avons pas encore défini exactement ces structures phrastiques que nous appelons proposition.

Suivant le modèle de la grammaire anglaise, Hashimoto a essayé de définir la proposition comme une unité comportant à la fois un sujet et un prédicat, ce qui a été largement critiqué par beaucoup de linguistes comme nous l'avons déjà vu dans la section 6.1.

Cependant, aucun de ces opposants n'a réussi à proposer de définition satisfaisante nous permettant de repérer réellement les propositions. En effet, ils semblent avoir réussi à distinguer la proposition de la phrase sur un plan conceptuel, mais ils ne sont pas arrivés (ou n'ont pas cherché) à trouver d'indications formelles de cette distinction.

Dans les travaux contemporains, Masuoka & Takubo (1992) définissent d'abord la phrase simple comme une phrase constituée d'un seul prédicat et la phrase complexe comme une phrase constituée de plusieurs prédicats. Ils appellent alors propositions les unités de la phrase complexe constituées d'un prédicat et de ses compléments.

Néanmoins, cette définition, très grossière, ne permet pas non plus la détermination exacte de la proposition, dans la mesure où la notion de prédicat est également floue. En effet, toutes les occurrences des mots variables ne peuvent pas être considérées comme des prédicats, noyaux de la proposition. Devant cette difficulté, certains comme Mikami et Garnier ont renoncé à la notion de proposition.

En plus de ce problème de distinction difficile entre les propositions et les syntagmes non-propositionnels à mot variable, il en existe un autre lié à la définition des éléments connecteurs. En effet, pour relier deux constructions à mot variable (y compris les propositions) en japonais, existent deux possibilités : d'une part marquer la connexion par une forme connective du mot variable subordonné et, de l'autre, réaliser la connexion à l'aide d'un élément suivant une forme autonome du mot variable de la subordonnée. Le second problème concerne donc les éléments assurant la connexion dans les structures de ce second type.

Nous nous appliquerons donc ici à élucider la notion de proposition en résolvant ces deux problèmes.

7.2 Premier problème : natures différentes des syntagmes à mot variable

Nous trouvons dans Mikami (1959) un passage dans lequel Mikami souligne l'impossible – ou même insensée – introduction de la notion de proposition dans l'analyse du japonais :

Dans la grammaire anglaise, les séquences de mots sont distinguées en deux catégories :

- *clause* (comprenant l'opposition sujet-prédicat, « proposition ») ;

- *phrase* (sans opposition sujet-prédicat, « syntagme »).
- [...] Comme il n'existe pas d'opposition sujet-prédicat, la distinction entre la proposition et le syntagme est aussi impossible. Ainsi, l'unité « proposition » comprenant l'opposition sujet-prédicat ne concerne pas la grammaire japonaise, et l'unité « syntagme » est donc suffisante, ou plutôt nous sommes obligés de nous contenter de cette unité.

Il semble cependant rester une possibilité, celle de définir la proposition par la présence d'un prédicat conjugué. Mais, lorsque nous essayons d'extraire les propositions avec des mots variables comme repères, nous obtenons des unités paraissant avoir une nature assez différente, d'autant plus que le japonais n'a pas non plus d'opposition entre les formes conjuguées et la forme infinitive pour ses mots variables. Il existe donc aussi bien des syntagmes contenant un mot variable – qui nous semblent très semblables à une phrase –, que d'autres qui sont presque des adverbes.

Si on refuse de renoncer complètement à la notion de proposition, cette absence de repère absolu pour la définition de la proposition pose un véritable problème pour distinguer une phrase complexe d'une phrase simple.

赤ん坊 が 歩く ようになる。

(*akanbô - ga - aruku - yô ni naru*)

(bébé - [ga] - marcher - devenir faire q.c.)

« Le bébé commence à marcher. »

「古都」という本を読んだ。

(*koto - to - iu - hon - wo - yonda*)

("koto" - [to (citation)] - dire - livre - [wo] - lire [passé])

« (J')ai lu le livre intitulé "koto". »

Ces deux exemples contiennent chacun deux mots variables, bien qu'intuitivement nous les considérons plus facilement comme des phrases simples que complexes.

Dans un ouvrage récent sur la phrase complexe de Noda (2002), par exemple, plusieurs critères supplémentaires sont proposés pour définir la proposition : le nombre minimum d'arguments, la présence de la notion de temps, la possession du sens concret, etc. Mais comme les auteurs le soulignent, cela reste une question de probabilité : il existe beaucoup de cas difficiles à juger.

Ainsi, il est très difficile de tracer une frontière entre les phrases simple et complexe, c'est-à-dire de définir la proposition.

7.3 État de l'art des travaux visant à définir la proposition

Nous présentons dans cette section les travaux sur la définition de la proposition. Nous allons aborder tout d'abord l'abandon de cette notion par Mikami (§ 7.3.1) et Garnier (§ 7.3.2), qui ont cherché plutôt à déterminer les différences

de nature entre tous les types de syntagmes contenant un mot variable, puis les tentatives de définition de cette unité par les autres, Minami (§ 7.3.3), Teramura (§ 7.3.4) et Noda (§ 7.3.5). Pour terminer l'exposé, la dernière partie sera consacrée à notre analyse critique de ces travaux antérieurs (§ 7.3.6).

7.3.1 Capacités phrasogénératrices des prédicats selon Mikami

Au lieu de regrouper certains syntagmes contenant un mot variable sous une même étiquette de proposition, Mikami a cherché à définir différentes catégories pour les syntagmes à mot variable.

Études des formes des mots variables

Pour lui, chaque forme de mot variable a une capacité phrasogénératrice de niveau différent et plus la capacité phrasogénératrice du mot variable du syntagme est élevée, plus ce syntagme est autonome, c'est-à-dire proche de la phrase. Il a représenté la capacité phrasogénératrice de différentes formes de mots variables – c'est-à-dire la capacité à constituer en tant que prédicat un constituant aussi phrastique que possible – par une valeur, comme suit (Mikami, 1953) :

Forme neutre	1/4
Forme autonome (emploi déterminant)	1/2
Forme de condition	3/4
Formes autonome et impérative	1

TAB. 7.1 – Capacités phrasogénératrices des formes des mots variables

Comme nous l'avons déjà vu dans la section 6.2.2, Mikami distingue quatre types de connexions dans la phrase japonaise. Mikami considère la première, forme neutre, comme produisant une connexion de style simple, la forme de condition celle de style complexe souple et les formes autonome et impérative celles de style complexe dur. La forme autonome en emploi déterminant ne réalise pas de connexion avec le syntagme à mot variable, mais elle produit une connexion particulière appelée déterminante.

Comme l'affirme Mikami (1953), cette idée de classification des formes des mots prédicatifs est inspirée des travaux de Mio (1942) qui a présenté le résultat d'études, réalisées avec un corpus, sur la possibilité de transformation en forme polie des mots prédicatifs précédant une particule conjonctive.

Études des emplois des formes des mots variables

Nous constatons cependant une évolution de cette catégorisation au cours des années de recherche de l'auteur. En effet, chaque forme, en particulier la forme autonome, produit différents types de connexions selon son emploi effectif.

Reprenons les exemples de Mikami (1955). Les trois phrases suivantes sont toutes constituées de la partie exprimant la cause « je me suis réveillé trop tard » introduite par trois connecteurs introducteurs de cause différents (*tameni*, *node*, *kara*), suivie du prédicat principal « (j')étais en retard ».

寝坊した (<i>nebôshita</i>) (se réveiller trop tard [passé])	ために (<i>tame ni</i>) ので (<i>node</i>) から (<i>kara</i>)	遅刻した (<i>chikoku shita</i>) (être en retard [passé])
--	---	--

En dépit de l'identité de la forme, la construction que constitue le premier verbe « se réveiller trop tard » avec son introducteur produit différents types de connexions selon le connecteur. Cette différence est claire avec un test d'enchâssement dans la structure déterminante. Quand on rajoute « *koto wa sukunai* (le fait - [wa] - peu) » à la fin des trois phrases ci-dessus, la phrase avec *tameni* est totalement enchâssée dans la structure déterminante :

nebôshita tameni chikokushita *koto* *wa sukunai*
 « Le fait que je sois en retard parce que je me suis réveillé trop tard est rare »

En revanche, celle avec *kara* sort complètement de l'enchâssement et constitue une phrase sémantiquement incorrecte :

**nebôshita kara* *chikokushita* *koto* *wa sukunai*
 « *Le fait que je sois en retard est rare, parce que je me suis réveillé trop tard »

La phrase avec *node* sort généralement en dehors de l'enchâssement comme la phrase avec *kara*, mais elle peut être interprétée comme la phrase avec *tameni*. Le résultat de ce test est donc comme suit :

- Verbe à la forme autonome (emploi déterminant) + *tame ni*
= connexion de style simple ;
- Verbe à la forme autonome (emploi conclusif) + *node*
= connexion de style complexe souple ;
- Verbe à la forme autonome (emploi conclusif) + *kara*
= connexion de style complexe dur.

Dans Mikami (1959), il présente la catégorisation non pas des formes mais des emplois des formes (cf. tableau 7.2 (voir page suivante)), en cinq ensembles : emplois infinitif, neutre, de condition, conclusif non final et conclusif final. L'emploi conclusif non final désigne le cas où la forme conclusive ne se situe pas en fin de phrase mais qu'elle est suivie d'une particule conjonctive avec laquelle elle constitue une construction phrastique dépendant syntaxiquement du prédicat principal.

Emploi infinitif	0
Emploi neutre	1/4
Emploi de condition	1/2
Emploi conclusif (non final)	3/4
Emploi conclusif (final)	1

TAB. 7.2 – Capacités phrasogénératrices des emplois des mots variables

L'emploi de condition est le seul emploi de la forme de condition et c'est la seule forme ayant l'emploi de condition. Il en va de même pour la forme neutre et l'emploi neutre. L'infinitif est un des emplois de la forme autonome, forme qui a en outre les emplois conclusifs final et non final ainsi que l'emploi déterminant. Les formes volitive et impérative ont également l'emploi conclusif mais la forme impérative n'a pas d'emploi conclusif non final.

Lexicalisation des formes des mots variables

Par ailleurs, il parle également de la lexicalisation de certaines formes des mots variables. C'est le cas de l'exemple présenté dans la section précédente décrivant la difficulté de la définition (§ 7.2) :

「古都」と いう 本 を 読んだ。
 (koto - to - iu - hon - wo - yonda)
 ("koto" - [to (citation)] - dire - livre - [wo] - lire [passé])
 « (J')ai lu le livre intitulé "koto". »

Dans cet exemple, le verbe « *iu* (dire) » a perdu la véritable fonction de prédicat et constitue avec la particule *to* qui le précède, un connecteur introduisant des éléments déterminants. Il cite quelques conditions empêchant cette lexicalisation, comme par exemple : possession d'un ou plusieurs compléments, avec particules de cas ou non, et ce notamment avec un acteur animé ; présence du complément en *ga*.

7.3.2 Les trois classes des syntagmes à mot variable de Garnier

Garnier (1982) cherche également à saisir la nature différente des syntagmes à mot variable.

Deux nouvelles unités

Renonçant à toutes les unités « classiques » mal définies, elle définit tout d'abord ses propres unités syntaxiques dont les deux principales, supérieures au syntagme minimal et inférieures à la phrase :

- séquence : unité formée par un mot prédicatif et ses compléments, y compris le thème en *wa* ;
- unité de discours : toute séquence qui n'a aucune contrainte pour les trois possibilités maximales : la réception de suffixes fonctionnels, la réception de segments en statut de complétisation et la factorisation (ou thématisation) de segments.

Elle étudie ensuite les phrases à deux (ou plusieurs) mots variables. En analysant combien de séquences et unités de discours elles contiennent, elle réalise une catégorisation de chaque forme des mots variables.

Les trois classes des syntagmes à mot variable

Ses travaux catégorisent finalement les syntagmes à mot variable en trois types¹ :

1. syntagme qui ne peut constituer une séquence fonctionnant comme un élément de phrase qu'avec un autre syntagme à mot variable en fin de phrase :

*sôshite - sanbo wo - tabeyô **to** - shimashita*

(et - Sanbo - manger - (faire))

« Et il s'apprêta à avaler Sanbo. »

2. syntagme qui peut constituer à lui seul une séquence, mais qui ne peut constituer une unité de discours qu'avec une autre séquence qui le suit :

*atarashii kôsha ga - dekita - **toki** - minna - taihen - yorokobimashita*

(nouvelle école - être fini - moment - tous - très² - se réjouir)

« Quand la nouvelle école fut finie, tout le monde se réjouit. »

3. syntagme qui peut constituer à lui seul, non seulement une séquence, mais aussi une unité de discours :

*zepetto ga - tomemashita - **ga** - pinokio wa - kikimasen deshita*

(Gepetto - arrêter - [ga] - Pinocchio - ne pas écouter)

« Gepetto voulut l'arrêter mais Pinocchio ne l'écouta pas. »

Syntagmes à mot variable qui ne sont pas des séquences

Étant donné la définition, la différence entre la séquence et l'unité de discours est nette, mais celle entre la séquence et le segment à mot variable qui ne peut pas constituer tout seul un élément de phrase est plus difficile à saisir.

Il y a deux types de syntagmes à mot variable de ce genre :

- hyper-séquence : constituée de deux syntagmes à mot variable qui se complètent ;

¹Exemples et traductions tirés de Garnier (*Ibid.*). La transcription de la particule *wo* est « o » dans l'original. Le mot variable concerné est souligné et le connecteur ou le substantif cheville est mis en gras par nous-mêmes, en modifiant le style de l'original.

²Absence de traduction dans l'original.

- semi-séquence incluse : ayant la possibilité de posséder ses propres compléments – avec tout de même quelques restrictions – qui ne peut fonctionner cependant qu'en tant qu'élément d'une séquence.

Hyper-séquence Ce sont deux syntagmes à mot variable contigus. Le premier comporte des compléments alors que le deuxième, verbe (ou mot prédicatif) support, sert à exprimer le temps, l'aspect et la modalité. Ces deux syntagmes se complètent et ne peuvent fonctionner en tant qu'élément de phrase qu'ensemble.

1. *sôshite - sanbo wo - tabeyô to - shimashita*
(et - Sanbo - manger - (faire))
« Et il s'apprêta à avaler Sanbo. »
2. *okane wo - ikura - haraeba - yoi deshô ka*
(argent - combien - payer - bien c'est)
« Combien dois-je payer ? »

Le deuxième mot variable restant toujours un syntagme minimal – donc aucune insertion possible entre les deux mots variables –, la détermination des hyper-séquences est relativement aisée.

Semi-séquence incluse Il s'agit de syntagmes à mot variable avec leurs compléments, qui sont enchâssés à l'intérieur d'une séquence et qui constituent un élément de cette dernière.

1. *yagate - morikara - dongajisan ga - taiko wo - narashi nagara - arawareta*³
(bientôt - forêt - le vieux Donga - tambour - faire résonner - se montrer)
« Bientôt, le vieux Donga, tapant sur son tam-tam sortit de la forêt. »
2. *beruto no ue de - jidôsha ga - sukoshizutsu - kumitaterarete - dandan - dekigatte ikimasu*
(chaîne sur - voiture - morceau par morceau - être assemblé - peu à peu - se terminer)
« Les voitures se construisent petit à petit, montées pièce par pièce sur la chaîne. »

Ce sont des syntagmes suivis d'un mot connecteur *nagara* (cf. ex. 1) ou *mama*, ou terminés par une forme adverbiale (cf. ex. 2).

L'analyse de Mikami pour ce type de phrase diffère. Il interprète les compléments antérieurs au premier mot variable comme dépendant de ce dernier. Mais la connexion que produit ce dernier est telle que les compléments franchissent ce premier mot prédicatif et dépendent à nouveau du deuxième mot variable. C'est la définition même de la connexion de style simple.

Malgré cette différence d'analyse, la conclusion de ces deux auteurs sur l'autonomie de ces mots variables semble très proche : la capacité phrasogénératrice de ces mots variables (ou les formes de ces mots variables) est très faible.

Pour l'analyse des compléments communs, nous suivons le modèle de Mikami. Ce choix provient cependant d'une raison pratique d'analyse automatique⁴

³L'encadrement est absent dans l'original, mais rajouté pour présenter clairement la structure enchâssée dont parle Garnier.

⁴Tous les compléments antérieurs au premier mot variable – sauf bien évidemment les éléments extérieurs tels que le thème ou les éléments indépendants en tête de phrase – dépendent soit de ce

et la détermination de leur appartenance du point de vue linguistique nécessiterait une étude beaucoup plus poussée que nous ne pouvons pas réaliser dans le cadre de la présente thèse.

7.3.3 Les propositions subordonnées de Minami

Dans ses travaux destinés à définir quatre couches constituant la phrase japonaise – par l'étude de la possibilité d'apparition ou non des éléments dans différentes structures subordonnées –, Minami définit tout d'abord les propositions subordonnées.

Définition des subordonnées

En expliquant que la subordonnée est un type d'unité ayant une structure proche de la phrase et qui apparaît dans une phrase, il énumère dans Minami (1993) les unités qu'il considère comme des subordonnées :

1. unité se terminant par un mot variable à la forme neutre :

- 大臣は その時には Aさんであり、また、事務次官は Bさんでありました。

(*daijin wa - sonotoki ni wa - A san deari - mata - jimujikan wa - B san dearimashita*)

(ministre [*wa*] - à cette époque [*wa*] - M./ Mme. A [copule] - de plus - vice-ministre [*wa*] - M./Mme. B [copule passé])

Le ministre était, à cette époque, Monsieur A et le vice-ministre était Monsieur B.

2. unité se terminant par une particule conjonctive telle que :

が (*ga*), から (*kara*), けれど (も) (*keredo(mo)*), し (*shi*), たら (*tara*), だら (*dara*), て (*te*), で (*de*), ては (*tewa*), では (*dewa*), ても (*temo*), でも (*demo*), と (*to*), ながら (*nagara*), なら (*nara*), ので (*node*), のに (*noni*), ば (*ba*).

- よく かきまぜながら、煮る。

(*yoku - kakimaze - nagara - niru*)

(bien - mélanger [simultanéité] - mijoter)

(On le) mijote en mélangeant bien.

3. unité se terminant par certains substantifs formels⁵ tels que :

あげく (*ageku*), ため (*tame*), ところ (*tokoro*)

- 部品を 取り替えてみた ところ、エンジンが 作動する ようになった。

(*buhin wo - torikaete - mita - tokoro - enjin ga - sadōsuru - yōninatta*)

(pièce détachée [*wo*] - changer - pour voir - [sub formel *tokoro*] - moteur - fonctionner -

dernier uniquement, soit également du mot variable postérieur. Mais ils ne dépendent pas forcément tous du deuxième mot prédicatif, ce qui complique considérablement l'analyse automatique.

⁵形式名詞 (*keishikimeishi*, substantifs formels) : il s'agit d'un type de substantif qui, « tout en ayant des caractéristiques nominales, n'a que très peu de sens et n'est donc utilisé qu'avec des éléments déterminants » (Masuoka & Takubo, 1992).

devenir [passé])

Quand j'ai changé la pièce détachée pour voir, le moteur s'est remis à fonctionner à nouveau.

Après avoir ainsi défini les propositions subordonnées, il les classe en trois groupes A, B et C définis sur la base de sa théorie de la structure en couches de la phrase japonaise, que nous avons déjà vue dans la section 6.2.2.

Exception des syntagmes à mot variable non-propositionnels

Bien qu'appartenant à l'une de ces classes, les éléments constituant une locution figée, ou lexicalisés comme s'ils étaient des adverbes, sont considérés comme des exceptions.

I. constituant une locution :

a) précédant *なる* (*naru*, devenir) ou *する* (*suru*, faire) :

– 手 が 動かなく (なる)

(*te - ga - ugokanaku - (naru)*)

(main - [ga] - bouger [négation] - (devenir))

Les mains ne bougent plus.

b) précédant d'autres expressions :

– ~て/ても (いい、かまわない、...)

(*X te/temo - (ii, kamawanai, ...)*)

(*X* [forme neutre en *te*] / même si *X* - (bien, pas gênant, ...))

Vous pouvez faire *X* / cela ne me gêne pas que tu fasses *X*, ...

II. lexicalisés :

– 飛んで (帰る)

(*tonde - kaeru*)

(voler - rentrer)

Rentrer à toute vitesse.

D'autres constructions phrastiques non subordonnées

Il cite d'autres constructions phrastiques qu'il ne considère pas comme des subordonnées, telles que les constructions se terminant par une impérative, les phrases incidentes, les constructions constituant des éléments indépendants de la grammaire scolaire, les discours rapportés ou les constructions phrastiques déterminantes.

7.3.4 Les deux types de phrases simples de Teramura

Teramura (1982a) fait la remarque que, malgré une décision contraire vis-à-vis de la notion de proposition, Mikami et Minami sont arrivés finalement à une

conclusion semblable⁶. D'après Teramura, la catégorie de type simple de Mikami correspond aux subordonnées du groupe A de Minami, celle de style complexe souple aux subordonnées du groupe B et enfin celle de style complexe dur aux subordonnées du groupe C.

Teramura trace ensuite la limite de la proposition, *grosso modo* entre les groupes A et B de Minami et considère seulement les subordonnées appartenant aux groupes B et C comme des propositions subordonnées.

Par ailleurs, certaines propositions du groupe B de Minami sont également considérées comme non-propositionnelles lorsqu'elles partagent des compléments avec la proposition racine. En d'autres termes, les constructions se terminant par un mot variable à la forme neutre (adverbiale) doivent comporter plus d'un complément dépendant syntaxiquement directement de son prédicat et dans Teramura (1991), l'auteur parle plus précisément du partage, non pas de n'importe quel complément, mais du complément en *ga*. Cette condition supplémentaire montre qu'un des critères formels de Teramura pour reconnaître le statut de proposition était la présence explicite du complément en *ga*.

Ainsi, Teramura (1982a) définit deux types de phrases simples :

– Type 1 :

1. phrase contenant un seul prédicat qui s'associe avec ses compléments et qui se termine par une forme conclusive ;
2. phrase terminée par un verbe à la forme neutre suivi d'un second verbe exprimant l'aspect.

– Type 2 :

1. phrase contenant, en plus du prédicat, un syntagme terminé par un verbe à la forme en *te* qui fonctionne comme un adverbe ;
2. phrase contenant, en plus du prédicat, un syntagme terminé par un verbe à la forme neutre, qui partage son/ses complément(s) avec le prédicat principal (dans Teramura (1991), précisément le complément nominatif).

7.3.5 Les frontières entre phrase simple et phrase complexe selon Noda

Nous pouvons également trouver dans Noda (2002) une discussion très riche sur la proposition. Noda y présente trois typologies de la proposition, dont l'une est réalisée selon le niveau de ressemblance à la phrase de la structure de proposition.

La catégorisation est réalisée par un test examinant la présence possible ou non dans la proposition de six types d'éléments :

1. éléments de voix ;

⁶Minami (1993) fait lui-même la même remarque en disant que ses trois catégories des groupes A, B et C correspondent aux classes proposées par Mikami, respectivement les styles simple, complexe souple et complexe dur. Mais, il doute que leur classification soit tout à fait identique et présente quelques exemples pour lesquels leurs analyses divergent.

2. éléments d'aspect ;
3. éléments de négation ;
4. éléments de temps ;
5. éléments de modalité orientée vers le contenu ;
6. éléments de modalité orientée vers l'interlocuteur.

Les propositions comportant les éléments d'un niveau donné peuvent contenir tous les éléments des niveaux inférieurs. En d'autres termes, les propositions contenant les éléments du niveau 5, peuvent comporter également des éléments des niveaux 1 à 4.

Suite à ce test, les propositions sont distinguées en six types selon la possibilité d'apparition de ces éléments. Les syntagmes à mot variable ne pouvant comprendre aucun de ces éléments sont considérés comme des syntagmes non-propositionnels.

L'auteur compare son classement avec la catégorisation de Minami comme suit :

Propositions pouvant comporter	Catégorisation de Minami
des éléments 1	groupe A
des éléments 2	groupe B
des éléments 3	groupe B
des éléments 4	groupe B
des éléments 5	groupe C

7.3.6 Analyse critique

Tout comme Teramura a constaté une correspondance entre les travaux de Mikami et ceux de Minami, nous pouvons également reconnaître certains points communs dans les études de Garnier. Le tableau 7.3 page ci-contre est un comparatif des typologies des syntagmes à mot variable de ces trois chercheurs.

Frontière de la proposition

Nous pouvons maintenant tracer, sur le modèle de Teramura, la limite de la proposition entre la catégorie de style simple et celle de style complexe de Mikami. Cette frontière correspond à celle entre le groupe A et le groupe B de Minami et à celle entre la catégorie « 1 séquence - 1 unité de discours » et la catégorie « 2 séquences - 1 unité de discours » de Garnier.

Mais quels sont les critères concrets pour reconnaître les syntagmes non-propositionnels, qui n'appartiennent pas à la catégorie limitée par la frontière que nous venons de tracer ?

Minami se contente d'énumérer les exemples types de syntagmes à mot variable qu'il ne considère pas comme des propositions. Le critère de la possibilité

Comparaison des analyses des syntagmes à mot variable

Garnier	Mikami	Minami
1 séquence – 1 unité de discours ・～たりする ・～ようとする ・（連用）に行く ・（連用）ながら ・（連体）ままV／にしておく ・（連体）ようにする／なる ・（仮定）ばよいでしょう（か）／ばならない ・（連用）たら／～てよいでしょう（か） ・（終止）といいです／いけないです ・（終止）からです ・～ていけない／ならない ・～て₁／（連用-手段）₁V	Capacité phrasogénératrice nulle ・ Formes lexicalisées ・ Emploi infinitif (forme autonome)	Non objet de l'examen (考察対象外) ・ なる／するに続くもの ・ ～たら（いい、こまる、だめだ） ・ ～て（いい、かまわない） ・ ～ては（いけない、こまる、だめだ） ・ ～ても（いい、かまわない、だめだ） ・ ～と（いい、こまる） ・ ～なければ（ならない、だめだ） ・ ～ば（いい、さいわいだ） ・ 等
	Style simple (単式) ・ Emploi neutre (forme neutre)	A ・ （連用）ながら ・ ～つつ ・ ～て₁／連用₁ ・ 連用形反復（なめなめ）
2 séquences – 1 unité de discours ・ （連体）とき、あと、まで、まえ、ころ（等）、ほど ・ （談話）と／か ・ （連体）ため、ように、には ・ （仮定）ば／たら ・ （終止）と ・ （連体）なら、のに、ので ・ ～て_{2,3}／連用_{2,3}V ・ ～ても、てから	Style complexe souple (複式・軟式) ・ Emploi de condition (forme de condition) ・ Emploi conclusif non final (forme autonome)	B ・ ～て₂／連用₂ ・ ～と ・ ～ながら（逆接） ・ （仮定）ば／たら ・ （連体）なら、のに、ので ・ ～て₃／連用₃ ・ ～ず（に） ・ ～ないで ・ （？）形式名詞で終わる句
2 séquences – 2 unités de discours ・ （終止）し、けれども、が、から ・ ～て₄／連用₄V	Style complexe dur (複式・硬式) ・ Emploi conclusif final (forme autonome)	C ・ （終止）し、けれど、が、から ・ ～て₄／連用₄
	Connexions particulières ・ Style flottant (遊式) ・ Style fermé (ト式)	Élément semblable à la phrase (文に似た部分) ・ 命令形、已然形で終わるもの ・ 挿入的な性格のもの ・ 独立部 ・ 引用句 ・ 連体修飾句

TAB. 7.3 – Comparaison des typologies des syntagmes à mot variable

d'insertion d'un élément entre deux mots variables proposé par Garnier et le critère de la possibilité d'ajout au minimum d'un élément de voix proposé par Noda ne sont efficaces que pour certains syntagmes.

Quatre types de syntagmes à mot variable non-propositionnels

En effet, il existe quatre types de syntagmes à mot variable non-propositionnels à reconnaître :

1. mots variables supports ou auxiliaires tels que :

V なければならぬ

(V - *nakere* - ba - *naranai*)

(V - [négation] - [condition] - ça ne se fait pas)

« il faut V »

2. syntagmes à mot variable avec un complément lexicalisés tels que :

N に関して

(N *ni* - *kanshite*)

(N [ni] - concerner)

« concernant N »

3. syntagmes avec le mot variable à une forme neutre non-propositionnels tels que :

落ち着いて N に取り組む

(*ochitsuite* - N *ni* - *torikumu*)

(se calmer - N [ni] - s'attaquer)

« s'attaquer calmement à N »

4. verbes composés tels que :

N1 に N2 を 移し 替える

(N1 *ni* - N2 *wo* - *utsushi* - *kaeru*)

(N1 [ni] - N2 [wo] - déplacer - changer)

« transférer N2 vers N1 ».

Pour le troisième type de syntagme, Mikami, Garnier et Teramura considèrent que le partage des compléments est un des critères. Mais, dans la phrase japonaise où tous les éléments peuvent être omis selon le contexte, le partage des compléments est un phénomène très courant et ne peut pas être un test efficace pour distinguer la nature d'un syntagme. Mikami affirme qu'il est important de distinguer les cas où les compléments sont partagés par les deux mots prédicatifs du fait du caractère grammatical de la connexion que produit le premier mot variable, d'une part, et les cas où les compléments sont partagés selon le contexte, de l'autre. Mais, il ne parle pas de la façon de les distinguer. Garnier explique que les deux actions décrites par les deux syntagmes à mot variable non-propositionnels (qui ne constituent pas tout seuls la séquence, selon sa terminologie) relèvent toujours d'un actant unique.

Afin de discriminer de manière automatique ces syntagmes à mot variable démunis du statut de proposition, il nous faudrait déterminer des critères concrets et précis.

7.4 Critères de détermination des syntagmes à mot variable non-propositionnels

Examinons maintenant pour ces quatre types de syntagmes à mot variable non-propositionnels les critères concrets de détermination possibles.

7.4.1 Mots variables supports ou auxiliaires

V なければならぬ

(V - *nakere* - *ba* - *naranai*)

(V - [négation] - [condition] - ça ne se fait pas)

« il faut V »

Ce sont des mots variables suivant, directement ou presque directement par l'intermédiaire d'un connecteur, un autre mot variable de manière à constituer un verbe composé ou une locution prédicative. Ce sont les hyper-séquences de Garnier. Nous pouvons donc employer le critère proposé par cette dernière et définir la règle permettant de constituer préalablement la liste des mots variables supports comme suit :

Règle 1 (mots variables supports) Lorsqu'un mot variable suit toujours un autre mot variable quelconque à une forme donnée et qu'il ne peut prendre aucun complément qui lui soit propre, il est considéré comme mot variable support ou auxiliaire et constitue le prédicat avec le mot variable qui le précède.

7.4.2 Syntagmes à mot variable avec un complément lexicalisés

N に関して

(N *ni* - *kanshite*)

(N [ni] - concerner)

« concernant N »

Comme Mikami le signale, il existe beaucoup de constructions comportant un mot variable – essentiellement un verbe – éventuellement suivi d'un connecteur, qui constituent une unité semblable à une particule. Mikami les appelle particules de cas composées. On ne trouve pas de discussion sur ces éléments dans les travaux de Garnier. C'est sans doute dû à la nature de son corpus, des manuels scolaires de l'école primaire. En effet, ce sont souvent des expressions utilisées plutôt à l'écrit et surtout dans un style académique ou journalistique.

La détermination de ces syntagmes à mot variable lexicalisés que nous traitons ici n'est en fait que la partie visible de l'iceberg : elle est liée à la définition d'un autre type d'unité, 連語 (*rengo*) que nous traduisons par « locution ».

Problèmes des *rengo*

La définition de ce terme est assez floue : dans un sens large, il désigne les constructions à plusieurs mots y compris les syntagmes et les propositions ; dans un sens strict, il désigne les unités constituées de plusieurs mots équivalentes à un mot. Dans ce dernier sens, se posent des problèmes de distinction avec le terme « mot composé », 複合語 (*fukugô-go*), désignant les mots constitués de plusieurs mots simples. Aujourd'hui, le terme *rengo* est également utilisé pour la traduction du terme « collocation ».

Nous utilisons dans la présente étude le terme locution pour désigner l'unité constituée de plusieurs mots ayant perdu plus ou moins leurs sens initiaux et ne constituant un sens stable équivalent à celui d'un mot qu'ensemble.

Dans les dictionnaires, figurent généralement les locutions. Étant donné qu'il est impossible d'énumérer toutes les locutions dans le sens large du terme, ce terme est utilisé sans doute dans un sens plus ou moins proche de ce que nous venons de définir. Mais, comme le montre Yazawa (1995), les entrées étiquetées comme des locutions varient extrêmement un dictionnaire à l'autre. De plus, même à l'intérieur d'un dictionnaire, on constate difficilement une cohérence entre les entrées étiquetées comme des locutions, les critères de cet étiquetage ne semblant pas être bien définis.

Critères de détermination des locutions à mot variable

Nous nous occupons dans la présente thèse uniquement des locutions qui nous concernent, les locutions à mot variable, et définissons les critères de leur détermination.

Le critère de Noda qui définit la possibilité d'inclusion d'éléments de voix comme condition minimale requise pour la proposition semble utile au premier abord. Mais, dans la mesure où l'acception des éléments de voix dépend également de la nature lexicale du verbe, il faut utiliser ce test avec prudence.

Nous n'avons pas trouvé jusqu'ici d'autres pistes sur ce sujet dans la littérature. Nous avons rencontré seulement un passage de Mikami que nous avons déjà présenté précédemment (cf. § 7.3.1) qui parle de quelques conditions empêchant la lexicalisation des constructions. Il y dit que les syntagmes verbaux avec le complément en *ga* ne subissent jamais de lexicalisation. C'est une condition nécessaire mais non suffisante. Pour trouver des conditions nécessaires et suffisantes, il faudrait plus d'études avec l'analyse de beaucoup d'exemples. Nous définissons de manière provisoire les règles suivantes pour reconnaître les locutions à mot variable :

Règle 2 (locutions à mot variable) Lorsqu'un mot variable à une forme non autonome – notamment une forme neutre – prend un seul complément (qui n'est pas complément en *ga*) dont le syntagme nominal joue une fonction syntaxique vis-à-vis du prédicat dont le mot variable dépend syntaxiquement, et que le mot variable ne partage pas le complément en *ga*, implicite ou explicite, de ce prédicat, la construction constituée de ce mot variable et la particule introduisant son complément sont considérées comme une locution assimilée à la particule.

Exemple d'application de la règle

Considérons l'exemple suivant :

A社は、企業買収を めぐり 虚偽の 発表を した

(*A sha wa - kigyôbaishû wo - meguri - kyôgi no - happyô wo - shita*)

(société A [wa] - rachat d'entreprises [wo] - faire le tour - faux - déclaration [wo] - faire [passé])

« La société A a fait de fausses déclarations concernant ses rachats d'entreprises »

Le syntagme « *kigyôbaishû wo meguri* » n'a pas de complément en *ga*, et le complément nominatif du prédicat principal, implicite car thématique « *A sha (ga)* », n'est pas non plus son complément nominatif. On doit considérer que le verbe « *meguri* » ne possède pas, purement et simplement, de complément nominatif. En revanche, on peut analyser que le SN « *kigyôbaishû* » joue le rôle de complément accessoire vis-à-vis du prédicat principal. On considère donc « *wo meguri* » comme une locution équivalente à une particule.

Cas délicats

Pour certaines expressions, il est difficile de juger si elles ne partagent pas le complément en *ga* avec le prédicat. Dans ce cas, on applique la règle d'extension suivante.

Règle 2' (règle d'extension pour les locutions) Après avoir mis le mot variable à une forme conclusive, si la séquence créée ne peut jamais apparaître toute seule dans un texte quelconque, cette séquence est considérée comme une locution.

Par exemple,

Aは、Bに 対して 結果を 報告した

(*A wa - B ni - taishite - kekka wo - hôkoku shita*)

(A [wa] - B [ni] - faire face - résultat [wo] - rapporter [passé])

« A a rapporté le résultat à B »

Le complément nominatif du verbe « *taishite* » peut être « *A (ga)* » qui est le complément nominatif implicite du prédicat principal. On transforme alors le verbe en forme conclusive pour constituer une phrase autonome, i.e. *taisuru*. Mais, « *A ga B ni taisuru* » ne peut pas être considérée comme une phrase indépendante –

du moins dans le japonais moderne. On considère donc que « *ni taishite* » est une locution fonctionnant comme une particule.

Locutions variables

Certaines locutions, ainsi déterminées, sont moins figées que d'autres, conservant alors leurs propres variations.

Par exemple, supposons que l'expression *N* に関して (*N ni kanshite*) vérifie les conditions présentées précédemment et qu'elle soit considérée comme une locution. Cette expression conserve sa propre variation et apparaît également sous forme de *N* に関し (*N ni kanshi*) ou *N* に関する (*N ni kansuru*) dans la structure déterminante. En revanche, l'expression *N* について (*N ni tsuite*) n'a que la forme *N* につき (*N ni tsuki*).

Règle 2'' (locutions variables) Les expressions conservant leurs propres variations sont considérées comme des locutions sous l'ensemble de leurs formes.

L'ensemble des règles 2, 2' et 2'', primitives et plutôt expérimentales, doit être bien évalué avec les résultats de l'analyse automatique afin d'être amélioré.

7.4.3 Syntagmes avec le mot variable à une forme neutre non-propositionnels et verbes composés

落ち着いて N に 取り組む

(*ochitsuite* - *N ni* - *torikumu*)

(se calmer - *N* [*ni*] - s'attaquer)

« s'attaquer calmement à *N* »

N1 に N2 を 移し 替える

(*N1 ni* - *N2 wo* - *utsushi* - *kaeru*)

(*N1* [*ni*] - *N2* [*wo*] - déplacer - changer)

« transférer *N2* vers *N1* »

Contrairement aux deux types de syntagmes présentés précédemment, pour ceux que nous traitons ici, il est beaucoup plus difficile voire impossible de constituer des listes préalables car les possibilités sont beaucoup plus importantes. Il est donc nécessaire, pour ces syntagmes, de définir des règles permettant de les reconnaître de manière dynamique lors de l'analyse.

Examinons d'abord le cas des syntagmes avec le mot variable à une forme neutre. Pour la détermination de la nature des syntagmes à mot variable terminés par une forme neutre, il existe un critère proposé par Mikami et utilisé par Tera-mura : la présence du complément en *ga*. Nous adoptons une règle un peu plus large : la présence d'au moins un complément. En effet, aussi bien dans les travaux de Minami que dans ceux de Garnier, les emplois de forme neutre se distinguent en plusieurs emplois et chaque emploi appartient à une catégorie fort différente.

Ne connaissant pas pour l'instant les critères formels permettant de distinguer ces différents emplois, nous adoptons une règle plus large mais qui semble efficace.

Il faut tenir compte des deux cas :

1. V_1 (forme neutre) + Complément(s) + V_2
2. Complément(s) + V_1 (forme neutre) + V_2

Dans le premier cas, le mot variable V_1 n'a aucun complément, tandis que dans le second, c'est le mot variable V_2 qui apparaît sans aucun complément.

La règle pour le second cas permet également de reconnaître les verbes composés dans lesquels aucun complément n'apparaît entre les verbes simples constituants.

Par ailleurs, la règle pour le premier cas permet de détecter ceux qui sont lexicalisés du type, 続いて (*tsuzuite*, puis) ou 併せて (*awasete*, en même temps) – qui fonctionnent comme des mots de liaison – lorsqu'ils sont traités comme des verbes par l'analyseur morphologique.

Règle 3 (pour l'identification dynamique) Lorsqu'un syntagme terminé par un mot variable à une forme neutre ne comprend aucun complément, il est considéré comme un syntagme non-propositionnel dépendant du prédicat apparaissant à une position postérieure. Lorsqu'un syntagme terminé par un mot variable conclusif ne comprend aucun complément et qu'il est précédé directement par un mot variable à une forme neutre, il est considéré comme constituant un mot variable composé avec celui qui le précède directement.

7.5 Nos définitions des unités : proposition et sous-phrase

Nous définissons maintenant la proposition ainsi que quelques autres unités en tenant compte de tout ce que nous avons étudié jusqu'ici. Mais, avant de présenter nos définitions, nous allons passer brièvement en revue les problèmes liés à la notion de proposition principale.

Discussion sur la notion de proposition principale

Après avoir défini la phrase complexe comme une phrase comprenant plusieurs prédicats qui constituent les propositions, Masuoka & Takubo (1992) distinguent les propositions en deux grands types : proposition principale et proposition connectée. La proposition principale est celle qui régit, en se mettant en fin de phrase, l'ensemble de la phrase. La proposition connectée est celle qui n'est pas une proposition principale mais qui lui est reliée par une certaine relation.

Noda récuse tout d'abord la définition usuelle de la phrase complexe comme construction réalisée par assemblage de plus de deux propositions ayant une forme semblable à la phrase. Il considère comme phrase complexe la phrase dans laquelle une partie de la phrase simple devient une proposition lors d'un développement. Ainsi, il rejette la notion de proposition principale et appelle la struc-

ture constituée autour du prédicat principal, phrase principale. Les constructions phrastiques jouant le rôle d'élément dans une phrase principale sont appelées propositions.

Comme dans les études linguistiques du français, du fait de l'inexactitude de l'analyse classique totalement linéaire, nous évitons le terme de « proposition principale » et utilisons le terme de « proposition racine ».

Définitions

Proposition Nous appelons proposition toute construction à mot variable satisfaisant la condition de capacité régissant des compléments, à l'exception des constructions terminées par une forme neutre sans aucun complément. Nous appelons par ailleurs **prédicat** uniquement les mots variables capables de constituer autour d'eux une proposition.

Proposition racine Nous appelons les propositions se terminant par un mot variable à une forme conclusive qui ne dépendent syntaxiquement d'aucun élément de phrase, propositions racines.

Proposition subordonnée Les propositions subordonnées sont des propositions suivies d'un connecteur (particules conjonctives, substantifs formels) ou terminées par une forme connective.

Sous-phrase Par ailleurs, nous appelons sous-phrases les constructions – dépendant du prédicat principal de la phrase – constituées d'une proposition et d'un thème ainsi qu'éventuellement d'éléments externes précédant le thème et qui n'entrent pas dans la proposition.

Sous-structure subordonnée Nous appelons sous-structures subordonnées (ou plus simplement subordonnées) l'ensemble des propositions subordonnées et des sous-phrases. Les propositions subordonnées peuvent également être appelées subordonnées, tant que la confusion avec le sens plus large ne gêne pas l'interprétation.

Connecteur souple et connecteur dur Suivant la typologie de Mikami, nous posons comme hypothèse que les constructions produisant une connexion de type complexe souple peuvent comporter seulement des propositions, mais pas de sous-phrase. En revanche, les constructions produisant une connexion de type complexe dur peuvent comporter des sous-phrases. Nous appelons le connecteur susceptible de constituer les constructions produisant une connexion de type complexe souple, connecteur souple, et celui susceptible de constituer les constructions produisant une connexion de type complexe dur, connecteur dur.

7.6 Second problème : catégorisation imprécise des éléments suivant une forme conclusive du mot variable

7.6.1 Description du problème

Teramura (1982a) signale que la sous-structure dans une phrase peut être constituée par des mots variables, non seulement à une forme connective, mais aussi à une forme conclusive (voir aussi § 7.9.1). Il énumère quatre cas où le mot variable à une forme conclusive constitue le prédicat non principal de la phrase :

1. lorsqu'il constitue une subordonnée déterminant le substantif qu'il précède ;
2. lorsqu'il constitue une subordonnée déterminant un substantif formel et constituant ensemble une unité équivalente à un substantif ;
3. lorsqu'il constitue, en étant suivi par une particule conjonctive, une subordonnée conjonctive ;
4. lorsqu'il constitue, en étant suivi par la particule *to*, une subordonnée de citation.

Outre ces quatre possibilités, il existe en réalité une cinquième possibilité de construction des subordonnées conclusives : constitution avec une particule adverbiale. L'absence de description de cette dernière possibilité peut être un simple oubli, mais elle pourrait aussi être due à la définition généralement floue des catégories concernées : particules conjonctives, particules adverbiales, particule *ka*, substantifs formels, ou encore auxiliaires.

Comme nous l'avons vu dans la section 5.3.2, ces catégories constituent une zone totalement floue et désordonnée. Mais, le mot prédicatif étant incapable d'assurer lui-même la connexion par sa forme, dans le cas des subordonnées terminées par un mot prédicatif à une forme autonome conclusive, c'est l'élément suivant la forme autonome qui réalise le lien syntaxique. Aussi considérons-nous ces éléments suivant la forme autonome comme des connecteurs syntaxiques.

Afin de bien décrire tous les types de propositions japonaises, l'analyse correcte de ces connecteurs est indispensable. Nous allons donc maintenant essayer de réorganiser cette zone désordonnée afin de bien définir chacune de ces catégories.

7.6.2 Connecteurs syntaxiques des propositions

Comme nous venons de le voir, nous considérons comme des connecteurs syntaxiques les éléments suivant la forme autonome, qui assurent la connexion syntaxique des subordonnées terminées par un mot prédicatif à une forme autonome conclusive (subordonnées conclusives ci-après).

Mais les subordonnées terminées par un mot prédicatif à une forme non conclusive peuvent être elles-aussi suivies d'un autre élément – notamment les

particules de mise en relief. Ces éléments suivant une forme connective du mot prédicatif doivent-ils être considérés eux-aussi comme des connecteurs ?

Éléments suivant le mot prédicatif à une forme connective

Dans la mesure où, dans le cas des subordonnées terminées par un mot prédicatif à une forme connective, le mot prédicatif est à une forme lui permettant d'assurer lui-même la connexion, l'élément suivant le mot prédicatif est sans doute un élément non chargé de la réalisation du lien, c'est-à-dire un élément non obligatoire sur le plan syntaxique. Nous les considérons donc non pas comme des connecteurs syntaxiques mais comme des marqueurs du lien sémantique.

Différents connecteurs suivant le mot prédicatif à une forme autonome

Rappelons encore une fois les cinq possibilités de constitution des subordonnées terminées par un mot prédicatif à une forme autonome :

1. constitution d'une subordonnée déterminante ;
2. constitution d'une subordonnée déterminant un substantif formel (subordonnée assimilée substantive) ;
3. constitution d'une subordonnée conjonctive avec une particule conjonctive ;
4. constitution d'une subordonnée avec une particule adverbiale ;
5. constitution d'une subordonnée de citation avec la particule *to*.

Les subordonnées résultant de la première possibilité sont les seules subordonnées conclusives sans connecteur. Dans les autres types, la connexion est réalisée par le connecteur constitué respectivement par le substantif formel, la particule conjonctive, la particule adverbiale et la particule *to*.

Le problème que nous rencontrons ici est, encore une fois, que la définition des catégories concernant les mots constituant les trois premiers connecteurs étant assez floue, nous ne savons finalement pas déterminer exactement ces connecteurs.

Afin d'élucider ces notions obscures, nous allons maintenant passer en revue les travaux antérieurs sur la catégorisation de ces mots.

7.7 État de l'art des travaux sur la catégorisation des mots suivant une forme autonome

7.7.1 Les mots agglutinants de Sakuma

Sakuma (1940b) a regroupé les mots suivant le mot prédicatif à une forme autonome sous le nom de *kyûchakugo*, mots agglutinants, après avoir découvert leur point commun : être peu voire non autonome et capable de constituer un élément autonome et complet sémantiquement une fois déterminé.

Mais, Sakuma n'a pas défini de catégorie spécifique à part : *kyûchakugo* restait la simple appellation d'un ensemble de mots de catégories différentes regroupés selon une certaine caractéristique commune qu'ils possédaient.

Mikami définit en revanche une catégorie spécifique couvrant l'ensemble de ces mots qu'il appelle 準詞 (*junshi*, mot assimilé). Les *junshi* se distinguent eux-même en deux types : mots assimilés mots prédicatifs (準用詞, *jun'yô-shi*) et mots assimilés substantifs (準体詞, *juntai-shi*)

7.7.2 Les études comparatives de Teramura

Sans essayer de réorganiser les catégories existantes ni d'en définir une nouvelle, Teramura montre tout simplement la différence et la continuité de ces catégories. D'après l'auteur, cette continuité entre les substantifs constituant la base d'une subordonnée déterminante d'un côté et les particules conjonctives de l'autre, est formée par les substantifs constituant la base d'une subordonnée déterminante, plus ou moins lexicalisés, ayant ainsi un caractère formel de niveau différent.

Il propose plusieurs types de tests pour évaluer le caractère substantif de chaque mot. Il en existe deux types : tests pour la possibilité qu'un mot accepte la détermination d'un mot donné (déterminant démonstratif, qualificatif, juxtaposition d'un substantif, etc.) et tests pour la possibilité qu'un mot assume une fonction donnée (cas *ga*, cas *wo*, cas *no*, cas *kara*, cas *ni*, cas zéro).

Le tableau 7.4 (voir page suivante), repris de Teramura (1978), présente le résultat de ces tests réalisés par Teramura.

7.7.3 La réorganisation complète proposée par Okutsu et Numata

Okutsu propose dans l'introduction de Okutsu et al. (1986) une réorganisation complète des catégories existantes, notamment les différents types de particules. Il renonce à la catégorie de particule et les particules sont catégorisées dans différentes classes nouvellement définies, telles que les substantifs formels, les ad-
verbes formels ou les *toritate-shi* (mots de mise en relief).

Substantifs formels et adverbies formels d'Okutsu

La catégorie des substantifs formels d'Okutsu est beaucoup plus limitée que celle définie antérieurement, alors que celle des adverbies formels (Okutsu, 1986) est relativement large incluant une grande partie des particules adverbiales et conjonctives. En effet, Okutsu détermine la catégorie équivalente selon les fonctions que peuvent jouer les syntagmes constitués par ces mots. Ainsi, beaucoup de substantifs formels et de particules adverbiales sont classés dans la catégorie d'adverbies formels par Okutsu.

表 1

	承							接						
	① コレガ ダ	② [名] ノ	③ コノ、 ソノ	④ [形] イ	⑤ コンナ、 ソナ	⑥ [名]	⑦ コレ、 ソレ	⑧ ガ	⑨ ヲ	⑩ ノ	⑪ カラ	⑫ ニ	⑬ φ	
ト	×	○	○	○	○	×	×	○	○	○	○	○	○	
ア	×	×	×	×	×	×	×	○	○	○	○	○	○	
コ	×	×	×	×	×	○	○	△	×	○	×	×	○	
以	×	×	×	×	×	○	○	○	○	○	×	○	×	
以	×	△	△	×	×	○	○	○	○	○	×	×	○	
カ	×	×	×	×	×	×	×	×	×	×	×	×	○	
マ	×	○	○	○	△	×	×	×	△	○	×	○	○	
タ	×	○	○	○	○	×	×	×	×	×	×	×	○	
度	×	○	○	○	○	×	×	×	×	×	×	×	○	
場	×	○	○	○	○	×	×	×	×	×	×	×	○	
目	×	○	○	○	○	×	×	×	×	×	×	×	○	
タ	×	○	○	○	○	×	×	×	×	×	×	×	○	
セ	×	○	○	○	○	×	×	×	×	×	×	×	○	
理	×	○	○	○	○	×	×	×	△	×	×	×	○	
カ	×	○	○	○	○	×	×	×	×	×	×	×	○	
ユ	○	○	○	○	○	×	×	○	○	○	×	×	○	
ワ	×	○	○	○	○	×	×	×	×	○	×	×	○	
原	×	○	○	○	○	×	×	×	×	○	×	×	○	
結	×	○	○	○	○	×	×	×	×	○	△	×	○	
末	×	○	○	×	×	×	×	×	×	○	×	○	○	
ゲ	×	○	○	○	○	×	×	×	×	○	×	○	○	
ア	×	○	○	○	○	×	×	×	×	○	×	○	○	
ウ	×	○	○	○	○	×	×	×	×	○	×	○	○	
マ	×	○	○	○	○	×	×	×	×	○	×	○	○	
コ	×	○	○	○	○	×	×	○	○	○	×	○	○	
程	×	×	×	○	×	△	○	○	○	○	×	×	○	
ホ	×	○	×	○	×	×	×	○	○	○	×	×	○	
限	×	×	×	△	×	×	×	×	×	○	×	×	○	
ダ	×	○	○	○	○	×	×	×	×	○	×	×	○	
カ	×	×	×	×	×	×	×	×	×	×	×	×	○	
キ	×	×	×	×	×	×	×	×	×	×	×	×	○	
ナ	×	×	×	×	×	×	×	×	×	×	×	×	○	
マ	×	○	○	○	○	×	×	○	○	○	×	×	○	
ク	×	○	○	○	○	×	×	×	×	○	×	○	○	
ト	×	○	○	○	○	×	×	×	×	○	×	○	○	
ワ	×	○	○	○	○	×	×	×	×	○	×	○	○	
様	×	○	○	○	○	×	×	×	×	○	×	○	○	
ヨ	×	○	○	○	○	×	×	×	×	○	×	○	○	

TAB. 7.4 – Tableau comparatif réalisé par Teramura (1978)

Mots de mise en relief de Numata

Les particules adverbiales et les *kakari-joshi* qui ne sont pas considérées comme des adverbes et des substantifs formels sont catégorisées pour la plupart dans la classe des とりたて詞 (*toritate-shi*), mots de mise en relief. Numata (1986) définit les mots de mise en relief comme des unités mettant en relief différents éléments de la phrase (appelés les internes) tout en représentant les relations logiques qu'entretiennent ces éléments avec d'autres éléments du même paradigme (appelés les externes). Ainsi, certaines particules adverbiales et d'autres particules *kakari-joshi* sont classées dans cette nouvelle catégorie.

Les mots regroupés ainsi sont caractérisés par quatre particularités syntaxiques :

1. grandes possibilités de positionnement ;
2. caractère facultatif ;
3. possibilité d'apparition dans les subordonnées déterminantes ;
4. caractère non substantif.

et ces quatre caractères sont considérés comme les conditions nécessaires et suffisantes pour la détermination des mots de mise en relief.

7.7.4 Analyse critique

Malgré tout l'intérêt que présentent ces travaux, la tentative de réorganisation des catégories d'Okutsu et la définition et les études des mots de mise en relief, *toritate-shi*, de Numata ne conviennent pas à nos travaux d'analyse automatique des phrases.

Problèmes de la définition des adverbes formels d'Okutsu

Comme nous l'avons vu, Okutsu détermine la catégorie équivalente selon les fonctions que peuvent jouer les syntagmes constitués par le mot en question. Mais, cette définition annulant la frontière entre les analyses en termes de catégories et en termes de fonctions, n'est pas la solution la plus économique.

Par exemple, Okutsu considère ため (*tame*) comme un adverbe formel, du fait notamment que ce mot, déterminé par une subordonnée, constitue une proposition de cause et de but. Mais la séquence terminée par *tame* peut former, en étant suivi de la particule *no*, le constituant déterminant. Si on définit le mot *tame* comme un adverbe formel, on doit également définir un autre substantif formel *tame* ou un qualificatif formel *tame no*.

Toutefois, il nous semble plus économique et même cohérent de considérer ce mot comme un substantif formel ayant un emploi adverbial, emploi possédé par beaucoup de substantifs japonais.

Comme l'affirme Le Goffic (1993a), nous considérons que les « étiquetage en termes de fonctions et étiquetage en termes de catégories doivent être parallèles et distincts » et nous ne pouvons pas approuver la position d'Okutsu.

Inconvénients de la définition des mots de mise en relief de Numata

La catégorie des mots de mise en relief définie par critères sémantiques comme décrit précédemment, regroupe des unités ayant des comportements syntaxiques si divers que Numata considère comme une des caractéristiques leur liberté de positionnement. Mais ce n'est que le résultat de l'utilisation de critères non syntaxiques mais sémantiques. Dans le cadre de nos travaux, nous préférons accorder plus d'importance aux aspects syntaxiques et distinguer au moins les unités capables de suivre une forme autonome et d'assurer la connexion syntaxique, des autres qui ne sont pas en mesure de remplir ces deux conditions.

Numata utilise tout de même également un critère syntaxique : le caractère facultatif. Mais ce critère peut remettre en cause la justesse de cette catégorisation. Par exemple, dans la phrase :

可能性が ある かどうか すら も 分からない。

(*kanôsei ga - aru - ka dôka - sura - mo - wakaranai*)

(possibilité [*ga*] - exister - [interrogation totale indirecte] - même - [*mo*] - savoir [né-gation])

« (Je) ne sais même pas s'il y a des possibilités. »

les mots *sura* et *mo*, classés dans les mots de mise en relief, sont tous les deux effectivement facultatifs syntaxiquement.

En revanche, dans la phrase (reprise de Numata (1986)) :

太郎は 働く だけ を 生き甲斐 としている。

(*tarô wa - hataraku- dake - wo - ikigai - to shiteiru*)

(Taro [*wa*] - travailler - [*dake*] - [*wo*] - raison d'être - considérer comme)

« Taro considère comme sa seule raison d'être le fait de travailler. »

la suppression de *dake* rend la phrase agrammaticale. Pour respecter le critère de caractère facultatif, il est possible de considérer que le mot *dake* suivant une forme autonome est un substantif formel et distinct du *dake* catégorisé dans les mots de mise en relief. Mais dans ce cas, on ne peut pas s'empêcher de se demander l'intérêt de distinguer deux unités ayant une forme et un sens identiques, pour n'en regrouper qu'une avec d'autres mots similaires.

Une autre caractéristique syntaxique des mots de mise en relief est leur caractère non substantif. Les mots de mise en relief comportent en fait des particules adverbiales considérées dans beaucoup de grammaires comme des mots assimilés substantifs. Mais Numata ne reconnaît pas le caractère substantif de ces particules. Afin de justifier sa position, elle utilise un test consistant à transformer la phrase en subordonnée déterminante de manière à faire du syntagme contenant la particule en question la base déterminée⁷. Par exemple, la phrase d'exemple précédente est transformée en :

*太郎が 生き甲斐 としている 働く だけ

(*tarô ga - ikigai - to shiteiru - hataraku- dake*)

(Taro [*ga*] - raison d'être - considérer comme - travailler - [*dake*])

⁷Ce test est proposé par Okutsu (1974).

et comme la séquence résultant de cette transformation est agrammaticale, l'auteur considère que le mot *dake* n'a pas de caractère substantif.

Mais ce test nous laisse sceptiques quant à sa justesse. En effet, avec ce test, nous ne pouvons pas non plus reconnaître le caractère substantif du substantif formel *no*, dont personne ne nie la capacité nominalisatrice.

Le syntagme nominalisé par *no* (souligné dans l'exemple) :

太郎は 働く の を 生き甲斐 としている。
 (tarô wa - hataraku- no - wo - ikigai - to shiteiru)
 (Taro [wa] - travailler - [no] - [wo] - raison d'être - considérer comme)
 « Taro considère comme sa raison d'être le fait de travailler. »

ne peut pas constituer la base d'une subordonnée déterminante, tout comme le syntagme terminé par *dake* :

*太郎が 生き甲斐 としている 働く の
 (tarô ga - ikigai - to shiteiru - hataraku- no)
 (Taro [ga] - raison d'être - considérer comme - travailler - [no])

Le plus grand défaut de cette méthode d'évaluation du caractère substantif est sans doute qu'elle ne tient pas compte de l'existence des deux aspects différents du « caractère substantif ». Ce caractère doit, en effet, être évalué, comme Teramura l'a fait avec ses différents tests, sur les deux capacités distinctes : capacité à régir d'autres éléments (c'est-à-dire le nombre de manières de qualifier le mot en question) et capacité à être régi (c'est-à-dire le nombre de fonctions que le mot en question peut assumer).

Nous essayons maintenant de réaliser une catégorisation détaillée de l'ensemble des mots agglutinants en tenant bien compte de ces deux aspects différents du caractère substantif.

7.8 Notre catégorisation et ses critères

7.8.1 Méthodologie

Nous nous basons principalement sur le résultat de différents tests réalisés par Teramura et présenté dans le tableau 7.4 page 286. Pour les mots qui ne figurent pas dans le tableau de Teramura ou dont l'analyse nous semblait contestable, nous avons réalisé nos propres tests avec le corpus Shincho (représentant 66 899 phrases, cf. Liste des corpus utilisés p. 550 et suivantes) et de plusieurs articles du corpus Yomiuri (3 237 phrases) à l'aide d'un concordancier élémentaire⁸ que nous avons développé spécifiquement (cf. fig. 7.5 (voir page suivante)).

⁸Il indique les occurrences d'une séquence saisie par l'utilisateur, avec indication de ses contextes gauche et droit de 20 caractères.

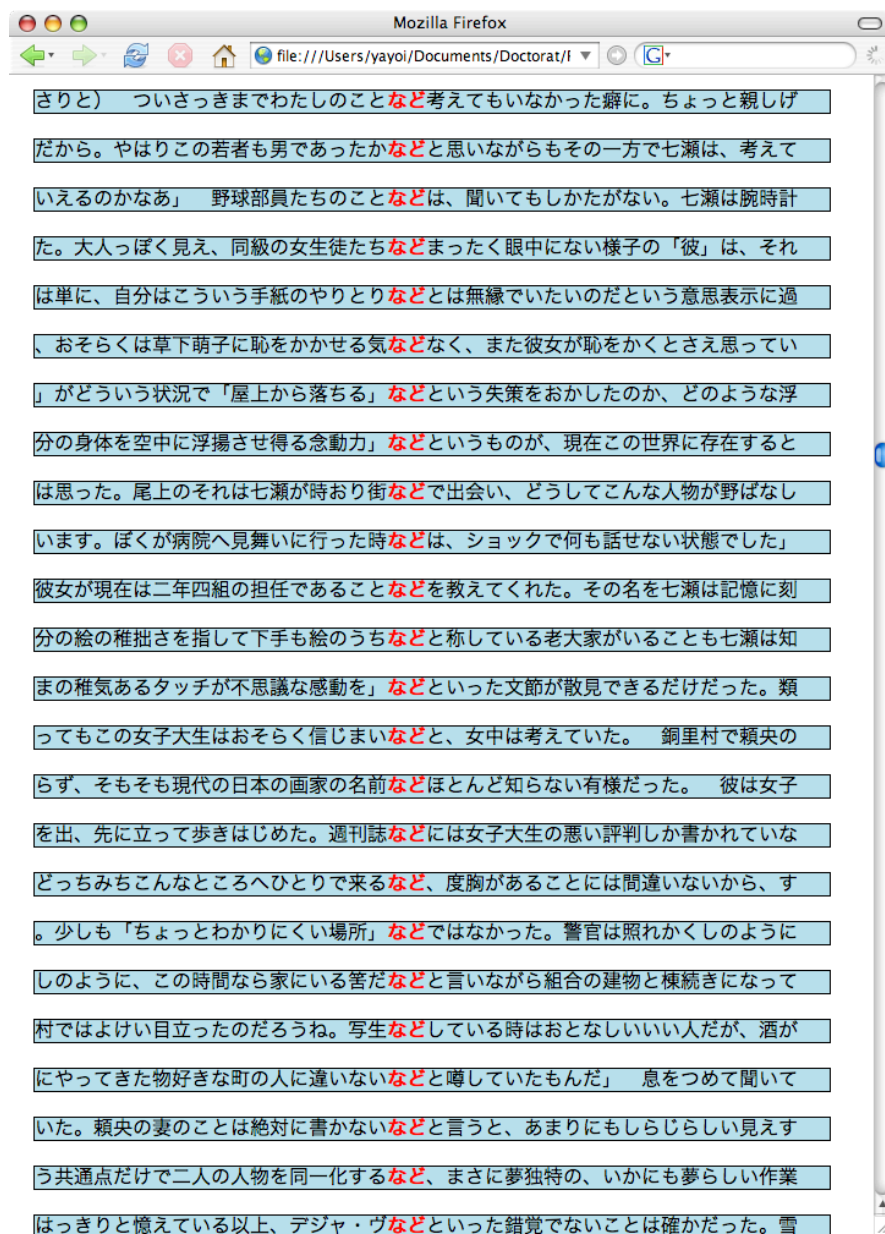


FIG. 7.5 – Résultat d'analyse par notre concordancier

7.8.2 Définition des connecteurs agglutinants

Parmi les mots qui peuvent suivre une forme autonome, nous considérons comme des connecteurs agglutinants, ceux qui ne sont ni substantifs autonomes, ni auxiliaires, ni particules conjonctives.

Nous éliminons d'abord avec des critères concrets les deux premières unités n'appartenant pas aux connecteurs agglutinants. Puis, nous examinons le reste des mots pour les catégoriser finalement selon leur caractère substantif en quatre classes – l'une regroupe les particules conjonctives, et les trois autres correspondent à des sous-catégories des connecteurs agglutinants.

La catégorisation se déroule comme suit :

- étape de catégorisation 0 : exclusion des éléments ne suivant pas une forme autonome ;
- étape de catégorisation 1 : exclusion des éléments autonomes et définition des mots agglutinants ;
- étape de catégorisation 2 : exclusion des auxiliaires ;
- étape de catégorisation 3 : classement en quatre catégories.
- étape de catégorisation 4 : définition des connecteurs agglutinants.

Catégorisation 0 : exclusion des éléments ne suivant pas une forme autonome

Nous distinguons tout d'abord les mots qui peuvent suivre une forme autonome de ceux qui ne le peuvent pas. Suite à ce test, les particules *koso*, *sae*, *shika*, *sura*, *wa*, *mo*, *zo*, définies comme *kakari-joshi* dans le dictionnaire ipadic⁹, sont catégorisées quasiment toutes comme des éléments ne suivant pas une forme autonome.

Nous avons tout de même constaté quelques cas d'exception.

sae apparaît deux fois après la forme autonome comme « *omoidasu sae* » (se souvenir - sae) dans les œuvres de Mishima et de Tanizaki. Mais, ces exemples très peu fréquents (2 occurrences contre 33 pour la structure comportant un nominalisateur) ne peuvent pas être un motif de reconnaissance de l'emploi *sae* suivant directement la forme autonome. Nous n'avons donc pas retenu la particule *sae* comme un élément suivant une forme autonome.

Nous avons également constaté des emplois de *shika* (ne ... que) suivant une forme autonome. Mais ces occurrences précèdent presque toujours directement le prédicat principal *nai* (ne pas exister), ou avec au plus le syntagme nominatif, *hōhō ga/wa* ou *te ga/wa* (moyen), et/ou le syntagme locatif, *hoka ni* (ailleurs), telles que :

のぼりつめていく しか 手は ない。

(*nobori tsumete iku - shika - te wa - nai*)

(monter continuellement - [*shika*] - moyen [*wa*] - ne pas exister)

« (Je) n'ai aucun moyen (autre que) de monter continuellement. »

⁹Dictionnaire électronique pour le TAL utilisé par l'analyseur morphologique ChaSen que nous utilisons dans notre réalisation informatique.

Cette construction avec *shika* étant assez figée, nous ne traitons pas dans la présente étude cette particule comme un élément assurant le lien syntaxique, position contestable nécessitant peut-être des réexamens futurs.

Catégorisation 1 : exclusion des éléments autonomes et définition des mots agglutinants

Nous utilisons un test de Teramura : si le mot *N* peut constituer la phrase « *kore ga N desu* » (c'est *N*), le mot *N* est considéré comme un substantif autonome.

Les autres mots (c'est-à-dire les mots capables de suivre une forme autonome mais qui ne sont pas autonomes) sont des *kyûchakugo*, mots agglutinants.

Catégorisation 2 : exclusion des auxiliaires

Les mots agglutinants qui peuvent constituer le prédicat principal éventuellement avec la copule et qui ne peuvent constituer un élément de phrase (y compris une proposition subordonnée) qu'à l'aide d'une particule conjonctive ou par la variation de leur forme, sont considérés comme des auxiliaires.

Catégorisation 3 : classement en quatre catégories

La classification est réalisée suite à l'évaluation du caractère substantif de chaque mot sur les deux aspects dont nous avons parlé précédemment.

Les mots sont examinés du point de vue, d'une part de leur capacité à régir d'autres éléments (c'est-à-dire le nombre de manières de qualifier le mot en question), et d'autre part, de leur capacité à être régi (c'est-à-dire le nombre de fonctions que le mot en question peut assumer).

Pour le premier aspect, nous recourons essentiellement au test vérifiant si le mot *N* en question est capable de constituer le syntagme nominal « (substantif) *no N* ». Par exemple, dans la mesure où il est tout à fait possible de dire « *kodomo no toki* » (enfant - [*no*] - quand ; temps), on considère que le mot *toki* a un caractère substantif élevé du point de vue de la capacité à être déterminé. En revanche, le mot *dake* pouvant être déterminé non pas avec la particule *no* « **kodomo no dake* » (enfant - [*no*] - seulement) mais seulement par juxtaposition « *kodomo dake* » (enfant - seulement), on considère qu'il a un caractère substantif faible du point de vue de la capacité à être déterminé.

Pour le second aspect, nous évaluons si le mot *N* en question est capable ou non de constituer – une fois déterminé – les constituants de la phrase du cas *ga*, cas *wo*, cas *no*, cas *kara*, cas *ni* et cas zéro. Par exemple, le mot *dake* pouvant constituer tous les constituants sauf celui du cas *kara* (d'après l'analyse de Teramura), on considère qu'il a un caractère substantif élevé du point de vue de la capacité à assumer des fonctions. En revanche, le mot *kiri* (depuis) ne pouvant constituer que le syntagme nu (cas zéro), on considère qu'il a un caractère substantif très faible du point de vue de la capacité à assumer des fonctions.

En combinant ces deux analyses, nous pouvons établir quatre classes comme présenté dans le tableau 7.6.

capacité à être déterminé \ capacité à assumer des fonctions		caractère substantif	
		fort	faible
caractère substantif	fort	substantif formel	adverbe / qualificatif substantifs
	faible	particule nominalisatrice	particule conjonctive

TAB. 7.6 – Catégorisation des éléments suivant une forme autonome du mot variable

Nous appelons substantifs formels les mots ayant un caractère substantif élevé pour les deux aspects tels que *toki* (quand ; temps) ou *mama* (tel quel).

Sont appelés particules nominalisatrices les mots assurant plusieurs types de fonctions comme les substantifs, mais pour lesquels les types de déterminants qui peuvent les déterminer sont limités, tels que *dake* (seulement), *ka* ([marqueur d'interrogation]) ou *no* ([nominalisateur]).

Les mots acceptant plusieurs types de déterminants mais qui sont figés quant au choix des fonctions qu'ils assurent, se distinguent en deux types : qualificatifs substantifs tels que *yô* (manière) et adverbessubstantifs comme *wari (ni)* (relativement) ou *kuse (ni)* (malgré)¹⁰.

Enfin, les mots n'ayant qu'un faible caractère substantif sur les deux aspects sont des purs connecteurs de propositions : des particules conjonctives.

Catégorisation 4 : définition des connecteurs agglutinants

Nous considérons comme connecteurs agglutinants les substantifs formels, les particules nominalisatrices et les adverbessubstantifs (sur fond coloré dans le tableau 7.6), lorsqu'ils réalisent la nominalisation de la proposition qui les précède. La connexion syntaxique à proprement parler est généralement assurée par la particule de cas qui les suit (ou par son absence, ou encore par la terminaison) mais nous considérons la construction constituée de ces mots et éventuellement de la particule comme des connecteurs reliant la proposition qu'ils introduisent au prédicat postérieur.

Les particules nominalisatrices apparaissant également après différentes unités autres que les formes autonomes du mot variable (sauf la particule *ka*) sont en fait des particules usuellement classées dans les particules de mise en relief.

¹⁰ Ces adverbessubstantifs sont si figés qu'ils ne peuvent pas non plus constituer le prédicat principal de la phrase contrairement aux autres mots agglutinants qui peuvent, eux, former le prédicat de la phrase en étant suivi de la copule.

Les particules nominalisatrices peuvent être considérées comme constituant une sous-catégorie des particules de mise en relief.

7.8.3 Résultat général de notre catégorisation

Nous avons proposé ici une manière de tracer des frontières entre les unités, frontières assez floues jusqu'aujourd'hui, avec des critères tout à fait concrets.

Sur un plan pratique, la catégorisation a été réalisée sur la base du résultat des analyses réalisées par Teramura, que nous avons complétées par l'analyse de quelques mots supplémentaires.

Cette première expérimentation nous a fourni un résultat qui confirme dans les grandes lignes la justesse de la catégorisation traditionnelle.

Cette analyse devrait cependant être réalisée, dans des travaux futurs, de manière plus systématique et pour un plus grand nombre de mots par l'analyse d'un corpus de taille importante.

7.8.4 Caractéristiques et problèmes des *kyûchakugo*

Fonction des mots agglutinants assimilés à des auxiliaires

Outre des connecteurs semblables aux particules conjonctives, ces mots agglutinants peuvent également constituer, comme le montre Teramura (1978), des prédicats, devenant ainsi des éléments semblables aux auxiliaires, *jodôshi*.

Dans la grammaire de Masuoka & Takubo (1992), il existe une sous-catégorie des auxiliaires dite « contenant un substantif formel ».

Nous considérons les mots agglutinants suivis directement de la copule non comme des connecteurs, mais comme constituant avec la copule un auxiliaire.

Problèmes liés aux mots agglutinants : polysémie

Teramura (1978) signale que ces mots n'ont pas toujours le même caractère formel et qu'ils expriment un sens plus ou moins stable et autonome selon le contexte. La polysémie pose toujours des problèmes difficiles lors de l'analyse automatique. Dans le cadre de la présente thèse, nous ne prétendons pas proposer de solution à ce problème de polysémie des mots agglutinants, mais nous observerons l'influence de cette polysémie afin d'obtenir des données bénéfiques à nos futurs travaux.

7.9 État de l'art sur les typologies des subordonnées

Nous allons maintenant passer en revue les typologies des subordonnées proposées par différents linguistes.

7.9.1 Typologie selon la forme de connexion

Teramura (1982a) distingue, comme nous l'avons abordé dans la section 7.6.1, les subordonnées constituées du mot variable à une forme connective de celles se terminant par une forme conclusive et il définit d'abord quatre types de subordonnées terminées par un mot prédicatif à une forme conclusive :

1. subordonnée déterminant le substantif qui la suit, appelée **subordonnée déterminante** ;
2. subordonnée déterminant un substantif formel (tel que *no*, *koto* ou *ka*) et constituant ensemble une unité équivalente à un substantif, qu'il appelle **subordonnée assimilée à un substantif** (et qu'il considère comme un type de subordonnée déterminante) ;
3. subordonnée terminée par une particule conjonctive, appelée **subordonnée conjonctive** ;
4. subordonnée terminée par la particule *to*, appelée **subordonnée de citation**.

En tenant compte de ces quatre types de subordonnées terminées par une forme conclusive, ainsi que des subordonnées terminées par une forme connective, il définit finalement cinq types de phrases complexes selon la subordonnée qu'elles contiennent :

- Type 1 :
phrase contenant une subordonnée coordonnée terminée par un verbe à la forme neutre, ayant ses propres compléments différents de ceux du prédicat principal ;
- Type 2 :
 1. phrase contenant une subordonnée terminée par un mot prédicatif à une forme de condition ;
 2. phrase contenant une subordonnée terminée par un mot prédicatif à une forme autonome suivie de la particule *to* produisant l'expression de condition ;
- Type 3 :
phrase contenant une subordonnée déterminante (y compris subordonnée assimilée à un substantif) ;
- Type 4 :
phrase contenant une subordonnée conjonctive ;
- Type 5 :
phrase contenant une subordonnée de citation.

Plus le chiffre est grand, plus le caractère phrasogénérateur de la proposition subordonnée est élevé, c'est-à-dire plus elle est proche de la phrase autonome.

7.9.2 Typologies selon les fonctions des subordonnées dans la phrase

Nous constatons des typologies selon les fonctions des subordonnées dans la phrase dans les travaux de Noda et dans ceux de Masuoka et Takubo. Les travaux de Noda (2002) étant très brefs et ne faisant pas ressortir de point de vue particulier, nous ne présentons ici que les travaux de ces derniers, la grammaire de Masuoka & Takubo (1992) et les deux ouvrages plus récents de Masuoka (1997, 2002).

Masuoka & Takubo (1992) catégorisent l'ensemble des propositions comme suit :

1. Principale;
2. Connectée :
 - a) Proposition subordonnée ;
 - i. Proposition complétive ;
 - ii. Proposition adverbiale ;
 - iii. Proposition déterminante ;
 - b) Proposition coordonnée.

Les propositions connectées (cf. § 7.5) sont catégorisées en deux classes : subordonnée et coordonnée. La première a elle-même trois sous-classes : propositions adverbiales, propositions complétives et propositions déterminantes (i.e. relatives).

Cette catégorisation est modifiée dans les travaux postérieurs de Masuoka (1997, 2002).

Dans ces nouveaux travaux, les propositions sont à nouveau catégorisées comme suit :

1. Principale;
2. Subordonnée :
 - a) substantive ;
 - b) adverbiale ;
 - c) déterminante ;
 - d) coordonnée ;
 - e) flottante.

La première grande différence est la disparition de l'opposition subordonnée-coordonnée. Dans la nouvelle catégorisation, les coordonnées sont un type de subordonnée.

La deuxième différence est que les propositions appelées complétives sont regroupées selon la nouvelle catégorisation sous le nom de substantives. Les propositions dites interrogatives appartiennent toujours à cette classe, mais les propositions de citation sont classées maintenant dans les subordonnées adverbiales.

Le troisième changement est la définition des subordonnées flottantes. Ce sont des propositions fonctionnant comme des éléments externes à la proposition tels que les éléments indépendants de la grammaire scolaire.

Propositions subordonnées complétives

鍵 を 忘れた こと に 気がついた。

(*kagi - wo - wasureta - koto - ni - kigatsuita*)

(clés - [wo] - oublier [passé] - le fait de [substantif formel] - [ni] - se rendre compte [passé])

« (Je) me suis aperçu que (j')avais oublié mes clés. »

Ce sont des propositions fonctionnant comme des compléments du prédicat. Il en existe trois types.

Le premier est la proposition terminée par un substantif dit formel tel que *koto* dans l'exemple. Cette structure est identique à celle d'une subordonnée déterminant un substantif normal et l'ensemble formé par la proposition déterminante et par le substantif formel déterminé est donc équivalent à un SN. Si bien qu'il est généralement suivi d'une particule de cas qui marque sa fonction syntaxique comme le fait la particule de cas *ni* dans l'exemple.

Le deuxième type est la proposition dite interrogative telle que :

何を していた か (が) 知りたい。

(*nani wo - shiteita - ka - (ga) - shiritai*)

(quoi [wo] - faire [état, passé] - [interrogation] - [ga] - vouloir savoir)

« (Je) veux savoir ce qu'(il) faisait. »

Les auteurs notent que les particules *ga* – marquée entre parenthèses dans l'exemple – et *wo* suivant la subordonnée sont souvent omises.

Le troisième type est la proposition introduite par l'expression de citation telle que :

誰も いない と 思った。

(*dare mo - inai - to - omotta*)

(personne - se trouver [négaration] - [citation] - penser [passé])

« (J')ai pensé qu'il n'y avait personne. »

Dans les travaux plus récents de Masuoka, ce dernier type est classé dans les subordonnées adverbiales, du fait notamment de l'existence de l'emploi tel que :

こんなことを しては いけない と ワープロの ス
イッチを 入れた。

(*kon'na koto wo - shiteite wa - ikenai - to - wâpuro no - suicchi wo - ireta*)

(telle chose [wo] - faire [état] - il ne faut pas - [citation] - machine de traitement de textes [no] - bouton [wo] - appuyer pour allumer [passé])

« (J'ai) allumé le traitement de textes (en me disant) que je n'avais pas que ça à faire. »

Propositions subordonnées adverbiales

時間 が あれば 出席します。

(*jikan - ga - areba - shusseki shimasu*)

(temps - [ga] - se trouver [condition] - assister [non passé])

« Si (j')ai le temps, (j')y assisterai. »

Il existe six formes :

1. proposition terminée par le prédicat à une forme non conclusive telle qu'une forme neutre ou de condition ;
2. proposition suivie d'une particule de mise en relief ;
3. proposition déterminant un substantif formel suivie d'une particule de cas ;
4. proposition suivie d'une particule conjonctive ;
5. proposition suivie d'un suffixe ;
6. proposition suivie d'une locution conjonctive.

Les auteurs proposent également des sous-catégories selon le sens (temporel, causal, etc.).

Propositions subordonnées déterminantes

きのう 借りた 本 を 読んだ。

(*kinô - karita - hon - wo - yonda*)

(hier - emprunter [passé] - livre - [wo] - lire [passé])

« (J)'ai lu le livre que (j')avais emprunté hier. »

Les subordonnées déterminantes sont distinguées en trois classes selon la relation syntaxique existant entre elles et leur base.

Lorsque la base a une fonction syntaxique de complément vis-à-vis du prédicat de la subordonnée déterminante, cette dernière est appelée 補足語修飾節 (*hosokugo shûshoku setsu*, proposition qualifiant le complément).

Les auteurs appellent substantifs relatifs les substantifs tels que 前日 (*zenjitsu*, la veille) ou あと (*ato*, après) dont le sens a une valeur relative. Lorsque la base est un substantif relatif, la subordonnée qui la détermine est appelée 相対名詞修飾節 (*sôtai meishi shûshoku setsu*, proposition qualifiante avec un substantif relatif). La subordonnée suivante en est un exemple :

フランスへ 発つ 前日

(*furansu e - tatsu - zenjitsu*)

(France [e] - partir - la veille)

« la veille (du jour) (où il) partira/est parti pour la France ».

Le complément supprimé de la subordonnée est « le jour », adverbe de temps pour le prédicat « partir ». La base déterminée par la subordonnée, « la veille », est un substantif relatif ayant une valeur relative par rapport à ce complément supprimé de la subordonnée, « le jour ».

Le troisième type est la proposition appelée 内容節 (*naiyô setsu*, proposition de contenu). Dans cette proposition, la base n'a pas de fonction syntaxique. La proposition exprime le contenu de l'élément auquel se réfère la base.

実験が 失敗した 報告

(*jikken ga - shippai shita - hōkoku*)

(expérience [ga] - échouer [passé] - rapport)

« rapport (informant) (que) l'expérience a échoué ».

La subordonnée exprime le contenu « l'expérience a échoué » du « rapport ». Les auteurs expliquent que ce dernier type de proposition est souvent réalisé avec des expressions de citation telles que « *to iu* ». Dans les travaux postérieurs de Masuoka, l'ensemble des déterminantes est d'abord distingué en deux types selon la nature de connexion : sans connecteur et avec connecteur « *to iu* » ou « *yōna* ».

Masuoka semble avoir retravaillé ce classement, et dans ses ouvrages plus récents, l'auteur définit une nouvelle catégorisation comme suit :

1. déterminantes en relation interne (内の関係, *uchi no kankei*) ;
2. propositions de contenu ;
3. déterminantes réduites (縮約連体節, *shukuyaku rentai setsu*).

Le premier type concerne les subordonnées qu'il a appelées dans ses travaux antérieurs *hosokugo shūshoku setsu*, proposition qualifiant le complément. La relation interne¹¹ est celle constatée entre cette déterminante et sa base : cette dernière a une fonction syntaxique de complément vis-à-vis du prédicat de cette première.

Le troisième type, le nouveau, désigne des subordonnées déterminantes dont la relation avec leur base est implicite. Par exemple, dans la structure :

復習しなかった 報い

(*fukushū shinakatta - mukai*)

(réviser [négation, passé] - punition)

« punition qui résulte de mon absence de révision »

la relation entre « punition » et « ne pas avoir révisé » qui est « A comme résultat de B » est implicite et l'interprétation par « punition qui résulte de mon absence de révision » nécessite des données extra-linguistiques.

Propositions coordonnées

父 が フランス人 で、母 が 日本人 です。

(*chichi - ga - furansu jin - de - haha - ga - nihon jin - desu*)

(père - [ga] - français - [copule : neutre] - mère - [ga] - japonais - [copule : conclusive, polie])

« Mon père est français, ma mère japonaise. »

Ce sont des propositions qui sont sur un pied d'égalité avec la principale.

¹¹ La notion de relation interne, opposée à la relation externe (外の関係, *soto no kankei*), provient des travaux sur les subordonnées déterminantes de Teramura, très captivants. L'analyse précise des subordonnées déterminantes ne concernant pas directement nos travaux actuels, nous ne décrivons pas ici les détails de ses travaux, quoiqu'intéressants et bénéfiques. Ses principaux articles sur la subordonnée déterminante sont repris dans Teramura (1993).

Il existe deux types de formes : propositions terminées par le prédicat à une forme neutre et propositions avec un connecteur, en particulier une particule conjonctive.

Les auteurs signalent que la première forme sert également à constituer la subordonnée adverbiale.

Enrichissement avec les travaux réalisés dans le cadre d'une application au TAL

Dans le cadre de la conception du système CBAP (*Clause Boundaries Annotation Program*) (Maruyama et al., 2004) – système de détection automatique des frontières de proposition –, les auteurs complètent la liste des propositions de Masuoka et Takubo afin d'obtenir une liste exhaustive permettant de détecter les propositions de manière automatique¹².

Du fait de leur caractère appliqué et surtout automatique, ces travaux présentent le grand avantage d'être complets. Les auteurs ont d'abord défini dix sous-classes de propositions connectées en sous-catégorisant les quatre classes de subordonnée définies dans Masuoka & Takubo (1992) :

- Proposition complétive :
 1. proposition complétive (2) ;
 2. proposition de citation (5) ;
 3. proposition interrogative indirecte (3) ;
- Proposition adverbiale :
 4. proposition de condition et de concession (23) ;
 5. proposition de cause (8) ;
 6. proposition de temps (21) ;
 7. proposition de manière (12) ;
 8. autres (38) ;
- Proposition déterminante :
 9. proposition déterminante (15) ;
- Proposition coordonnée :
 10. proposition coordonnée (12).

Les valeurs entre parenthèses indiquent le nombre de patrons définis pour chaque classe.

Ainsi, sont finalement définis 139 types de propositions à reconnaître automatiquement.

Outre ces propositions, les compléments extra-prédicatifs sont aussi extraits à part et le nombre de types de frontières de proposition s'élève à 147.

¹²La description plus détaillée du système CBAP sera présentée dans § 10.2.1 et nous aborderons également le résultat de notre propre évaluation de ce système dans § 11.1.

7.9.3 Autres typologies

Les typologies des subordonnées selon les couches auxquelles elles appartiennent – du type les travaux de Minami (cf. 7.3.3) – sont abondantes, et nous ne citerons que Masuoka (1997) et Noda (2002). Nous ne réalisons pas un état de l'art plus détaillé de ces travaux dont nous ne pouvons profiter dans nos travaux. Masuoka (1997) présente également une typologie selon le niveau de dépendance de la subordonnée vis-à-vis de la principale.

7.9.4 Récapitulation et analyse critique

Le tableau 7.7 (voir page suivante) est une synthèse comparative des études présentées.

Tableau comparatif

La colonne « Teramura » indique la classe à laquelle appartient la subordonnée de l'exemple selon la catégorisation de Teramura (1982a), la colonne « Masuoka & Takubo (92) » celle selon la grammaire de Masuoka & Takubo (1992), et la colonne « Masuoka (1997, 2002) » celle selon les travaux de Masuoka (1997, 2002).

Le point d'interrogation signale que la subordonnée équivalente à l'exemple n'est pas traitée dans les travaux concernés et que la classe indiquée résulte de notre interprétation réalisée d'après la définition des auteurs.

Analyse des typologies

La catégorisation de Masuoka et Takubo, que nous avons qualifiée de catégorisation selon la fonction, est en réalité beaucoup moins cohérente que ne le laisse entendre cette qualification.

En effet, la définition même de coordination est plutôt sémantique que syntaxique. Masuoka (1997) dit lui-même que les subordonnées coordonnées peuvent être interprétées comme adverbiales, mais qu'étant donné que du point de vue sémantique, les propositions coordonnées sont sur un pied d'égalité avec leurs principales, il est incorrect de les traiter comme des adverbiales.

De plus, le statut des subordonnées substantives est également discutable car ce regroupement est réalisé avec des critères non pas du niveau fonctionnel mais du niveau lexical, à savoir les catégories lexicales équivalentes. Dans la mesure où un mot d'une catégorie donnée peut assumer différentes fonctions, la classe substantive peut coïncider avec les autres classes définies par des critères fonctionnels. C'est justement la raison pour laquelle les propositions de citation posent autant de problèmes dans leur catégorisation. Les auteurs n'abordent pas ce point dans leurs ouvrages, mais certaines subordonnées classées dans les adverbiales sont également des substantives (ou complétives selon la terminologie de leurs anciens travaux).

7. ÉTUDE DE LA PHRASE COMPLEXE

	Teramura	Masuoka & Takubo (92)	Masuoka (1997, 2002)
花子が 詩を 書き、太郎が 作曲した。 <i>hanako ga - shi wo - kaki - tarô ga - sakkyokushita</i> Hanako [gô] - parole [wo] - écrire - Tarô [gô] - composer « Hanako a écrit les paroles et Tarô a composé »	T1	Crd	Crd
雨が 降って、試合が 中止に なった。 <i>ame ga - futte - shiai ga - chûshi ni - natta</i> pluie [gô] - tomber - match [gô] - cessation [ni] - devenir « Il a plu et le match a été annulé (le match a été annulé à cause de la pluie) »	T1	Crd ou Adv	Crd ou Adv
時間が あれば、出席します。 <i>jikan ga - areba - shussekishimasu</i> temps [gô] - exister [condition] - assister [politesse] « Si j'ai le temps, j'y assisterai »	T2	Adv	Adv
鍵を 忘れた こと に 気がついた。 <i>kagi wo - wasureta - koto - ni - kigatsuita</i> clés [wo] - oublier [passé] - le fait [substantif formel] - [ni] - se rendre compte [passé] « (Je) me suis aperçu que (j')avais oublié mes clés »	T3	Cmp	Sub
きのう 借りた 本を 読んだ。 <i>kinô - karita - hon wo - yonda</i> hier - emprunter [passé] - livre [wo] - lire [passé] « (J')ai lu le livre que (j')avais emprunté hier »	T3	Dét	Dét
太郎は 英語が 上手だ し、花子は スペイン語が 話せる。 <i>tarô wa - eigo ga - jôzuda - shi - hanako wa - supeingo ga - hanaseru</i> Tarô [wo] - anglais [gô] - fort - [addition] - Hanako [wo] - espagnol [gô] - savoir parler « Tarô est fort en anglais, et en plus Hanako sait parler espagnol »	T4	Crd	Crd
何度も 説明した のに 理解してもらえなかった。 <i>nando mo - setsumeishita - noni - rikaishite morae nakatta</i> plusieurs fois - expliquer [passé] - [concession] - pouvoir faire comprendre [négation, passé] « (J'ai eu) beau expliquer à plusieurs reprises, je n'ai pas pu (lui) faire comprendre »	T4	Adv	Adv
何を していた か が 知りたい。 <i>nani wo - shiteita - ka - ga - shiritai</i> quoi [wo] - faire [état, passé] - [interrogation] - [gô] - vouloir savoir « (Je) veux savoir ce qu'il faisait »	T5 ?	Cmp	Cmp
誰も いない と 思った。 <i>dare mo - inai - to - omotta</i> personne - se trouver [négation] - [citation] - penser [passé] « (J')ai pensé qu'il n'y avait personne »	T5	Cmp	Adv ?
こんなことを しては いけない と ワープロの スイッチを 入れた。 <i>kon'na koto wo - shiteite wa - ikenai - to - wôpuro no - suicchi wo - ireta</i> telle chose [wo] - faire [état] - il ne faut pas - [citation] - machine de traitement de textes [no] - bouton [wo] - appuyer pour allumer [passé] « (J'ai) allumé le traitement de textes (en me disant) que je n'avais pas que ça à faire »	T5	Cmp ou Adv ?	Adv ?
話が まとまった かどうか 早く 結果が 知りたい。 <i>hanashi ga - matomatta - ka dôka - hayaku - kekka ga - shiritai</i> propos [gô] - conclure [passé] - [interrogation totale] - vite - résultat [gô] - vouloir savoir « (Je) veux vite connaître le résultat (pour savoir) s'ils se sont mis d'accord »	T5 ?	Cmp ou Adv ?	Flt

Crd = coordonnée, **Cmp** = complétive, **Sub** = substantive, **Adv** = adverbiale, **Dét** = déterminante, **Flt** = flottante

TAB. 7.7 – Comparaison des typologies des subordonnées

En revanche, la typologie de Teramura basée uniquement sur des critères formels est tout à fait cohérente. Nous adopterons donc pour nos travaux la typologie proposée par Teramura.

7.10 Notre typologie des subordonnées

Nous définissons donc les différentes subordonnées comme suit :

- Subordonnées sans connecteur :
 1. **subordonnées neutres** : propositions ou sous-phrases terminées par un mot variable à une forme neutre, éventuellement suivi de particules de mise en relief ;
 2. **subordonnées de condition** : propositions terminées par un mot variable à une forme de condition, éventuellement suivi de particules de mise en relief ;
 3. **subordonnées déterminantes sans connecteur** : propositions terminées par un mot variable à une forme conclusive.
- Subordonnées avec connecteur :
 1. **subordonnées avec particule conjonctive** : propositions ou sous-phrases terminées par un mot variable à une forme conclusive suivi d'une particule connective ;
 2. **subordonnées avec connecteur agglutinant** : propositions constituées d'une déterminante et d'un connecteur agglutinant, éventuellement suivi de particules de cas et/ou de mise en relief. Ces propositions se distinguent elles-mêmes en deux types :
 - a) à fonction de complément essentiel : suivies d'une particule de cas ;
 - b) à fonction de complément accessoire : utilisées seules sans particule de cas ;
 3. **subordonnées de citation** : propositions ou sous-phrases terminées par un mot variable à une forme conclusive terminées par la particule *to* ;
 4. **subordonnées déterminantes avec connecteur** : propositions ou sous-phrases terminées par un mot variable à une forme conclusive suivi d'un connecteur déterminant.

Distinction des connecteurs durs et souples

Notre définition ne tient pas encore compte de la différence entre les connecteurs durs et souples.

Leur distinction exacte – considérée comme non indispensable pour notre sujet actuel – ne sera pas réalisée dans le cadre de la présente thèse, mais l'état de l'art présenté dans ce chapitre pourrait constituer la base de futurs travaux.

Cas de la double apparition des connecteurs

Il existe des cas où plusieurs connecteurs se succèdent les uns après les autres, comme par exemple la phrase interrogative terminée par la particule *ka* (connecteur agglutinant) introduite dans son ensemble par la particule *to* de citation. Dans ce cas, nous considérons comme connecteur celui apparaissant en dernier.

7.11 Récapitulation : définition formelle de la phrase

D'après tout ce que nous avons vu jusqu'ici, nous pouvons écrire formellement la structure de la phrase japonaise comme suit¹³ :

phrase → sous-phrase? phrase
 phrase → élément-externe* thème? proposition
 proposition → subordonnée? proposition
 proposition → racine
 racine → complément* prédicat-forme-conclusive
 sous-phrase → élément-externe* thème? subordonnée-neutre particule-mise-en-relief?
 sous-phrase → phrase particule-conjonctive
 subordonnée → (subordonnée-condition | subordonnée-connecteur-agglutinant | subordonnée-neutre) particule-mise-en-relief?
 subordonnée → proposition particule-conjonctive
 subordonnée-déterminante → (subordonnée-déterminante particule-mise-en-relief?) | (phrase connecteur-déterminant)
 subordonnée-citation → phrase particule-*to*.
 complément → subordonnée-citation.
 complément → subordonnée-connecteur-agglutinant particule-de-cas.
 ...
 SN → déterminant N.
 ...
 déterminant → subordonnée-déterminante.
 ...
 élément-externe → subordonnée.
 ...

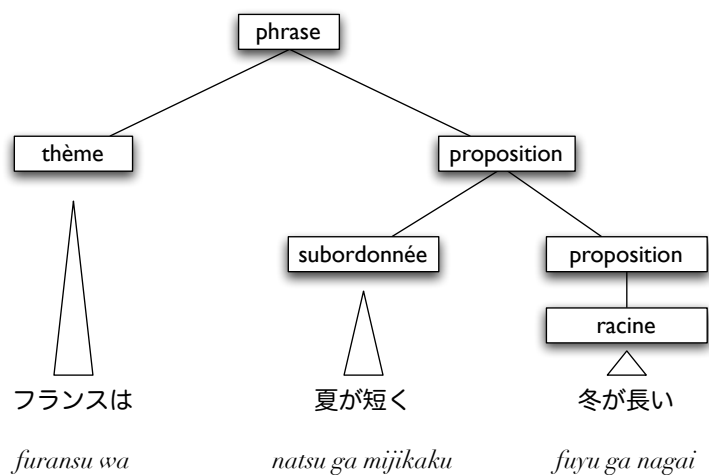
Considérons maintenant deux exemples d'analyse de phrases selon cette définition.

Exemple d'analyse 1 : phrase contenant une subordonnée

フランス	は	{夏 が 短く、}{冬 が 長い。}
(furansu - wa -		natsu - ga - mijikaku - fuyu - ga - nagai)
(la France - [wa] - été - [ga] - être court - hiver - [ga] - être long)		

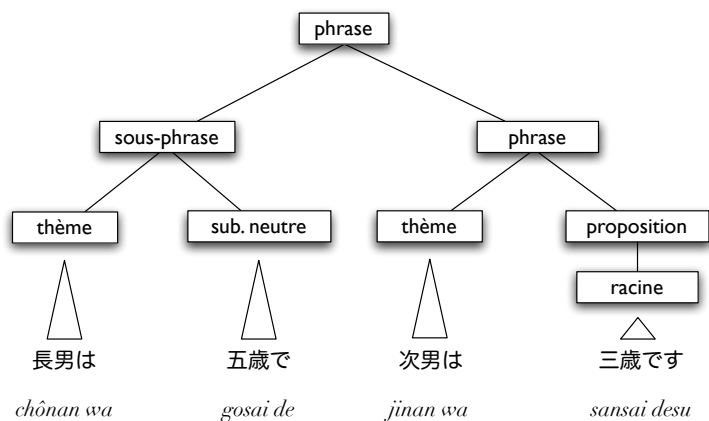
¹³Du fait du sujet de nos travaux, seules les règles concernant les structures propositionnelles sont décrites. Cette définition formelle a pour but d'obtenir cohérence et précision dans notre définition des subordonnées et nous ne cherchons nullement à construire une grammaire pour l'analyse syntaxique complète. Il est évident que beaucoup d'autres règles seraient à définir pour une analyse syntaxique complète.

« En France, l'été est court et l'hiver est long. »



Exemple d'analyse 2 : phrase comportant une sous-phrase

長男 は 五歳 で、 次男 は 三歳 です。
 (chônân - wa - go sai - de - jinan - wa - san sai - desu)
 (fils aîné - [wa] - 5 ans - [copule] - deuxième fils - [wa] - [copule : poli])
 « Notre fils aîné a cinq ans et notre deuxième a trois ans. »



7.12 Relations entre le syntagme thématisé et les subordonnées

Comme nous l'avons vu dans la section 6.4.3, le syntagme thématisé a généralement une double fonction. Il entre en relation, en tant que thème de la phrase,

avec l'ensemble du prédicat principal régissant tous ses compléments. Le syntagme thématisé joue, en plus, vis-à-vis du radical du mot prédicatif, un rôle de complément.

Nous allons maintenant étudier les relations que le syntagme thématisé entretient avec le prédicat de chaque subordonnée constituant la phrase.

7.12.1 Mécanisme général

中国は 国力が 伸びて {自分たちの 国際的地位を 高める}戦略を とっている。

(*chûgoku - wa - kokuryoku ga - nobite - jibuntachi no - kokusaiteki chii wo - takameru - senryaku wo - totteiru*)

(Chine [*wa*] - pouvoir national [*ga*] - grandir - leurs - statut sur le plan international [*wo*] - élever - stratégie [*wo*] - adopter [progressif])

« Son pouvoir national ayant grandi, la Chine adopte une stratégie permettant d'élever son statut sur le plan international »

La proposition s'opposant au thème de cette phrase est constituée de la subordonnée neutre et de la proposition racine (encadrées).

Le syntagme thématisé « *chûgoku wa* (la Chine [*wa*]) » assure, en plus de celui de thème de la phrase, le rôle de complément vis-à-vis des mots prédicatifs de ces sous-propositions. Il assure le rôle de complément secondaire du nominatif « *kokuryoku* (pouvoir national) » du premier prédicat « *nobite* (grandir) » et le cas de *ga* vis-à-vis du prédicat de la racine « *totteiru* (adopter) ».

La proposition racine contient elle-même une subordonnée déterminant son complément en *wo*. Dans la phrase d'exemple, le syntagme thématisé n'assure aucune fonction cumulative vis-à-vis du prédicat de la déterminante contenue dans la racine, mais il est tout à fait possible qu'il joue un rôle de complément vis-à-vis du prédicat situé au plan secondaire.

Détermination de la fonction cumulative du thème

Afin de déterminer la fonction cumulative du syntagme thématisé, on cherche la place du cas vide dans chaque proposition constituant la phrase, comme représenté figure 7.8 page suivante.

Les informations sur les cas nécessaires à chaque prédicat sont prédéfinies. Lors de l'analyse, on prévoit les emplacements des cas nécessaires à chaque prédicat. Les compléments explicites sont ensuite associés à ces emplacements.

Dans l'exemple étudié (cf. figure 7.8 page ci-contre), le prédicat principal « *totteiru* » a entraîné la génération des emplacements pour le cas *ga* et pour le cas *wo* (encadrés sur la figure). L'emplacement ouvert par le prédicat de la subordonnée « *nobite* » se limite au cas *ga*, mais le syntagme nominal a également ouvert un emplacement pour son complément secondaire.

Pour automatiser cette opération, différentes informations seraient nécessaires et devraient être stockées dans des bases de données du type diction-

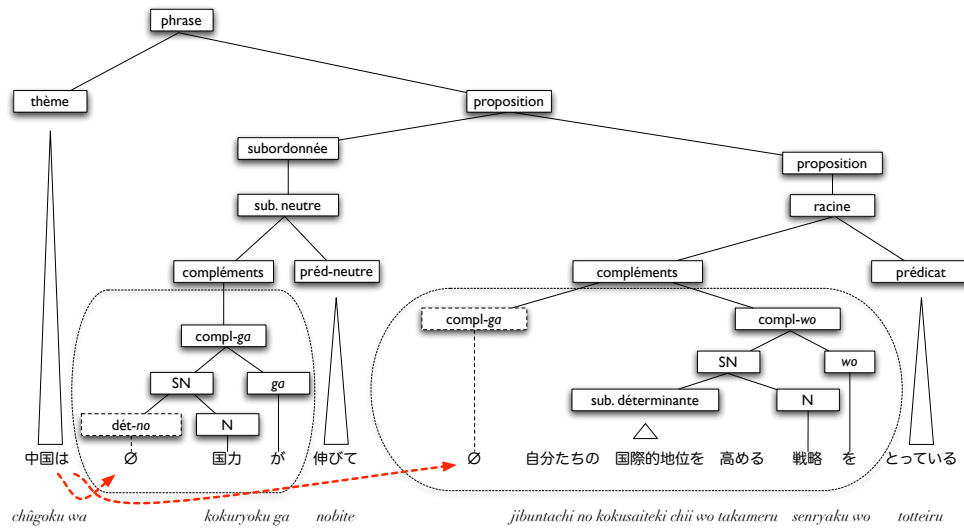


FIG. 7.8 – Détermination de la fonction cumulative du thème

naire : types de compléments nécessaires pour chaque mot variable, caractères de chaque substantif constituant chaque complément, etc.

De plus, beaucoup de travaux restent à réaliser afin de déterminer précisément les différentes règles : conditions d'ouverture des emplacements de complément secondaire, accessibilité du syntagme thématisé, etc.

7.12.2 Problème lié à la portée du thème dans la structure introduite par la particule *to*

Il existe des phrases où le syntagme thématisé est considéré comme appartenant à la subordonnée. Comme nous pouvons le voir dans notre définition formelle de la phrase, certains connecteurs, notamment celui de citation, peuvent introduire une sous-phrased ou même une phrase contenant un thème.

Masuoka & Takubo (1992) analysent le thème des deux phrases suivantes par celui appartenant à la subordonnée de citation introduite par la particule *to* (ou le connecteur déterminant comportant cette particule)¹⁴ :

- P1 計画 は 中止する べきだ という 意見 が 多かった。
 (*keikaku - wa - chūshi suru - beki da - to iu* - *iken - ga - ôkatta*)
 (projet - [wa] - annuler - il faut que - [connecteur déterminant] - opinion - [ga] - être nombreux [passé])
 « Les opinions considérant qu'il fallait annuler le projet étaient nombreuses. »

¹⁴L'encadrement n'est pas effectué dans l'ouvrage, mais ajouté ici pour mieux comprendre la structure.

P2 鈴木さん は まだ 学校 に いる と 思う。
 (*suzuki san - wa - mada - gakkô - ni - iru* - to - omou)
 (M. Suzuki - [wa] - encore - école - [locatif] - se trouver - [citation] - je pense)
 « Je pense que M. Suzuki est encore à l'école. »

Ambiguïté d'interprétation

La figure 7.9 représente cette interprétation – par laquelle le thème appartient à la sous-phrasedu premier exemple sous forme d'un arbre syntaxique selon notre définition.

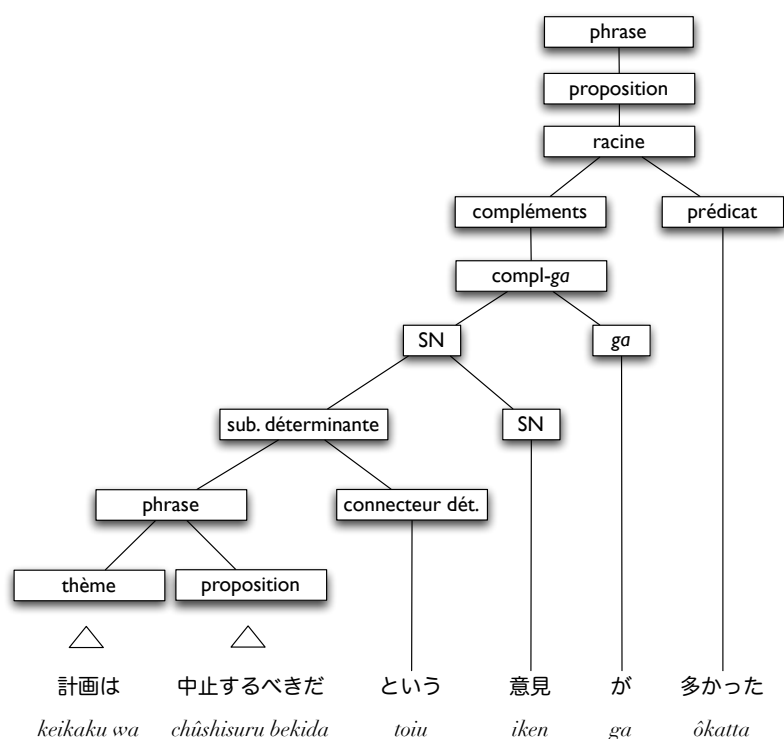


FIG. 7.9 – Interprétation de P1 - 1

Mais l'appartenance du thème à la subordonnée peut-elle vraiment être déterminée selon le type de subordonnée? En effet, il est tout à fait possible d'imaginer une situation où l'on parle d'un projet et que quelqu'un explique qu'il y avait beaucoup d'opinions défavorables vis-à-vis de ce projet lors d'un référendum en disant : « Ce projet, il y avait plus de gens qui pensaient qu'il fallait l'annuler. » Dans ce cas, le thème « *keikaku - wa* » est bien celui de la phrase et non pas seulement de la sous-phrasede citation. La figure 7.10 page ci-contre représente cette seconde interprétation.

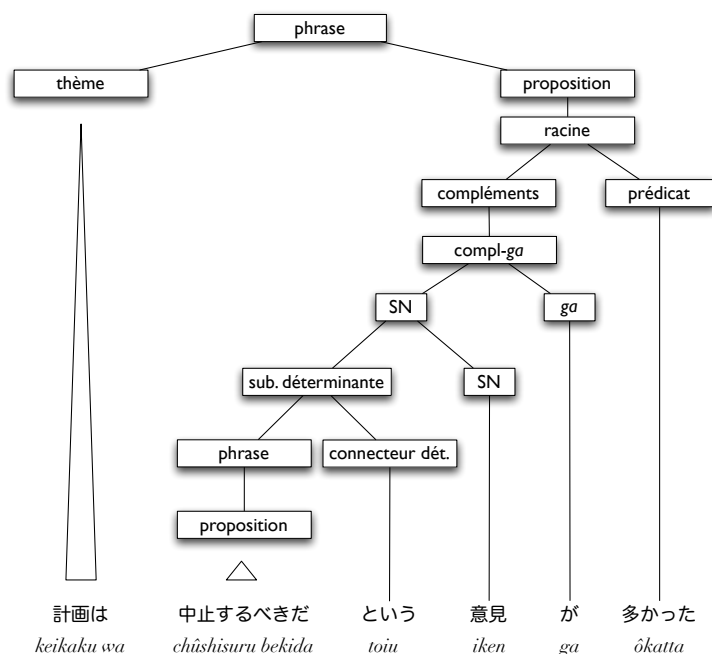


FIG. 7.10 – Interprétation de P1 - 2

Analyse du syntagme thématisé

L'appartenance du thème à une subordonnée ne peut pas être déterminée par le type de cette subordonnée, qui fournit seulement une information sur la possibilité ou non d'appartenance du thème. En d'autres termes, si le type de connecteur (soit le connecteur proprement dit, soit la forme du mot variable réalisant la connexion) peut introduire une sous-structure contenant le thème, on installe le nœud « sous-phrase » (ou « phrase »), et dans le cas contraire, on ne peut installer que le nœud « subordonnée » duquel ne peut pas dériver la branche « thème ».

Toutefois, la détermination de l'appartenance effective du thème – dans le cas où la subordonnée peut constituer la sous-phrase contenant le thème – dépendant totalement du contexte ou de la situation, il est extrêmement difficile voire impossible de résoudre cette ambiguïté de manière automatique.

7.12.3 Notre position pour la réalisation

La détermination de la fonction cumulative du thème semble indispensable à l'analyse syntaxique de la phrase japonaise, en particulier dans un cadre multilingue avec traitement de plusieurs langues ayant des structures complètement différentes.

Néanmoins, une implémentation complète demanderait des recherches ex-

trêmement poussées. Ce point constitue un sujet à part entière, dépassant le cadre de la présente thèse. Aussi, nous contentons-nous d'examiner les résultats d'alignement sans traitement de résolution de fonction cumulative, pour analyser sa réelle influence et évaluer correctement son utilité.

7.13 Problèmes liés au phénomène d'ellipse

Tout comme pour la phrase française, le phénomène d'ellipse est une source de problèmes lors de la définition de la proposition du japonais.

7.13.1 Omission du prédicat

Étudions un exemple cité dans Mikami (1959) :

三月 に 梅 が、四月 に 桜 が 咲きます。
(*sangatsu - ni - ume - ga - shigatsu - ni - sakura - ga - sakimasu*)
(mars - [*ni*] - prunier - [*ga*] - avril - [*ni*] - cerisier - [*ga*] - fleurir [*poli*])
« Les pruniers fleurissent en mars, et les cerisiers en avril. »

Si on détermine les propositions en fonction de la présence d'un prédicat, cette phrase est constituée d'une seule proposition à un prédicat auquel sont attachés deux ensembles coordonnés de compléments en *ga* et en *ni*. Mais, comme le dit Mikami, il est tout à fait possible de considérer qu'elle est constituée de deux propositions et que dans la première le prédicat est omis.

Il existe également une structure similaire avec coordination d'autres compléments. Comparons les deux exemples suivants :

私 が 朝日を、夫 が 読売 を 読んでいます。
(*watashi - ga - asahi - wo - otto - ga - yomiuri - wo - yonde imasu*)
(moi - [*ga*] - Asahi - [*wo*] - mon mari - [*ga*] - Yomiuri - [*wo*] - lire [*habitude*])
« Je lis l'Asahi et mon mari, le Yomiuri. »

ワイン を 二杯、ビール を 三本 飲んだ。
(*wain - wo - ni hai - biru - wo - san bon - nonda*)
(vin - [*wo*] - deux verres - bière - [*wo*] - trois bouteilles - boire [*passé*])
« J'ai bu du vin, deux verres, et de la bière, trois bouteilles. »

Le premier exemple dans lequel les sujets sont coordonnés pourrait être considéré comme constitué de deux propositions dans sa traduction française, alors que le second exemple ne peut être défini que comme une phrase d'une seule proposition.

Toutefois, dans le cas du japonais, il est impossible – du moins difficilement justifiable – de reconnaître ce type de différence entre ces deux structures. Les auteurs du système CBAP considèrent les structures de coordination partielle, telles que les exemples ci-dessus, comme des phrases complexes aussi bien pour

la coordination des compléments en *ga* que pour celle d'autres compléments. Néanmoins, les concepteurs de ce système reconnaissent également l'extrême difficulté de déterminer automatiquement les frontières de propositions pour ces structures.

7.13.2 Omission de la partie variable du prédicat

Il existe également une structure dans laquelle la partie variable du premier prédicat n'apparaît pas dans la représentation du niveau de surface. Cette construction est extrêmement courante dans l'ensemble du style écrit mais plus particulièrement dans les textes journalistiques.

二階西側 の 部屋 から 出火、二階建て住宅 を 全焼 した。

(*nikai nishigawa - no - heya - kara - shukka - nikai date jûtaku - wo - zenshō - shita*)

(le coté ouest du premier étage - de - appartement - [*kara* : point de départ] - apparition du feu - immeuble de deux étages - [*wo*] - destruction totale par l'incendie - [verbe support : passé])

« Le feu a pris dans l'appartement du premier étage du coté ouest, et l'immeuble de deux étages a été entièrement détruit (par cet incendie). »

高速道路 で トラック が 衝突、三人 が 死亡 した。

(*kōsokudōro - de - torakku - ga - shōtotsu - san'nin - ga - shibō - shita*)

(autoroute - [*de*] - camion - [*ga*] - collision - trois personnes - [*ga*] - décès - [verbe support : passé])

« (Deux) camions se sont heurtés sur l'autoroute et trois personnes ont été tuées. »

Les prédicats de cette structure sont, la plupart du temps, des substantifs ayant un sens représentant une action¹⁵. Ces substantifs constituent un verbe en s'attachant au verbe support, *suru* (する, faire)¹⁶.

Dans cette structure, l'ensemble des éléments de la proposition, y compris le mot prédicatif, est coordonné sauf la partie variable du prédicat, *suru*. Contrairement aux exemples de la section précédente que nous hésitions à définir comme des phrases complexes du fait de l'absence de prédicat sur la représentation du niveau de surface, ces phrases conservant le radical du prédicat dans la première proposition peuvent être facilement considérées comme des phrases complexes.

Mais la détection automatique de cette structure est très difficile. Les systèmes tels que CBAP, basés uniquement sur les informations locales d'une zone assez limitée, ne peuvent pas reconnaître les frontières de ces propositions.

¹⁵ Il s'agit souvent de mots constitués en *kanji* (*kango*) que Matsushita considérait comme des verbes et appelait verbes invariables. Mais Minami (1993) fait remarquer qu'il existe également des mots purement japonais (*wago*) de ce type et qu'aujourd'hui on rencontre beaucoup d'exemples avec des mots emprunts constitués en *katakana*.

¹⁶ Ces substantifs peuvent également constituer un autre type de prédicat en s'associant avec la copule. Minami appelle la phrase ainsi construite phrase pseudo-substantive (voir la section 6.3.2) et présente la particularité de sa structure, semblable à celle de la phrase verbale.

7.13.3 Notre position pour la réalisation

Étant donné que nous utilisons la présence d'un mot variable comme condition minimum pour la détection des propositions, la reconnaissance des propositions dans lesquelles le mot variable constituant le prédicat est omis est impossible.

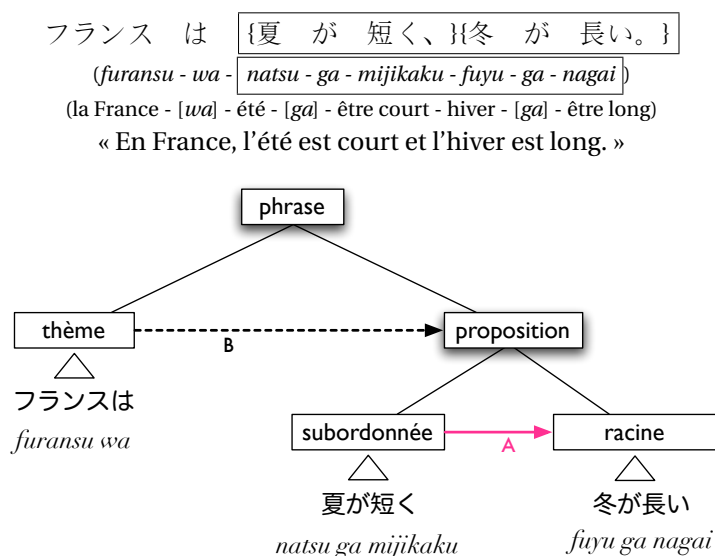
Nous reportons la réalisation de la détection des subordonnées terminées par le radical du mot prédicatif, en réalité indispensable, à de prochains travaux.

7.14 De l'arbre des constituants à la représentation en graphe des relations de dépendance des propositions

Jusqu'ici, nous avons utilisé les arbres des constituants pour représenter l'analyse structurale de la phrase. Mais comme nous exploiterons les relations de dépendance entre les propositions lors de l'opération postérieure d'alignement, nous allons présenter dans cette section comment une structure peut être représentée sous forme d'un graphe des relations de dépendance.

7.14.1 Arbre des constituants et relations de dépendance

Reprenons l'arbre déjà présenté dans la section 7.11 représentant la phrase :



Dépendance entre les constituants d'un nœud « proposition »

Nous appellerons relations de dépendance les relations qu'entretiennent deux constituants dérivant d'un même nœud « proposition », telles que celle représentée par la flèche étiquetée A dans la figure. Nous dirons que le constituant apparaissant sur la feuille gauche dépend syntaxiquement du constituant apparaissant

sur la feuille droite et l'arc orienté représente cette relation en se dirigeant de l'élément dépendant vers l'élément régissant.

Relation entre les constituants d'un nœud « phrase »

En revanche, deux constituants dérivant d'un même nœud « phrase » ou « sous-phrase » tels que ceux reliés par la flèche discontinue étiquetée B dans la figure, n'entretiennent pas de relation de dépendance mais ils entrent en relation sur un pied d'égalité afin de constituer une unité supérieure. Mais, pour des raisons pratiques et afin de minimiser la complexité de la représentation, nous représentons cette relation de la même manière que la relation de dépendance entre deux constituants d'un nœud proposition, à la différence près que l'arc est représenté par une ligne discontinue.

Représentation des constituants intermédiaires

Pour la même raison, les constituants intermédiaires (i.e. les constituants représentés par un nœud intermédiaire qui n'est pas une feuille tels que β dans le graphe A de la figure 7.11, ou encore α et β dans le graphe B) ne seront pas représentés dans les graphes de dépendance.

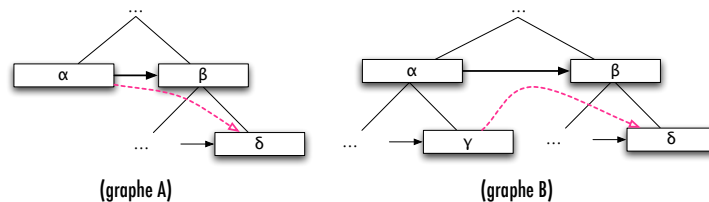


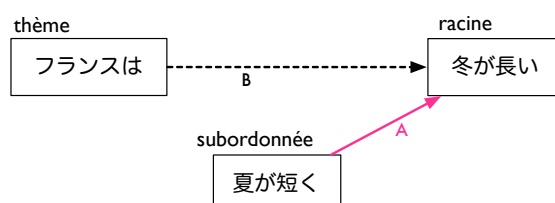
FIG. 7.11 – Relation de dépendance avec des constituants intermédiaires

Ainsi, lorsqu'un élément α entretient une relation avec un constituant intermédiaire β (cf. graphe A de la figure 7.11), nous la représenterons comme si l'élément α était en relation avec le constituant apparaissant sur la feuille tout à droite δ dérivant du nœud du constituant intermédiaire en question β . Dans le cas de la phrase d'exemple représentée dans la figure précédente, l'arc reliant le thème au nœud proposition doit alors être prolongé jusqu'à la feuille « racine ».

De même, les relations entre les deux constituants intermédiaires, α et β (cf. graphe B de la figure 7.11), seront représentées par l'arc reliant les deux constituants γ et δ , apparaissant sur la feuille tout à droite des nœuds intermédiaires α et β .

7.14.2 Graphe des relations de dépendance

En tenant compte des définitions présentées précédemment, la phrase d'exemple peut être présentée en relations de dépendance comme suit :

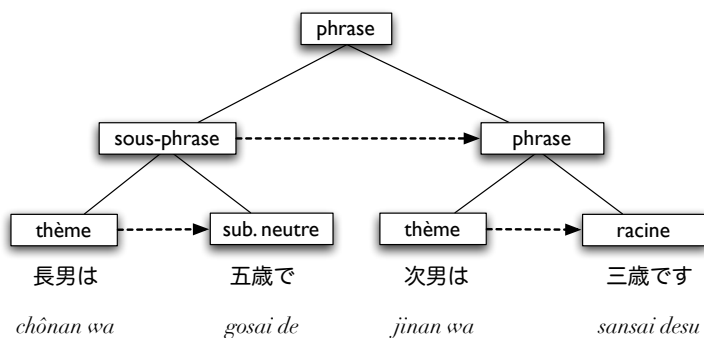
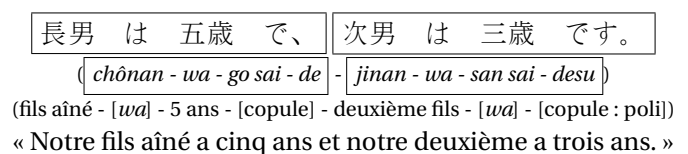


7.14.3 Exemple

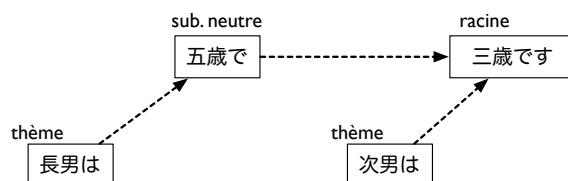
Considérons pour mieux illustrer la définition précédente encore deux autres exemples.

Exemple 1 :

Reprenons d'abord l'exemple également présenté dans la section 7.11 :



La représentation en graphe des relations de dépendance (au niveau des propositions) pour cette phrase est :

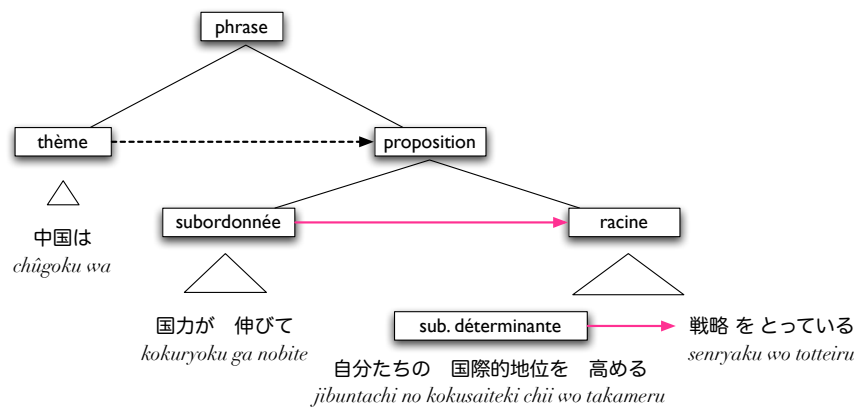


Exemple 2 :

Reprenons maintenant l'exemple utilisé dans la section 7.12.1 :

中国は 国力が 伸びて {自分たちの 国際的地位を 高める}戦略を とっている。
 (chûgoku - wa - kokuryoku ga - nobite -
 jibuntachi no - kokuzaiteki chii wo - takameru - senryaku wo - totteiru)
 (Chine [wa] - pouvoir national [ga] - grandir -
 leurs - statut sur le plan international [wo] - élever - stratégie [wo] - adopter [progressif])
 « Son pouvoir national ayant grandi, la Chine adopte une stratégie permettant d'élever son statut
 sur le plan international »

La représentation arborescente en constituants (simplifiée aux niveaux inférieurs aux propositions) est :



La représentation en graphe des relations de dépendance (au niveau des propositions) pour cette phrase est :

