

Bracketing Guidelines for Vietnamese Treebank (0.1)

Quy T. Nguyen and Yusuke Miyao

July, 2016

Table of Contents

Table of Contents.....	2
1. Introduction.....	4
2. Annotation scheme.....	4
2.1. Constituency labels.....	4
2.2. Nhãn chức năng.....	5
2.3. Null elements.....	5
3. Annotation guidance.....	6
3.1. Clause and sentence.....	6
3.1.1. Simple sentence.....	6
3.1.1.1. Descriptive sentence.....	11
3.1.1.2. Inferential sentence.....	19
3.1.1.3. Negative form.....	20
3.1.2. Question.....	21
3.1.3. Special sentence (SPL).....	27
3.1.3.1. Special sentence having main component which is a noun phrase.....	27
3.1.3.2. Special sentence whose main component includes only verb phrase or adjective phrase..	28
3.1.4. Compound sentence.....	35
3.1.4.1. Coordinating compound sentence.....	35
3.1.4.2. Reciprocal compound sentence (câu ghép qua lại).....	37
3.1.5. Mệnh đề phụ kết (SBAR).....	39
3.1.6. Note.....	42
3.2. Null elements.....	60
3.2.1. Trace of movement (*T*).....	60
3.2.2. Ellipses without trace (*E*).....	62
3.2.3. Null sentence component as it appears at other position of the sentence. (*).....	63
3.2.4. Null complementizer (*0*).....	64
3.2.5. Null passive verb (*P*).....	64
3.2.6. Null head word (*H*).....	64
3.2.7. Null post-modifier of a verb (*D*).....	66
3.2.8. Null post modifier as it is a part of a sum considered later (*SUM*).....	67
3.2.9. Note.....	68
3.3. Coindex.....	69
3.4. Phrase level.....	71
3.4.1. General norms.....	71
3.4.1.1. Independent or dependent construction.....	71
3.4.1.2. Coordination.....	78
3.4.1.2.1. Equal coordination.....	78
3.4.1.2.2. Reciprocal coordination.....	86
3.4.1.3. Sentence or Phrase.....	88
3.4.2. Noun phrase (NP).....	89
3.4.2.1. Form 1.....	89
3.4.2.1.1. Pre-modifier.....	89
3.4.2.1.2. Head word (component 0).....	91
3.4.2.1.3. Post-modifier.....	94
3.4.2.2. Form 2.....	100
3.4.2.3. Time phrase.....	101
3.4.2.3.1. Time point.....	101
3.4.2.3.2. Time period.....	103

3.4.2.3.3. <u>Lengh of time</u>	104
3.4.2.4. Measure phrase	104
3.4.2.4.1. <u>Pre-modifier</u>	104
3.4.2.4.2. <u>Head of measure phrase</u>	105
3.4.2.4.3. <u>Post-modifier</u>	105
3.4.3. Quantifier phrase (QP)	106
3.4.3.1. Pre-modifier	107
3.4.3.2. Head	108
3.4.3.3. Post-modifier	108
3.4.4. Verb phrase (VP)	109
3.4.4.1. Pre-modifier	110
3.4.4.2. Head word	112
3.4.4.3. Post-modifier	116
3.4.4.4. Caution	121
3.4.5. Adjective phrase (ADJP)	122
3.4.5.1. Pre-modifier	123
3.4.5.2. Head	123
3.4.5.3. Post-modifier	124
3.4.5.4. Caution	126
3.4.6. Prepositional phrase (PP)	126
3.4.6.1. Pre-modifier	127
3.4.6.2. Head	127
3.4.6.3. Post-modifier	131
3.4.7. Adjunct phrase (RP).....	132
3.4.8. Conjunction phrase (CONJP).....	133
3.4.9. Unlike coordinated phrase (UCP).....	134
3.4.10. Questioning phrase.....	135
3.4.11. Modal phrase (MDP).....	137
3.5. Function tags	137
3.5.1. Head of phrase (H)	137
3.5.2. Subject (SBJ).....	138
3.5.3. Logical subject (LGS)	140
3.5.4. Predicate that is not VP (PRD)	141
3.5.5. Object and Complement (CMP).....	143
3.5.5.1 Direct object (DOB).....	143
3.5.5.2. Indirect object (IOB) (giống nhãn BNF của PTB)	145
3.5.5.3. Complement (CMP).....	147
3.5.5.3. Caution	150
3.5.6. Topicalized (TPC)	153
3.5.7. Modal phrase (MDP)	159
3.5.8. Vocative (VOC).....	161
3.5.9. Parenthetical (PRN)	162
3.5.10.Subordinate components	167
3.5.10.1. Temporal (TMP).....	167
3.5.10.2. Locative (LOC)	169
3.5.10.3. Manner (MNR)	171
3.5.10.4. Purpose or reason (PRP).....	173
3.5.10.5. Adverbial (ADV).....	177
3.5.10.6. Cautions of annotating subordinate component.....	177
3.5.11. Condition (CND)	178
3.5.12. Concession (CNC)	178
3.5.13. Headline (HLN)	179

3.5.14. Title (TTL).....	179
3.5.15. Exclamative sentence (EXC).....	180
3.5.16. Imperative sentence (CMD).....	182
REFERENCES.....	184

1. Introduction

Vietnamese treebank is annotated with 3 layers, word segmentation, POS tagging and bracketing. This document is to guide for bracketing Vietnamese.

For each grammatical unit, we present the way to recognize and annotate, the ambiguous cases and how to distinguish them, examples and exceptions.

2. Annotation scheme

We use the following four types of labels: constituency labels indicating syntactic categories of the phrases, functional labels indicating syntactic functions and meanings (if any) of the phrases, null elements to mark ellipses, and reference indices to mark syntactic movement.

2.1. Constituency labels

2.1.1. Phrase tag set

No.	Tag	Meaning
1	NP	Noun phrase
2	QP	Quantitative phrase
3	VP	Verb phrase
4	ADJP	Adjective phrase
5	PP	Prepositional phrase
6	RP	Adjunct phrase
7	CONJC	Conjunction phrase
8	UCP	Unlike coordinated phrase
9	QNP	Questioning noun phrase
10	QADJP	Questioning adjective phrase
11	QRP	Questioning adjunct phrase
12	QPP	Questioning prepositional phrase
13	MDP	Modal phrase

2.1.2. Clause tag set

No.	Tag	Annotation
1	S	Simple sentence, compound sentence, or complex sentence
2	SQ	Question
3	SPL	Special sentence
4	SBAR	Relative clause

2.2. Nhận chức năng

No.	Tag	Annotation
1	H	Head of phrase
2	SBJ	Subject
3	LGS	Logical subject
4	PRD	Predicate that is not VP
5	DOB	Direct object
6	IOB	Indirect object
7	CMP	Complement
8	TPC	Topicalized
9	MDP	Modal phrase
10	VOC	Vocative
11	PRN	Parenthetical
12	TMP	Temporal
13	LOC	Locative
14	MNR	Manner
15	PRP	Purpose or reason
16	CND	Condition
17	CNC	Concessions
18	ADV	Adverbial
19	HLN	Headline
20	TTL	Title
21	EXC	Exclamative sentence
22	CMD	Imperative sentence

2.3. Null elements

No.	Tag	Nghĩa
1	*T*	Trace of movement
2	*E*	Ellipses without trace
3	*	Ellipses with trace
4	*0*	Null complementizer
5	*P*	Null passive verb
6	*H*	Ellipses with trace for head word
7	*D*	Ellipses with trace for direct object of reduced relative clause
8	*SUM*	Null post-modifier which is one of components of an aggregate

3. Annotation guidance

3.1. Clause and sentence

Sentence is a basic unit of speech expressing a complete idea. In this document, the sentences are classified on the basis of their constructions, including simple (single) sentence, special sentence, compound sentence, complex sentence, and question.

Clause is a subjective-predicative construction belonging to compound or complex sentence.

3.1.1. Simple sentence

Simple sentence includes 2 main components, subject and predicate. Where, subject indicates whom or what the sentence mentions about, predicate tells something about subject. Predicate can be a verb phrase, an adjective phrase, a prepositional phrase, etc.

Simple sentences are bracketed with S label.

Examples:

Examples:

Simple sentence includes 2 main components (without circumstance component (adverbial phrase)):

Tôi đi học.

(S (NP-SBJ Tôi)

(VP đi học)

(./PU))

Cô ấy rất đẹp.

(S (NP-SBJ Cô ấy)

(ADJP rất đẹp)

(./PU))

In addition, a single sentence can include one or more subordinate components (adverb phrases) that modify main components about temporal, location, manner, etc. The subordinate component can be stand at the beginning of the sentence, come after subject, come after object, etc.

Example:

Subordinate component expressing time stands at the beginning of the sentence:

Ngày mai, tôi đi học.

(S (NP-TMP Ngày mai)

(./PU)

(NP-SBJ tôi)

(VP đi học)

(./PU))

Subordinate component comes after subject:

If the adverbial phrase is splitted from main components by commas, the adverbial phrase is annotated the same level as the main components.

Tôi, ngày mai, đi học.

(S

(NP-SBJ Tôi)

(,/PU)

(NP-TMP ngày mai)

(,/PU)

(VP đi học)

(./PU))

If the adverb phrase is not splitted from the subject by a comma and meaning of the sentence does not change when we treat the adverb phrase as a post-modifier of the subject, the adverb phrase should be treated as a post-modifier of the subject.

(S

(NP-SBJ (Nn_swsp (Nc-H Nên) (Nn văn_minh))

(NP (Nn-H cổ_đại))

(PP-LOC (Cs-H trên)

(NP (Nn-H sao) (Nn Hoả))))

(VP (R đã) (Vv-H bị)

(VP (Vv-H quét) (R sạch))))

(PU .))

If the adverb phrase is not splitted from the subject by a comma and the sentence has a different meaning when we treat the adverb phrase as a post-modifier of the subject then

- the sentence will be annotated with structures that are appropriate for the context.
- If both annotation types (annotate the adverb phrase as the same level with subject and predicate; treat the adverb phrase as a post-modifier of the subject) are suitable with the context, treat the phrase as a post-modifier of the subject.

When subordinate component stands after object, it can be ambiguous that the subordinate component modifies head of predicate (example 1 below) or modifies object (example 2 below).

Examples:

1) *hôm qua* modifies *kể*.

Nam kể về vụ tai nạn hôm qua. {*Nam told about the accident yesterday.*}

(S

(NP-SBJ Nam)

(VP kể

(PP-DOB về vụ tai nạn)

(NP-CMP-TMP hôm qua))

(./PU))

2) *hôm qua* modifies *vụ*.

Nam kể về vụ tai nạn hôm qua. {*Nam tells about the yesterday's accident.*}

(S

(NP-SBJ Nam)

(VP kể

(PP-DOB về

(NP vụ tai nạn

(NP-TMP hôm qua))))

(./PU))

If different annotations create the sentences having different meanings, we select the annotation which is suitable for the context

However, if different annotations create the sentences having the same meanings, we treat the adverb phrase as a post-modifier of the immediately previous word (as example 2 above).

If the adverb phrase stands at the end of the sentence and splitted from the main components by commas, etc., the adverb phrase is annotated separately from the main components.

Example:

(SPL

(VP (Ve-H Có)

(NP-CMP (Nt-H lúc)

(SBAR (Cs *0*)

(S

(NP-SBJ (Nn-H giá)

(NP (Nn-H vàng)))

(VP (Vv-H giảm)

(NP-CMP

(QP (R gần) (Num-H 21))

(Nu-H USD[ounce]))

(PU ,)

(VP-PRN (Vv-H xuống)

(NP-CMP (Nn-H mức)

(ADJP (Aa-H thấp) (R nhất))

(NP-PRN (Num 1.562) (Nu-H USD[ounce]))))))))

(PU ,)

(PP-TMP (Cs-H sau)

(NP (Nq những) (Nn-H thông_tin)

(ADJP (Aa-H khả_quan))

(PP (Cs-H từ)

(NP

(Nn_swsp (Nc-H nền) (Nn kinh_tế))

(ADJP (Aa-H lớn) (R nhất)

(NP (Nn-H thể_giới))))))

(PU .))

There are 2 types of simple sentence, descriptive sentence and inferential sentence.

3.1.1.1. Descriptive sentence

This type of sentence describes status, characteristic, quality, etc. of the person/thing. There are the forms of descriptive sentences as follows:

- Subject is NP, and predicate is VP.
 - o Activity sentence: Head of predicate is transitive verb, intransitive verb, thinking verb, volition verb, or directional verb (Vv).

Examples:

Steve Jobs phát triển Apple.

(S (NP-SBJ Steve Jobs)

(VP phát_triển Apple)

(./PU))

Nhân dân rất tin tưởng Đảng cộng sản.

(S (NP-SBJ Nhân dân)

(VP rất tin tưởng Đảng cộng sản)

(./PU))

Họ đã vào Huế.

(S (NP-SBJ Họ)

(VP đã vào Huế)

(./PU))

- o Existing sentence:

Head of predicate is existing verb (Ve).

Examples:

Các đại biểu vẫn còn ở phòng họp.

(S (NP-SBJ Các đại biểu)

(VP vẫn còn ở phòng họp)

(./PU))

Nhà này có 3 người con trai và 1 người con gái.

(S (NP-SBJ Nhà này)

(VP có 3 người con trai và 1 người con gái)

(./PU))

Lòng em đã hết yêu anh rồi.

(S (NP-SBJ Lòng em)

(VP đã hết yêu anh rồi)

(./PU))

Head of predicate is not an existing verb. However, the sentence has existing meaning.

Examples:

Các đại biểu đã đến trong phòng họp.

(S (NP-SBJ Các đại biểu)

(VP đã đến trong phòng họp)

(./PU))

- Transformation sentence: Head of predicate is transformation verb.

Example:

Con chó hiền lành của tôi đã hoá dại.

(S (NP-SBJ Con chó hiền lành của tôi)

(VP đã hoá dại)

(./PU))

- Passive sentence: Head of subject is noun, and head of predicate is passive verb, *bị* or *được*.

Passive sentences are classified into 2 types:

- 1) The passive verb, *bị* or *được*, has to be used. There are 2 types of structure as follows.

Type 1: Post-modifier of the passive verb has to be a sentence.

Đất nước ta đã bị máy bay Mỹ tàn phá.

(S (NP-SBJ-1 Đất nước ta)
 (VP đã bị/Vv-H
 (S-CMP (NP-SBJ máy bay Mỹ)
 (VP tàn phá
 (NP-DOB *T*-1))))
 (./PU))

Type 2: Post-modifier of the passive verb is bracketed as VP.

Đất nước ta đã bị tàn phá bởi máy bay Mỹ.

(S (NP-SBJ-1 Đất nước ta)
 (VP đã bị/Vv-H
 (VP-CMP tàn phá *T*-1)
 (PP bởi máy bay Mỹ))
 (./PU))

Note: When *máy bay* is omitted, the sentence is annotated according to type 2. However, if the prepositional phrase (PP bởi ...) is not presented, we do not add a null element to present this PP.

Đất nước ta đã bị tàn phá.

(S (NP-SBJ-1 Đất nước ta)
 (VP đã bị/Vv-H

(VP-CMP tàn phá *T*-1))

(./PU))

Caution: Distinguishing passive sentence and simple sentence having subordinate component.

Examples:

Simple sentence having circumstantial constituent

(S (NP-SBJ Tôi) (VP bị mặ́ng) (PP-PRP (**Cs tại**) (NP anh)) (PU .))

(S (NP-SBJ Tôi) (VP bị mặ́ng) (PP-PRP (**Cs bởi**) (NP anh)) (PU .))

→ *bởi* in this case means *bởi vì*, we can replace *bởi* by *vì*

Passive sentence having logical subject ¹

(S

(NP-SBJ Tôi)

(VP bị

(VP mặ́ng)

(PP (Cs-H bởi) (NP-LGS anh)))

(PU .)

- 2) In the cases that post-modifier of the passive verb has passive meaning, passive verb can be omitted.

Examples:

Sentence with passive verb

Quả chuối này đã bị quá chín.

(S (NP-SBJ Quả chuối này)

(VP đã bị

(VP-CMP quá chín/VA))

¹ Refer to annotation of functional tags for more details.

(./PU))

Passive verb is omitted

Quả chuối này đã quá chín.

(S (NP-SBJ Quả chuối này)

(VP đã quá chín/VA)

(./PU))

3) Passive verb can be modified by a NP

Anh Nhân được xuất viện do chỉ bị sức ép nhẹ

(S (NP-SBJ-1 Anh Nhân)

(VP được

(VP xuất

(NP viện)

(SBAR do

(S (NP-SBJ *-1) (VP chỉ bị

(NP sức ép nhẹ))))))

(./PU))

- Subject is noun phrase, predicate is adjective phrase.

Sentence indicating quality of thing:

Example:

Em học sinh này rất là ngoan.

(S (NP-SBJ Em học sinh này)

(ADJP (RP rất là/M) ngoan)

(./PU))

Exiting sentence:

Example:

Lúa đang ẩm ấp hàng trăm bó trên sân.

(S (NP-SBJ Lúa)

(ADJP đang ẩm ấp hàng trăm bó trên sân)

(./PU))

- Subject is NP, and predicate is PP.

Prepositional phrase has head word is preposition, such as: *bằng, của, như, để, trong, or ngoài*.

Examples:

Hàng này không phải của công ty tôi phân phối.

(S (NP-SBJ Hàng này)

(PP-PRD (ADJP không phải) (Cs-H của)

(S-EXT (NP-SBJ công ty tôi)

(VP phân phối)))

(PU .))

(S (NP-SBJ *Tiền này*) (PP-PRD để đóng học phí) (./PU))

Prepositional phrase indicating cause or effect

When word indicating cause or effect, such as *vì, tại, do, or bởi* is head of predicative, its post-modifier is complement (CMP).

Examples:

Việc này do lỗi của anh.

(S (NP-SBJ Việc này)

(PP-PRD (Cs-H do)

(NP-CMP lỗi của anh))

(PU .))

Việc này do nó không cẩn thận.

(S (NP-SBJ Việc này)

(PP-PRD (Cs-H do)

(S-CMP nó không cẩn thận))

(PU .))

Việc này không phải tại nó gây ra.

(S (NP-SBJ Việc này)

(PP-PRD không phải tại

(S-CMP nó gây ra))

(PU .))

Hàng này không do nó phân phối.

(S (NP-SBJ Hàng này)

(PP-PRD không do

(S-CMP nó phân phối))

(PU .))

Caution: Distinguish above structure from:

- Topicalized

Việc này nó không cẩn thận.

(S (NP-TPC-1 Việc này)

(NP-SBJ nó)

(VP không cẩn thận

NP-DOB *T*-1)

(PU .))

Above sentence should be written as *Nó không cẩn thận việc này.*

However, moving *việc này* to the beginning of the sentence is to emphasize.

- Subordinate clause

Hàng do nó phân phối rất rẻ.

(S (NP-SBJ Hàng (SBAR (Cs do) (S nó phân phối)))

(ADJP-PRD rất rẻ)

(PU .))

- Subject or predicate of a descriptive sentence which is a clause having complete meaning will be bracketed with S label.

Examples:

Anh thành công khiến em rất vui mừng.

(S (S-SBJ Anh thành công)

(VP khiến em rất vui mừng)

(./PU))

Chiếc xe này máy đã hỏng.

(S (NP-SBJ Chiếc xe này)

(S-PRD máy đã hỏng)

(./PU))

- Subject and predicate of a single descriptive sentence can be linked by the conjunction *thì*.

Đảng viên thì phải gương mẫu.

(S (NP-SBJ Đảng viên)

(Cp thì)

(VP phải gương mẫu)

(PU .))

Phụ nữ Việt Nam thì trung hậu, đảm đang.

(S (NP-SBJ Phụ nữ Việt Nam)

(Cp thì)

(ADJP trung hậu, đảm đang)

(PU .))

3.1.1.2. Inferential sentence.

This type of sentence expresses an inference that shows that X is equal Y, where, X is subject and Y is object. X and Y are linked by copula verb *là*. There are the forms of inferential sentences as follows:

- X and Y usually have the same syntactic type.

Example:

Steve Jobs là chủ tịch Apple.

(S (NP-SBJ Steve Jobs)

(VP là

(NP-CMP chủ_tịch Apple))

(./PU))

- X is a sentence (S), and Y is a NP.

Example:

Em làm bài này tốt là dấu hiệu của sự tiến bộ.

(S (S-SBJ Em làm bài này tốt)

(VP là

(NP-CMP dấu hiệu của sự tiến bộ))

(./PU))

Above subject can be changed to noun phrase having post-modifier is a sentence.

Example:

Việc em làm bài này tốt là dấu hiệu của sự tiến bộ.

(S (NP-SBJ Việc/Nn

(S em làm bài này tốt))

(VP là

(NP-CMP dấu hiệu của sự tiến bộ))

(./PU))

The subject can be changed to subordinate component.

Example:

Em làm bài này tốt, đó là dấu hiệu của sự tiến bộ.

(S (S-ADV Em làm bài này tốt)

(./PU)

(NP-SBJ đó)

(VP là dấu hiệu của sự tiến bộ)

(./PU))

- X is NP, and Y can be annotated with S.

Example:

Điều đáng chú ý là cô ấy rất thông minh.

(S (NP-SBJ Điều đáng chú ý)

(VP là

(S cô ấy rất thông minh))

(./PU))

- X is NP, and Y is PP.

Examples:

Ly này (VP không phải (Vc-H là) (PP-CMP bằng thuỷ tinh)).

Việc này (VP (Vc-H là) (PP-CMP do nó không cẩn thận)).

Tiền này (VP (Vc-H là) (PP-CMP để đóng học phí)).

3.1.1.3. Negative form

Adjunct *không* comes before head of predicate.

Examples:

Tôi không biết chuyện riêng của cô ấy.

(S (NP-SBJ Tôi)

(VP không biết chuyện riêng của cô ấy)

(./PU))

Steve Jobs không là chủ tịch Apple.

(S (NP-SBJ Steve Jobs)

(VP không là chủ tịch Apple)

(./PU))

3.1.2. Question

The sentences to ask for information will be bracketed with SQ label. There are the types of questions as follows:

Optional question

The optional idea is expressed by:

- Conjunction *hay* in equal coordination

Examples:

Chúng ta đi luôn hay về?

(SQ (NP-SBJ Chúng ta)

(VP (VP đi luôn)

(Cp hay)

(VP về))

(PU ?))

Anh hay tôi làm?

(SQ (NP-SBJ (Nn Anh) (Cp hay) (Pp tôi))

(VP làm)

(PU ?))

Anh muốn đi xem phim hay đi ăn?

(SQ (NP-SBJ Anh)

(VP (Vv-H muốn)

(VP (VP đi xem phim)

(Cp hay)

(VP đi ăn)))

(PU ?))

- Adjunct: *chưa, không, có...không; đã...chưa, etc.*

Example:

Anh có đi không?

(SQ (NP-SBJ Anh)

(VP (R có) (Vv-H đi) (R không))

(PU ?))

- The adjunct can be omitted.

Examples:

Anh đi không?

(SQ (NP-SBJ Anh)

(VP (Vv-H đi) (R không))

(PU ?))

- Combine conjunction *hay* and adjunct.

Examples:

(SQ Hôm nay anh có đi làm không, hay anh ở nhà?)

- Optional question to ask an object is existing or not.

Example:

(SQ Có cái ví ở đây không?)

Non-optional question:

- Question including interrogative pronouns such as: *ai, gì, sao, nào, thế_nào, đâu, bao_giờ, etc.*

Examples:

Ai đang làm việc này?

(SQ (QNP-SBJ Ai)

(VP đang làm việc này)

(PU ?))

Bao giờ mẹ về?

(SQ (QNP-TMP Bao_giờ)

(NP-SBJ mẹ)

(VP về)

(PU ?))

Vì sao chưa ai đến?

(SQ (Cp Vì)

(QNP-PRP sao)

(NP-SBJ chưa ai)

(VP đến)

(PU ?))

Bằng cách nào anh vào được đây?

(SQ (QPP-MNR (Cs-H Bằng)

(NP cách

(NP nào)))

(NP-SBJ anh)

(VP vào được đây)

(PU ?))

Steve Jobs sẽ làm gì trên cương vị Chủ tịch Apple?

(SQ (NP-SUB Steve_Jobs)

(VP sẽ làm

(QNP gì)

(PP-MNR trên cương_vị Chủ_tịch Apple))

(PU ?))

Suppositional question²

Is constructed by:

- Add interrogative modifiers: *à, u, ạ, nhỉ, hử, hả, chứ, chứ, bộ, or so on* to the end of declarative sentence.

Examples:

Anh mới về à?

(SQ (S (NP-SBJ Anh)

(VP mới về))

(M à)

(PU ?))

(SQ (S Anh không đi nữa) hả/M ?)

- Place the interrogative expression, such as *có không, có phải không, đúng không, có đúng không* at the end of declarative sentence.

Examples:

(SQ (S (NP Tôi qua) (NP anh) (VP thức khuya)) (ADJP có phải không) ?)

- Có phải là ... hay không

(SQ

(NP-SBJ (Nt-H khi)

(SBAR (Cs *0*))

(S

² Corresponds to **tag question** in English

(NP-SBJ (NONE E))
 (VP
 (VP (R không) (Vv-H khắc_phục) (R ngay))
 (Cp mà)
 (VP (Vv-H đề)
 (SBAR-CMP (Cs *0*)
 (S
 (NP-SBJ
 (Nn_swsp (Nc-H con) (Nn đường)))
 (VP (Vv-H xuống_cấp)
 (ADJP (Aa-H thâm))))))))))
 (PU ,)
 (VP
 (ADJP (R có) (Aa-H phải))
 (Vc-H là)
 (UCP-CMP
 (VP
 (VP (Vv-H-1 góp)
 (VP (Vv-H thêm))
 (NP-DOB (Nn-H phần)
 (VP (Vv-H làm)
 (ADJP-CMP (Aa-H xấu) (R đi)
 (NP (Nn-H môi_trường) (Vv đầu_tư)
 (PP (Cs-H của)
 (NP (Nun-H thành_phố))))))))))

(Cp hay)
 (VP (R không) (Vv *H*-1)))
 (PU ,)
 (Cp và)
 (ADJP
 (ADJP (R cũng) (Aa-H-2 đồng_nghĩa)
 (PP (Cs-H với)
 (NP
 (Nn_swsp (Ncs-H việc)
 (VP (Ve-H mất) (M đi)
 (NP-CMP (Pp bao_nhiều) (Nn-H
 cơ_hội)
 (VP
 (VP (Vv-H làm_việc))
 (Cp và)
 (VP (Vv-H cải_thiện)
 (NP (Nn-H cuộc_sống))))
 (PP (Cs-H của)
 (NP
 (Nn_swsp (Nc-H người) (Nn
 lao_động))
 (PU -)
 (NP-PRN (Nn-H
 đồng_bào))))))))))
 (Cp hay)
 (ADJP (R không) (Aa *H*-2))))))

(PU ?))

Intonation question

Imperative sentences ending with a question mark are also bracketed with SQ.

Examples:

Còn người thứ 3?

(SQ (Cp còn)

(NP người thứ 3)

(PU ?))

3.1.3. Special sentence (SPL)

Special sentence itself is complete construction containing one main component (possibly having subordinate component or topicalized). Therefore, it is not necessary to add main components.

Special sentence is bracketed with the label SPL

3.1.3.1. Special sentence having main component which is a noun phrase

Examples:

Nước!

(SPL (NP (Nn-H Nước))

(PU !))

A! biển!

(SPL (E A)

(PU !)

(NP (Nn-H biển))

(PU !))

(S Như/Cp (SPL vậy) ,/PU (S chúng ta đã hoàn thành công việc) .)

3.1.3.2. *Special sentence whose main component includes only verb phrase or adjective phrase*

- Self-evident existing sentence

Examples:

Cháy!

(SPL (VP Cháy)

(PU !))

Mưa rồi!

(SPL (VP Mưa rồi)

(PU !))

Giỏi!

(SPL-EXC (ADJP Giỏi)

(PU !))

Tốt lắm!

(SPL-EXC (ADJP Tốt lắm)

(PU !))

- General existing sentence: indicates existing, non-existing or no longer existing things. It can includes VP or ADJP.

Verb phrase (VP)

Construction: Ve-H (*có, còn, mất, hết, chết*) + NP/ADJP/S

Examples:

Có 3 người trong phòng.

(SPL (VP (Ve-H Có)

(NP 3 người trong phòng))

(PU .))

Còn 1 tháng nữa là đến Tết.

(SPL (VP (Ve-H Còn)

(S 1 tháng nữa là/Vc đến Tết))

(PU .))

Có 3 người bị thương.

(SPL (VP (Ve-H Có)

(S 3 người bị thương))

(PU .))

Có đám mây che mặt trăng.

(SPL (VP (Ve-H Có)

(S đám mây che mặt trăng))

(PU .))

Caution:

Existing sentences completely having subject and predicate are annotated with S label.

Example:

(S (NP-SBJ (Pp-H Tôi))

(VP có/Ve-H

(NP tiền))

(PU .))

Existing sentence includes an adjective phrase (ADJP)

Construction: Aa-H (*nhiều, ít*) + NP

Example:

Nhiều người quá!

(SPL-EXC (ADJP (Aa-H Nhiều)

(NP người)

(M quá))

(PU !))

- Locative existing sentence

Locative existing sentence indicates things exist at specific places. It can include VP or ADJP. Existing sentence can have existing verb, topicalized, locative sunbordinate component or not.

Verb phrase (VP)

Constructions:

Circumstantial constituents (B) + Ve-H (*có, còn, hết, mất, chết*) + NP (A)

Circumstantial constituents (B) + Vc-H (*là*) + NP (A)

Circumstantial constituents (B) + Vv-H (*ngồi, mọc, nở, trông, bày, đặt, treo, kê, để, kết, ...*) + NP (A)

In above constructions, B indicates location. A and B can be interchanged. In some cases, B follow verb.

Examples:

Có cái tivi trên bàn.

(SPL (VP (Ve-H có)

(NP cái tivi)

(PP-LOC trên bàn)

(PU .))

Trên bàn, có cái tivi.

(SPL (PP-LOC Trên bàn) ,/PU

(VP (Vc-H có)

(NP cái tivi))

(PU .))

Trên bàn là cái tivi.

(SPL (PP-LOC Trên bàn)

(VP (Vc-H là)

(NP cái tivi))

(PU .))

Bàn có cái tivi.

(SPL (NP-LOC Bàn)

(VP (Ve-H có)

(NP cái tivi))

(PU .))

Trong hồ nở một bông sen.

(SPL (PP-LOC Trong hồ)

(VP (Vv-H nở)

(NP một bông sen))

(PU .))

The existing sentence “A có ở B” will change to sentence having relational dependent predicate if missing existing verb.

Examples:

Gạo trong kho.

TV ở bàn.

Mọi người trong phòng.

Constructions:

Topicalized (A) + Ve-H (*có, còn, hết, mất, chết*) + NP (C)

Topicalized (A) + Vv-H (*ngồi, mọc, nở, chín, tròng, bày, đặt, treo, kê, để, kết, ...*) + NP (C)

Examples:

Vịt còn 2 con. (A có C)

(SPL (NP-TPC Vịt)

(VP (Ve-H còn)

(NP 2 con))

(PU .))

Địch chết 2 tên.

(SPL (NP-TPC Địch)

(VP (Ve-H chết)

(NP 2 tên))

(PU .))

Tiền mất 2 triệu đồng.

(SPL (NP-TPC Tiền)

(VP (Ve-H mất)

(NP 2 triệu đồng))

(PU .))

Đến tham dự có giám đốc công ty.

(SPL (VP-TPC Đến tham dự)

(VP (Ve-H có)

(NP giám đốc công ty))

(PU .))

Adjective phrase (ADJP)

Construction:

Circumstantial constituents (B) + Aa-H (*nhiều, ít, đông, đầy, vắng, thừa, lác_đác, lủi lo*) + NP (A)

In this construction, B and A can be interchanged.

Examples:

Trong nhà khúc khích tiếng cười.

(SPL (PP-LOC Trong nhà)

(ADJP (Aa-H khúc_khích)

(NP tiếng cười))

(PU .))

Khúc khích tiếng cười trong nhà.

(SPL (ADJP (Aa-H khúc_khích)

(NP tiếng cười)

(PP-LOC Trong nhà))

(PU .))

Caution:

Another construction of special sentence which its main component is also ADJP

Topicalized (A) + Aa-H (*nhiều, ít, đông, đầy, vắng, thừa*) + NP (C)

Example:

Tranh ít bức đẹp.

(SPL (NP-TPC Tranh)

(ADJP (Aa-H ít)

(NP bức đẹp))

(PU .))

Caution: Distinguish between above structures and the cases of sentence whose head verb indicating possessive, whole-partitive meaning, or relational meaning.

These sentences are not special sentences and annotated with S label.

Example:

Tôi có cái tivi.

(S (NP-SBJ Tôi)

(VP có cái tivi)

(PU .))

- Appearing and disappearing sentence.

It can include VP or ADJP

Verb phrase

Construction:

Temporal or locative circumstantial constituents (B) + Vv-H (*xuất hiện, mất tích, biến mất, mất cắp, đi, chạy, ló, nhô, nổ, ...*) + NP (A)

Examples:

Từ dưới nước nhô lên một cánh tay.

(SPL (PP-LOC *Từ dưới nước*)

(VP *nhô lên một cánh tay*)

(PU .))

Chiều tối nổ ra cuộc đình công.

(SPL (NP-TMP *Chiều tối*)

(VP *nổ ra cuộc đình công*)

(PU .))

Adjective phrase

Construction:

Temporal or locative circumstantial constituents (B) + Aa-H (*ùng ục, lấp lánh, ...*) + NP (A)

Example:

Xa xa lấp lánh ánh đèn.

(SPL (ADJP-LOC *Xa xa*)

(ADJP *lấp lánh ánh đèn*)

(PU .))

In addition, special sentence can be formed by exclamation words³ and negative adjuncts.

Examples:

(SPL *Trời ơi!*)

(SPL *Không!*)

33.1.4. Compound sentence

The compound sentence is annotated with S label. It contains 2 or more clauses having forms of simple sentence (subjective-predicative construction) or special sentence. There are types of compound sentence as follows:

3.1.4.1. Coordinating compound sentence

Clauses are linked by commas or conjunctions (Cp): *và, rồi, hay, còn, mà, or so on*. There are types of relationship between the clauses as follows:

- Succeeding relationship: conjunctions (Cp), such as: *và* and *rồi* are used

Example:

Mây tan rồi mưa tạnh.

(S (S Mây tan)

(Cp rồi)

(S mưa tạnh)

(PU .))

- Listing relationship

Example:

Chim kêu, vượn hú, thác đổ ầm ầm.

(S (S Chim kêu)

(PU ,)

3

Refer to exclamative sentence for more details.

(S vượn hú)

(PU ,)

(S thác đổ ầm ầm)

(PU .))

- Selection relationship

Example:

Hôm nay, bạn đi học hay bạn đi làm.

(S (NP-TMP hôm nay)

(PU ,)

(S bạn đi học)

(Cp hay)

(S bạn đi làm)

(PU .))

- Symmetry relationship

Example:

An học giỏi còn Lan học dở.

(S (S An học giỏi)

(Cp còn)

(S Lan học dở)

(PU .))

Ba đi làm còn mẹ đi chợ.

(S (S Ba đi làm)

(Cp còn)

(S mẹ đi chợ

(PU .))

For the compound sentence having circumstant component, if circumstant component modifies all sentence, the sentence is annotated according to structure:

(S (circumstant component) (Cp) (S1) (Cp) (S2) (PU))

If circumstant component modifies one of clauses, the sentence is annotated according to structure:

(S (Cp) (S1) (Cp) (S2) (PU))

3.1.4.2. Reciprocal compound sentence (câu ghép qua lại)

This type of compound sentence has 2 clauses. Where, the first clause is main premise, and the other is conclusion. The clauses cannot be inverted.

Clauses can be linked by pairs of conjunctions or adjuncts.

- Adjuncts come before predicate of each clause:

Examples:

Người ta vừa mở miệng, anh đã cắt ngang.

(S (S Người ta vừa mở miệng)

(PU ,)

(S anh đã cắt ngang)

(PU .))

Ông ấy càng nói, người ta càng làm ồn.

(S (S Ông ấy càng nói)

(PU ,)

(S người ta càng làm ồn)

(PU .))

- Conjunction comes before each clause:

There are pairs of conjunctions such as: *nếu ... thì, hễ ... thì, miễn (là) ... thì, (bởi) vì ... (cho) nên, (tại) vì ... (cho) nên, do ... (cho) nên, nhờ ... (cho) nên, tuy ... nhưng, (mặc) dù ... nhưng, hễ ... là, sở dĩ ... (là) vì.*

Example:

Vì trời mưa nên tôi đi học trễ.

(S (Cp Vì)

(S trời mưa)

(Cp nên)

(S tôi đi học trễ)

(PU .))

Conjunction can be omitted or replaced by a comma.

Examples:

Trời mưa nên tôi đi học trễ.

(S (S trời mưa)

(Cp nên)

(S tôi đi học trễ)

(PU .))

Vì trời mưa, tôi đi học trễ.

(S (Cp Vì)

(S trời mưa)

(PU ,)

(S tôi đi học trễ)

(PU .))

- Both conjunction and adjunct are used.

Examples:

*Vì trời trở rét **nên** cửa hàng áo ấm **mới** đông khách thế.*

(S (Cp Vì)

(S trời trở rét)

(Cp nên)

(S cửa hàng áo ấm mới đông khách thế)

(PU .))

Nếu vậy, mọi người ắt phải đồng tình.

(S (Cp Nếu)

(SPL vậy)

(PU ,)

(S mọi người ắt phải đồng tình)

(PU .))

- Both conjunction and adjunct are not used.

Example:

Khi nào mẹ về, con báo tin cho ba.

(S (S Khi nào mẹ về)

(PU ,)

(S con báo tin cho ba)

(PU .))

3.1.5. Mệnh đề phụ kết (SBAR)

- In a sentence, subordinate clause is a subordinate constituent (including subject and predicate) that is introduced by a conjunction (possibly missing), such as: mà, rằng, do, là, cho. Subordinate clause modifies noun, verb or adjective with additional information.

Construction:

(SBAR (Cs) (S))

(SBAR (Cs) (PU) (S))

Subordinate clause in complete sentence

Examples:

Nó đã đọc (NP-DOB quyển sách (SBAR (Cs mà) (S anh đưa))).

(NP-SBJ Người (SBAR (Cs mà) (S ai cũng quý))) là chị tôi đó.

(NP-SBJ Việc (SBAR (Cs mà) (S em làm bài này tốt))) đã cho thấy sự tiến bộ của em.

(NP-TMP Khi (SBAR (mà) (S tôi còn bé))), ba đã bỏ đi rồi.

Anh có nhớ (NP-LOC nơi (SBAR (mà) (S mình gặp nhau lần đầu))).

(NP-SBJ Hàng hoá (SBAR (do) (S công ty này cung cấp))) rất rẻ.

Tôi rất thích đọc (NP-DOB chuyên mục (SBAR (do) (S chị phụ trách))).

Tôi biết (SBAR-DOB (SBAR (rằng) anh không có lỗi)).

Cô ấy hỏi (SBAR (rằng) ai ném quả bóng).

Hắn nói (SBAR (là) tôi không có tài).

Đừng cho (SBAR (là/rằng) mình đúng).

Chúc (SBAR (cho) bà sống lâu trăm tuổi).

Trang mới nhất của cuộc_sống Thái_Mỹ bày ra trước_mắt (SBAR (Cs rằng) (PU :) (Scác em học_sinh đang đến trường dưới con đường xanh_ngắt bóng tre) .

In the case that conjunction (Cs) is omitted, we add the null word (*0*) to position of Cs.

Example:

*Nó đã đọc (NP-DOB quyển sách (SBAR (Cs *0*) (S anh đưa))).*

Subordinate clause is elliptical clause

Examples:

*Hội đồng tiếp tục buổi họp (PP-TMP trong lúc (SBAR (mà) (S *E* ăn trưa))).*

*Hội đồng tiếp tục buổi họp (PP-TMP trong lúc (SBAR (*0*) (S *E* ăn trưa))).*

- Explanatory component which is a clause following the conjunction vì (bởi_vì/do/nhờ) is also bracketed with SBAR

Ví dụ:

Tôi đến trễ vì trời mưa.

(S

(NP-SBJ Tôi)

(VP (V_v-H đến)

(ADJP trễ (SBAR vì/Cs (S trời mưa)))))

(PU .))

Tôi đến trễ vì mưa.

(S

(NP-SBJ Tôi)

(VP (V_v-H đến)

(ADJP trễ (SBAR vì/Cs (S *E* mưa)))))

(PU .))

(S-CMP

(NP-SBJ-1 (Num 17) (Nn-H lao_động))

(VP (V_v-H trở_về)

(PP-LOC (Cs-H từ)

(NP (Nr-H Malaysia))))

(SBAR-PRP (Cs vì)

(S

(NP-SBJ (NONE *-1))

(VP (R không) (Ve-H có)

(NP-CMP (Nn-H việc_làm))))))

17 lao động trở về từ Malaysia vì không có việc làm và vì công ty không thực hiện đúng hợp đồng như cam kết

(S-CMP

(NP-SBJ-1 (Num 17) (Nn-H lao_động))
 (VP (Vv-H trở_về)
 (PP-LOC (Cs-H từ)
 (NP (Nr-H Malaysia)))
 (SBAR-PRP
 (SBAR (Cs vì)
 (S
 (NP-SBJ (NONE *-1))
 (VP (R không) (Ve-H có)
 (NP-CMP (Nn-H việc_làm))))))
 (Cp và)
 (SBAR (Cs vì)
 (S
 (NP-SBJ (Nn-H công_ty))
 (VP (R không) (Vv-H thực_hiện)
 (ADJP (Aa-H đúng))
 (NP (Nn-H hợp_đồng))
 (PP (Cs-H như)
 (VP (R đã) (Vv-H cam_kết))))))))))

3.1.6. Note

- Simple sentence has subject which is a sentence, a noun phrase including relative clause or a normal subordinate construction.
 - Construction: Sbj (NP) + Prd1 (VP/ADJP/...) + Prd2 (VP/ADJP/...)
 - Prd1 and Prd2 do not create a coordinate or subordinate construction
- Criteria:

- If Sbj + Prd1 has a complete meaning (can independently become a simple sentence (can be changed tenses)), Sbj + Prd1 is a simple sentence playing the role as subject. In this case, we can not insert the subordinate conjunction *mà/là/rằng/...* between Sbj and Prd1

Examples.

Nam học giỏi khiến cha mẹ vui lòng.

(S (S-SBJ (NP-SBJ Nam)

(VP học giỏi))

(VP khiến (SBAR *0*/Cs (S cha mẹ vui lòng)))

(PU .))

In some cases, we can transform the subjective sentence into the noun phrase.

Việc Nam học giỏi khiến cha mẹ vui lòng.

(S (NP-SBJ Việc

(SBAR (Cs *0*))

(S (NP-SBJ Nam)

(VP học giỏi))))

(VP khiến cha mẹ vui lòng)

(PU .))

If a conjunction such as *và, hoặc, etc.* is inserted between Prd1 and Prd2, they are two predicates.

(S (NP-SBJ Loài khủng long mới này)

(UCP-PRD (VP được đặt tên là A)

,/PU

(ADJP cổ xưa hơn loài khủng long B)))

- Prd1 is treated as a post-modifier of the preceding word (Sbj + Prd1 is a subordinate construction) when we can add a demonstrative

pronoun such as *này, đó, ấy, ect.* after Prd1 so that demonstrative pronoun and Prd1 are post-modifiers of the subjective noun phrase.

Examples:

Xi măng sản xuất theo phương pháp truyền thống bán rất chạy.

(S

(NP-SBJ (Nn-H-1 Xi_măng)

(VP (Vv *P*)

(VP (Vv-H sản_xuất)

(NP-DOB *D*-1)

(VP-CMP (Vv-H theo)

(NP (Nn-H phương_pháp)

(ADJP (Aa-H truyền_thống))))))

(VP bán rất chạy)

(PU .))

Xi măng sản xuất theo phương pháp truyền thống này bán rất chạy.

(S

(NP-SBJ (Nn-H-1 Xi_măng)

(VP *P* (VP (Vv-H sản_xuất)

(NP *D*-1)

(VP-CMP (Vv-H theo)

(NP (Nn-H phương_pháp)

(ADJP (Aa-H truyền_thống))))))

(Pd này))

(VP bán rất chạy)

(PU .))

Tình_hình tội_phạm liên_quan đến thẻ ngày_càng gia_tăng.

(S

(NP-SBJ (Nn-H tình_hình)

(NP (Nn-H tội_phạm)

(VP (Vv-H liên_quan)

(PP-CMP (Cs-H đến)

(NP (Nn-H thẻ))))))

(VP (R ngày_càng) (Vv-H gia_tăng))

(PU .))

We can add the subordinate conjunction *mà/là/rằng/sao/cho* before Prd1 to recognize that Prd1 is a reduced relative clause modifying the previous noun or a predicate of the sentence. If we can add the subordinate conjunction *mà/là/rằng/sao/cho* before Prd1, it is a reduced relative clause modifying the previous noun.

Example:

Xi_măng <mà> sản_xuất theo phương_pháp truyền_thống bán_rất chạy.

(S

(NP-SBJ (Nn-H-1 Xi_măng)

(VP (Vv-H *P*)

(VP (Vv-H sản_xuất)

(NP *D*-1)

(VP-CMP (Vv-H theo)

(NP (Nn-H phương_pháp)

(ADJP (Aa-H truyền_thống))))))

(VP bán_rất_chạy)

(PU .))

(NP-SBJ Cô gái <mà> (VP ném quả bóng)) là chị tôi.

Một dịch_vụ khá mới <mà> được các NH đưa vào áp_dụng qua NH điện_tử là gửi tiết_kiệm trực_tuyến.

(S (NP-SBJ-1 Một/Num dịch_vụ/Nn-H

(ADJP khá mới)

(VP được (S các NH đưa *T*-1 vào áp_dụng qua NH điện_tử)))

(VP là gửi tiết_kiệm trực_tuyến)

(PU .))

In above example, we can see that *T* can be inserted after *đưa* or *áp_dụng*. However, in our treebank, we stipulate that *T* is inserted after the first verb *đưa*.

- Construction: Noun Prd1 Prd2

Should we add a subject before Prd1 to create a SBAR when Prd1 is a reduced relative clause (subject of the relative clause is omitted)?

If Noun is not the subject who perform the action stated by Prd1, we add *P* before Prd1 and annotate the expression *P* + Prd1 as a post-modifier of Noun.

Examples:

(NP-SBJ Điều/Nn-H-1 (VP cần *P* chú ý *T*-1)) là giữ bí mật với cô ấy.

(S (NP-SBJ Điều/Nn-H-1

(VP (Vv-H *P*) (VP lưu ý *T*-1)))

(VP là giữ bí mật với cô ấy)

(PU .))

(S

(NP-SBJ (Nn-H-1 Xi_măng)
 (VP (Vv *P*)
 (VP (Vv-H sản_xuất)
 (NP *D*-1)
 (VP-CMP (Vv-H theo)
 (NP (Nn-H phương_pháp)
 (ADJP (Aa-H
 truyền_thống))))))
 (VP bán rất chạy) F
 (PU .))

If Prd1 is a complete relative clause, we annotate it as a relative clause modifying the previous noun.

However, if Prd1 is a reduced relative clause including an adverbial phrase standing at the beginning of Prd1, we treat Prd1 as a complete relative clause, insert null element to the ellipses subject (insert after the adverbial phrase).

Ex:

(NP (QP (R hàng) (Nq-H ngàn))
 (Nn_swsp (Nc-H người) (Nn lính))
 (VP (R đã) (Vv-H hi_sinh))
 (SBAR (Cs *0*)
 (S
 (NP-TMP (Nt-H nay))
 (NP-SBJ (NONE *E*))
 (VP
 (RP (R vẫn) (R-H chưa))
 (Vv-H tìm)

(VP-CMP (Vv-H thấy)

(NP (Nn-H thân_xác))))))

- Distinguishing between simple sentence having equal coordination and equal compound sentence.

Convention: There is not elliptical equal compound sentence. It means that when subject or predicate of each clause in equal compound sentence is reduced because they are the same, the existing construction is simple sentence having equal coordination.

Examples:

Equal compound sentence:

Bàn bằng gỗ và ghế bằng gỗ.

(S

(S (NP-SBJ Bàn) (PP-PRD bằng gỗ))

(Cp và)

(S (NP-SBJ ghế) (PP-PRD bằng gỗ))

(PU .))

Equal coordination:

Bàn và ghế bằng gỗ.

(S

(NP-SBJ (Nn-H Bàn) (Cp và) (Nn-H ghế))

(PP-PRD bằng gỗ)

(PU .))

When subject of the second clause of the compound sentence is omitted (because it is the same the subject of the first clause), if the second clause has an adverbial phrase standing at the beginning, this sentence is a compound sentence.

(S (S (NP-SBJ-1 mình) (VP có bao_nhiều tiền gửi vào cũng được))

(Cp và)

(S (NP-TMP hằng tháng) (NP-SBJ *-1) (VP có_thể nộp thêm để tăng phần tiết_kiệm của mình))

(PU .))

- Distinguishing between single sentence having circumstantial constituent which is a sentence and reciprocal compound sentence.
 - o The circumstantial constituent modify main component (subject and predicate) about maner, location, temporal, etc. Its position in the sentence can be changes (at the beginning of sentence, between subject and predicate, after object, etc.).

Examples:

(S (S-MNR Mắt đeo kính trắng)

(PU ,)

(NP-SBJ người đàn ông)

(VP nhìn về phía nhà ga)

(PU .))

Subject of circumstantial constituent can be similar to main subject.

Example:

Đứng bên cửa sổ, hấn đưa mắt nhìn xuống dưới đường.

(S (S (NP-SBJ *-1) (VP-MNR Đứng bên cửa sổ))

(PU ,)

(NP-SBJ-1 hấn)

(VP đưa mắt nhìn xuống dưới đường)

(PU .))

If the subject of the first clause is omitted (ellipses clause) and the ellipses clause does not modify the sentence about maner, location,

temporal, etc., this sentence is bracketed as a compound sentence including ellipses clause.

(S (S (NP-SBJ *-I) (VP Được khai quật tại Motana))

(PU ,)

(S (NP-SBJ-I các mảnh xương hóa thạch) (VP đã giúp các nhà khoa học có cái nhìn mới))

(PU .))

- In reciprocal compound sentence, two clauses have principle-accessory relationship. Inverting the clauses causes changing of the meaning of sentence. Two clauses can be linked by conjunctions.

Example:

Trời mưa nên tôi đi học trễ.

(S (S Trời mưa)

(Cp nên)

(S tôi đi học trễ)

(PU .))

Do anh mách sếp mà tôi bị mắ'ng.

(S (Cp Do)

(S-PRP anh mách sếp)

(Cp mà)

(S tôi bị mắ'ng)

(PU .))

- The conjunction *thì* is also used to link two clauses of a compound sentence.

Examples:

Thế thì thôi.

(SPL (SPL Thế)

(Cp thì)

(SPL thôi)

(PU .))

Mưa nhưng không lạnh.

(SPL (SPL Mưa)

(Cp nhưng)

(SPL không lạnh)

(PU .))

In above sentences, each clause has complete meanings by itself. Therefore, we do not treated them as elliptical sentences. We tagged each clause as special sentence.

- Distinguishing simple sentence having subject which is a sentence and compound sentence that subjects of the latter clauses are omitted because they are the same the subject of the first clause.

Without loss of generalization, we assume that the ambiguous sentence is constructed as follows:

(S Sbj Prd1 Prd2 PU)

Where, Prd1 and Prd2 are 2 predicates, and Sbj is subject.

- o In simple sentence having subject which is a sentence, Prd1 and Prd2 do not have equal relationship or principle-accessory relationship. Combination of Sbj and Prd1 which is a sentence having complete meaning is subject of action, quantity, etc. stated by Prd2. We also can not insert an subordinate conjunction such as mà, là, rằng, etc. between Sbj and Prd1.

Example:

Nam học giỏi khiến cha mẹ vui long.

(S (S-SBJ (NP-SBJ Nam)

(VP học giỏi))

(VP khiến cha mẹ vui lòng)

(PU .))

In some cases, we can convert the sentence playing the role of subject into noun phrase.

Examples.

Việc Nam học giỏi khiến cha mẹ vui lòng.

(S (NP-SBJ Việc

(S (NP-SBJ Nam)

(VP học giỏi)))

(VP khiến cha mẹ vui lòng)

(PU .))

- Equal compound sentence that subjects of the latter clauses are omitted because they are similar to subject of the first clause will be treated as a single sentence having equal coordination.

Examples:

Tre giữ làng, giữ nước.

(S (NP-SBJ Tre)

(VP (VP giữ làng)

(PU ,)

(VP giữ nước))

(PU .))

- In the subordinate compound sentences that subjects of the latter clauses are omitted because they are similar to subjects of the first clauses, Prd1 and Prd2 have principle-accessory relationships
 - If two clauses are combined by a pair of conjunctions where the first conjunction stands at the beginning of the sentence,

we mark the elliptical subject with * and annotate the sentence as a compound sentence.

Example:

Vì tôi bị bệnh nên nghỉ học.

(S (Cp Vi)

(S (NP-SBJ-1 tôi) (VP bị bệnh))

(Cp nên)

(S (NP-SBJ *-1) (VP nghỉ học))

(PU .))

Tuy tôi bị bệnh nhưng vẫn đi học.

(S

(Cp Tuy)

(S (NP-SBJ-1 Tôi) (VP bị bệnh))

(Cp nhưng)

(S (NP-SBJ *-1) (VP vẫn đi học))

(PU .))

- If the first conjunction is omitted, the sentence is annotated as a simple sentence having coordinating predicate (two predicates with equal combination)

Example:

Tôi bị bệnh nhưng vẫn đi học.

(S (NP-SBJ Tôi)

(VP (VP bị bệnh)

(Cp nhưng)

(VP vẫn đi học))

(PU .))

- Single sentences that their predicates linked by pair of adjuncts.

Ex:

Nó càng học càng giỏi.

(S (NP-SBJ Nó)

(UCP càng/R (VP học) càng/R (ADJP giỏi))

(PU .))

Nó vừa đẹp vừa giỏi.

(S (NP-SBJ Nó)

(ADJP-PRD vừa/R đẹp vừa/R giỏi)

(PU .))

- Không_những/không_chỉ ... mà còn ...

If *không_những ... mà còn ...* links 2 predicates, the sentence is treated as single sentence having reciprocal coordination.

Example:

Nó không những đẹp mà còn giỏi.

(S (NP-SBJ Nó)

(ADJP-PRD không_những/Cp

(ADJP đẹp)

(CONJP mà/Cp còn/R)

(ADJP giỏi))

(PU .))

If *không_những/không_chỉ ... mà còn ...* links 2 subjects, the sentence is treated as single sentence having reciprocal coordination.

Example:

Không chỉ cô mà tôi cũng mệt_mỏi.

(S (NP-SBJ (Cs Không_chỉ)

(NP cô)

(Cs mà)

(NP tôi))

(ADJP-PRD cũng mệt_mỏi)

(PU .))

Otherwise, the sentence is treated as compound sentence.

Examples:

Không_những nó chăm học mà nó còn chăm làm.

(S (Cp Không_những)

(S (NP-SBJ nó) (ADJP-PRD chăm học))

(Cp mà)

(S (NP-SBJ nó) (ADJP còn chăm làm))

(PU .))

Không_những chăm học, nó còn chăm làm.

(S (Cp Không_những)

(S (NP-SBJ *-1) (ADJP-PRD chăm học))

(Cp mà)

(S (NP-SBJ-1 nó) (ADJP-PRD còn chăm làm))

(PU .))

Không_những nó chăm học mà còn chăm làm.

(S (Cs Không_những)

(S (NP-SBJ-1 nó) (ADJP-PRD chăm học))

(CONJP mà còn)

(S (NP-SBJ *-1) (ADJP-PRD chăm làm))

(PU .))

Nó không_những chăm học mà nó còn chăm làm.

(S (NP-TPC-1 Nó)

(Cp không_những)

(S (NP-SBJ *-1) (ADJP-PRD chăm học))

(Cp mà)

(S (NP-SBJ nó) (ADJP-PRD còn chăm làm))

(PU .))

- Simple sentences including 2 predicates (S = Sbj Prd1 Prd2)

When Prd1 is a sentence or its post-modifier (modifier following the head verb) is a sentence, it is difficult to distinguish that Prd2 is a predicate or Prd1 and Prd2 create a suborsinate construction. We base on meaning to distinguish these structures. If Prd2 talk about Sbj, this sentence has 2 predicate. Otherwise, Prd2 is a component of Prd1.

Example:

(S

(NP-SBJ Lãnh_đạo tỉnh Bình_Định)

(VP

(VP đã triển_khai

(S các đoàn công_tác về các huyện , thị_xã ,
thành_phố trong tỉnh để triển_khai công_tác đối_phó
với bão số 4 , lên phương_án di_dời dân ở các vùng
triều_cường và lũ quét))

(PU ,)

(VP yêu_cầu các sở , ngành trong tỉnh tổ_chức trực 24[24])

(PU ...))

- The conjunction *thì*

Thì can link components of a simple sentence

(S (NP-SBJ Đơn vị thi công) (Cp thì) (ADJP-PRD quá chậm chạp) (PU .))

(S

(NP-SBJ Minh)

(VP (VP dư 5-7 triệu đồng trong tài_khoản)

(Cp thì)

(VP cứ để đó))

(PU .))

(S

(NP-SBJ Minh)

(VP (VP dư 5-7 triệu đồng trong tài_khoản)

(Cp thì)

(VP cứ để đó, (SBAR vì nghĩ *0* với số tiền ít mà * phải ra NH làm sổ tiết_kiệm thì mất thời_gian)))

(PU .))

Thì links the clauses of a compound sentence

(S (S Minh dư 5-7 triệu đồng trong tài_khoản)

(Cp thì)

(S mình cứ để đó)

(PU .))

- Object is a simple sentence, a subordinate construction, or a coordinating construction

This ambiguity occurs when object have the subject-predicate structure (NP VP/ADJP/...)

- o If we can insert mà/là/rằng/sao/cho/... before NP, NP VP/ADJP/... is a simple sentence

*Nam học giỏi (VP khiến *0* (S cha mẹ vui lòng)).*

- If we can insert *mà/là/rằng/sao/cho/...* before VP/ADJP/..., VP/ADJP/... is not a predicate

Tôi (VP cần (NP-DOB người (VP giúp việc) <mà> (VP giỏi nội trợ))).

- In the case that we can not insert a subordinate conjunction to recognize the correct construction, if we insert an adjunct indicating tense before VP/ADJP/... to change tense of the sentence, NP VP/ADJP/... is a simple sentence.

Mẹ (VP đưa (NP con) (VP đi học)).

- Double quotation mark

- Used to quote: externally marks quoted components.

(S “/PU (S Công ty đã không đủ khả năng tồn tại qua sang năm) ”/PU ,/PU (VP-MNR theo ông A) ./PU)

- Used to emphasize an phrase: externally marks emphasized phrase

(NP (Nq những) (PU ") (Nn-H giao_dịch)

(ADJP (Aa-H ngẫm))

(PU ")

(NP-DOB

(Nn_swsp (Nc-H-1 bài) (Nn báo))

(PU ")

(SQ-TTL

(NP (Num một) (Nn-H âm_mưu) (Vv tổng_tiền))

(PU ?))

(PU ")

(VP (Vv-H bị)

(ADJP-DOB (PU ") (Aa-H thui_chột) (PU ")

(NP (Num một) (Nn-H phần))

(PP (Cs-H về)

(NP (Nn-H thành_tích) (Nn-H chuyên_môn))))))

- Period is mark of quoted sentence or punctuation mark.

Complex sentences which the quoted component is at the end and not marked by double quotation mark or parenthesis are considered having no period. They are marked by punctuation mark (direct child of the root) at the end.

(S

(NP-SBJ (Pp-H Tui))

(VP (Vv-H hỏi)

(SBAR (Cs 0) (PU ,)

(SQ

(NP-SBJ

(Nn_swsp (Nc-H ngôi) (Nn nhà))

(ADJP (Aa-H to)))

(VP (R có) (Vv-H sinh_lợi) (R được) (R không))))))

(PU ?))

- The root is marked S in case of compound sentences which their clauses are special sentences.

(S

SPL

<conjunction>

SPL

(PU))

3.2. Null elements

3.2.1. Trace of movement (*T*)

This tag is to mark the gapping because of movement. The unit is moved to other positions in sentence to change its function (passive sentence, topicalized, quotation)

Examples:

Nó bị bạn bè chê cười.

(S (NP-SBJ-1 Nó)
(VP (V_v-H bị)
(S-DOB (NP-SBJ bạn_bè)
(VP chê cười
(NP-DOB *T*-1))))
(PU .))

Khuyết điểm này tôi đang cố khắc phục.

(S (NP-TPC-1 Khuyết_điểm_này)
(NP-SBJ tôi)
(VP đang cố khắc_phục
(NP *T*-1))
(PU .))

" Chỗ này xưa là đồn_bót nè ! Kia là những ngôi nhà vách đất ." - anh Phan_Văn_Du giới_thiệu với khách bằng hai cùi_tay cụt_lùn.

(S (PU "
(S-1
(S
(NP-SUB (Nn Chỗ) (Pd này))
(NP-TMP (Nt xưa))

(VP (Vc là)
 (NP (Nn đồn_bót)))
 (M nè)
 (PU !))
 (S
 (NP-SUB (Pd Kia))
 (VP (Vc là)
 (NP (Nq những) (Nc ngôi) (Nn nhà)
 (NP (Nn vách)
 (NP (Nn đất))))))
 (PU .)))
 (PU ")
 (PU -)
 (NP-SUB (Nn anh) (Nr Phan_Văn_Đu))
 (VP (Vv giới_thiệu)
 (PP (Cs với)
 (NP (Nn khách)))
 (PP-MNR (Cs bằng)
 (NP (Num hai) (Nn cùi_tay) (Aa cụt_lũn)))
 (SBAR (Cs *0*)
 (S *T*-1)))
 (PU .))

In passive sentence, direct object of the main verb should be inserted right behind the main verb.

(S
 (NP-TMP (Nt-H Hiện))
 (NP-SBJ-1 (Nn-H khoảng_cách)

(ADJP (Aa-H chênh_lệch)
 (NP (Nn-H giá))
 (PP (Cs-H giữa)
 (NP (Num hai) (Nn-H thi_trường))))))
 (VP (R lại) (Vv-H được)
 (S-DOB
 (NP-SBJ (NONE *E*))
 (VP (Vv-H nói)
 (NP-DOB-1 (NONE *T*))
 (ADJP-CMP (Aa-H rộng)
 (VP (Vv-H thêm) (R lên)
 (NP-CMP (Nn-H mức)
 (NP
 (QP (Num-H 4) (Nq-H triệu))
 (Nu-H đồng[lượng))))))))))
 (PU .))

3.2.2. Ellipses without trace (*E*)

This tag is to mark the omitting of main components of sentence (subject, predicate, object, or so on) without trace.

In passive sentence

Example:

We have a complete passive sentence as follows:

Nó bị bạn bè chê cười.

When active subject following passive verb is omitted, we mark the ellipsis with *E*.

Nó bị chê cười.

(S (NP-SBJ-1 Nó)

(VP (Vp-H bị)
 (S (NP-SBJ (NONE *E*))
 (VP chê cười
 (NP-DOB *T*-1))))

(PU .))

In elliptical sentence

Omitting predicate:

Example:

Ai đang đi phía trước?

*Nam (VP *E*)*

Omitting object:

Example:

*Cơ quan này không có quyền cấp (NP-DOB đất) (PP-IOB *E*).*

Caution

Ditransitive verbs (give-get), such as: *cho, tặng, đưa, trao, gửi, trả, cấp*, or so on require 2 objects as post-modifiers. Hence, when both or one of them disappears, have to replace by null elements *E* or *.

Example:

*Nếu anh tặng (NP-DOB-1 quà) (PP-IOB-2 cho cô ấy) thì tôi cũng tặng (NP-DOB *-1) (PP-IOB *-2).*

3.2.3. Null sentence component as it appears at other position of the sentence.

(*)

This label is a combination of *T* and *E*. It exists in compound sentence with topicalized.

Examples:

Vì nó ốm mà không đi làm được.

(S (Cs Vi)

(S (NP-SBJ-1 (Nn-H Nó)
 (ADJP-PRD-PRP ốm))
 (Cs mà)
 (S (NP-SBJ *-1)
 (VP không đi làm được))
 (PU .))

Elliptical clause

Example:

Đứng bên cửa sổ, hấn đưa mắt nhìn xuống dưới đường.

(S (S (NP-SBJ *-1) (VP-MNR Đứng bên cửa sổ))
 (PU ,)
 (NP-SBJ-1 hấn)
 (VP đưa mắt nhìn xuống dưới đường)
 (PU .))

3.2.4. Null complementizer (*0*)

If the conjunction (such as: *mà, rằng*) introducing a relative clause is omitted, the gap will be marked with *0*.

Example:

*Con đường (SBAR *0* tôi đi qua) có rất nhiều cây xanh.*

3.2.5. Null passive verb (*P*)

If passive verb (such as: *bị, được, ...*) is omitted, it's position is marked *P*.

Example:

(NP Xi măng (VP được (VP sản xuất theo dạng này))) bán rất chạy.

*(NP Xi măng (VP *P* sản xuất theo dạng này)) bán rất chạy.*

3.2.6. Null head word (*H*)

If head word of phrase is omitted, it's position is marked *H*.

Example:

Vốn ít đương_nhiên phải đi vay, nhưng nghịch_lý là ở nước ta vay các ngân_hàng làm nguồn chính .

(S
(NP-SBJ (Nn-H nghịch_lý))
(VP (Vc-H là)
(S-CMP
(PP-LOC (Cs-H ở)
(NP (Nn-H nước)
(NP (Pp-H ta))))
(NP-SBJ-3 (NONE-H *H*-4)
(VP (Vv-H *P*)
(VP (Vv-H vay)
(NP (Nq các) (Nn-H ngân_hàng))))
(VP (Vv-H làm)
(NP-CMP (Nn-H-4 nguồn)
(ADJP (Aa-H chính))))))
(PU .))

Cháu bị tóm lần nào chưa?

Rồi chú ạ.

(S
(NP-SBJ-1 (NONE *E*))
(VP (Vv *P*)
(VP-CMP (Vv *H*)
(NP-DOB (NONE *T*-1))
(R Rồi)))

(NP-VOC (Pp-H chú) (M ạ))

(PU .))

If phrase which refers to head work of the other phrase is omitted, it is also marked “*H*”.

(NP-SBJ

(QP (Num-H 3) (PU -) (Num-H 4))

(Nn-H-1 nhân_viên)

(SBAR (Cs *0*))

(S (NP-TMP (Nt-H ngày) (Pd nào))

(NP-SBJ (Nn-H *H*-1))

(VP (R cũng) (Vv-H chở)

(NP-DOB (Nn-H hàng))

(VP (Vv-H về)

(NP (Nr-H Biên_Hoà)))))))

3.2.7. Null post-modifier of a verb (*D*)

If direct object of an verb is omitted, it is marked *D*.

Example:

(S

(NP-SBJ (Nn-H-1 Xi_mãng)

(VP (Vv *P*)

(VP (Vv-H sản_xuất)

(NP-DOB *D*-1)

(VP-CMP (Vv-H theo)

(NP (Nn-H phương_pháp)

(ADJP (Aa-H truyền_thống))))))

(VP bán rất chạy)

(PU .))

Chú ý: If head word of an phrase has post-modifier (explanation component) is an passive clause with omitted subject, this clause is made VP. *D* is used to make omitted DOB. The head word considered above is referred by *D*. However, if the passive clause has no omitted component, *D* is replaced by *T* that refer to subject of the clause. (Take a look at examples in passive sentence part).

3.2.8. Null post modifier as it is a part of a sum considered later (*SUM*)

Structure:

(Phrase 1

(Phrase 2 <A>))

(Phrase 3

(Phrase 4))

(C = A + B)

In the data, A and B are not written and replaced by C, we marked omitted component *SUM* and mark reference indices as follow:

(Phase

(Phrase 1

*(Phrase 2 *SUM*-1))*

(Phrase 3

*(Phrase 4 *SUM*-1))*

(C-PRN-1 ...))

Example:

(VP (R đã) (V_v-H tổ_chức)

(VP-CMP

(VP (V_v-H bàn_giao)

(NP-DOB (N_n-H thư_viện) (N_n sách) (VP *SUM*-2))

(PP-IOB *-1))

(PU ,)

(VP (Vv-H tặng)

(NP-DOB

(NP (Num 450) (Nn-H áo_ấm) (VP *SUM*-2))

(PU ,)

(NP (Num 450) (Nn-H mũ) (VP *SUM*-2)))

(PP-IOB-1 (Cs-H cho)

(NP (Nn-H học_sinh)

(NP (Nn-H Trường) (Nn tiểu_học))))

(VP-PRN-1 (Vv-H trị_giá)

(NP-CMP

(QP (Num-H 150) (Nq-H triệu))

(Nu-H đồng))))))

3.2.9. Note

- Criteria of null element insertion:

Example:

Vốn ít đương_nhiên phải đi vay, nhưng nghịch_lý là ở nước ta vay các ngân_hàng làm nguồn chính .

Inserting null element as follow:

1) *Vốn ít đương_nhiên phải đi vay, nhưng nghịch_lý là ở nước ta **vốn được** vay các ngân_hàng làm nguồn chính .*

Or

2) *Vốn ít đương_nhiên phải đi vay, nhưng nghịch_lý là ở nước ta **người ta** vay **vốn** các ngân_hàng làm nguồn chính .*

Criteria:

- Base on meaning of existing supplementary component behind
- Insert null word based on existing words
- Inserting as less as possible (without accounting null of passive structure)

In the above example, omitted component is subject of *vay các ngân_hàng làm nguồn chính*. We see that, *nguồn is nguồn vốn* —> Inserting based on criterion 1).

- in case of omitted subject, if sentence has (thành phần ngoài nòng cốt) component (adverbial phrase), we add subject right behind adverbial phrase.
- Word types for null elements :
 - *0* gán Cs
 - *H* gán theo từ loại của từ mà nó tham chiếu tới
 - *P* gán Vv
 - *T*, *E*, *, *SUM* gán là NONE
 - *D* gán cùng từ loại với từ mà nó refer tới.

3.3. Coindex

Identity index: Ordinal number following constituent tags in order to mark this part having corresponding null elements *T*, *, *H*, or *D*.

Reference index: Ordinal number following null element that correspond to identity index in order to represent that it should have the missing constituent tags

Lưu ý:

Number assigned for identity index and reference index must be the same.

Examples:

Khuyết điểm này tôi đang cố khắc phục.

(S (NP-TPC-1 Khuyết_điểm này)

(NP-SBJ tôi)

(VP đang cố khắc_phục

(NP *T*-1))

(PU .))

Tôi không đi làm được vì ốm.

(S (S (NP-SBJ-1 Tôi)

(VP không đi làm được))

(Cs vì)

(S-PRP (NP-SBJ *-1)

(ADJP-PRD ốm))

(PU .))

If there are many references, Null component refers to the nearest phrase (base on structure).

(S (PU ")

(S

(S-TMP

(NP-SBJ (Nn-H Trắng))

(VP (Vv-H lặn)))

(NP-SBJ (Pp-H tui))

(VP (R mới) (Vv-H về)

(PP-LOC (Cs-H đến)

(NP (Nn-H chòi))))))

(PU ,)

(S

(S

(NP-SBJ (NONE *-1))

(VP-MNR (V_v-H thấy)
 (SBAR (Cs 0)
 (S
 (NP-SBJ (Num hai) (Nn-H mẹ) (Nn-H con))
 (VP (V_v-H ôm)
 (NP (Nn-H nhau))
 (VP (V_v-H khóc))))))
 (PU ,)
 (NP-SBJ (Nn-H ruột_gan)
 (**NP-1 (Pp-H tui)**))
 (VP (R cứ) (V_v-H dứt)
 (NP-DOB (Pp từng) (Nn-H đoạn))))
 (PU ")
 (PU .))

If there are many null elements in the sentence, reference indices are assigned from left to right.

If the sentence has quotation, the quotation is assigned reference index first, then null element from left to right.

3.4. Phrase level

Phrase is different from sentence. Phrase has only one main component. Main component can be modified by one or more supplement components. Relation between main component and supplement components is subordinator.

3.4.1. General norms

3.4.1.1. Independent or dependent construction

In a phrase including head word and modifiers, if modifiers (a word or a multi-word expression) describe characteristics of head word, it will be bracketed with suitable constituent tag.

Identifying describing component:

- if description is a word, we can insert supplement words to make it a phrase. It means we can insert words between head word and the description.

Example:

(NP mèo/Nn-H (ADJP (Aa đen)))

(NP mèo/Nn-H (ADJP (R rất) (Aa đen)))

(NP số_lượng/Nn-H (NP du_khách/Nn))

(NP số_lượng/Nn-H (PP của/Cs-H (NP du_khách/Nn)))

These cases are different from the following case:

(NP Nguyễn_tắc/Nn-H hoạt_động/Vv) → annotate cấu trúc phẳng vì không thể chèn kết từ hay phụ từ trước hoạt động để tạo thành ngữ

- If structure includes 2 words A and B, if C can be inserted behind B that modifies B (B C is expression that modifies A), b is bracketed as phrase.

Ex:

(NP Giá/Nn-H

(NP cước/Nn-H))

(NP Giá/Nn-H

(NP cước/Nn-H)

(VP vận chuyển hành khách nội địa))

However, the expressions to identify/name a person or thing should be annotated with flat structure, such as:

- Expression of classifier noun (Nc) and common noun

Example:

(NP con/Nc-H mèo/Nn)

- Expression of special classifier noun (Ncs) and noun/verb/adjective

Example:

(NP nhà/Ncs-H nghiên_cứu/Vv)

- Expressions of a common noun (Nn) and a personal noun (Nr)

Example:

(NP bà/Nn-H Hoa/Nr)

(NP công_ty/Nn-H Hoa_Mai/Nr)

- When sequence includes common noun (Nn) and personal name contains more than one word, if the personal name contains subsequence with subordinative relation, we annotate phrases properly. If personal name contains words with parallel relation, we use flat structure.

Examples:

(NP viện/Nn-H Kiểm_sát/Vv Nhân_dân/Nn)

(NP trường/Nn-H

(NP đại_học/Nn-H

(UCP (NP Khoa_học/Nn Xã_hội/Nn)

và/Cp

(ADJP Nhân_văn/Aa))

(NP Tp./Nun-H HCM))

(NP Cục

(NP Cảnh_sát

(VP (VP (VP phòng cháy) ,/PU (VP chữa cháy))

và/Cp

(VP (VP cứu nạn) ,/PU (VP cứu hộ))))))

If personal name contains noun refers to administrative agency, it is grouped to a phrase. This personal name shows name of an organization of an administrative agency.

(NP trường/Nn-H (NP đại_học/Nn-H (UCP Khoa_học/Nn Xã_hội/Nn và/Cp Nhân_văn/Aa) (NP Tp./Nun-H HCM/Nr)))

Note:

If there are describing factors between common noun and personal name, it is bracketed with the same level as head word.

Example:

(NP bà/Nn-H giám_đốc/Nn Hoa/Nr)

(NP nhà_hàng/Nn-H hải_sản Ngọc_Sương/Nr)

If the describing factor is subordinative construction, it is bracketed with proper phrase tag.

(NP công_ty/Nn-H (VP tư_vấn/Vv-H (VP du_học/Vv-H)) Á-Âu/Nr)

(NP công_ty/Nn-H (NP trách_nhiệm/Nn-H (ADJP hữu_hạn/Aa-H)) (NP 1/Num thành_viên/Nn-H) An_Hoà/Nr)

(NP Ông (UCP (NP chủ_tịch (NP hội_đồng_quản_trị)) (VP kiêm (S người sáng_lập_công_ty)))) Nguyễn_Văn_A)

(NP Viện/Nn-H (VP nghiên_cứu/Vv-H (VP quản_lý/Vv-H (NP kinh_tế)))) (NP TU/Nn-H)

- Expression of special unit noun (Nun) and Nr

Example:

(NP thành_phố/Nun-H Hồ_Chí_Minh/Nr)

Note:

In case of description between Nun and Nr, if description can not be converted to phrase by inserting word, it is considered as having the same level as head word.

(NP thành_phố/Nun-H biển/Nn Nha_Trang/Nr)

However, if the conversion by word insertion is possible, the description is bracketed as phrase.

Example:

(NP khu/Nun-H dân_cw/Nn (ADJP mới/Aa-H) Xuân_Hải/Nr)

(NP khu/Nun-H dân_cw/Nn (ADJP rất mới/Aa-H) Xuân_Hải/Nr)

If description between Nun and Nr is subordinative construction, It is bracketed constituent tag.

Example:

*(NP thành_phố/Nun-H (NP hoa/Nn-H phượng/Nn (ADJP (Aa-H đỏ)))
Hải_Phòng/Nr)*

- Time formulation

Example:

(PP trước/Cs (NP ngày/Nt-H 20-03-2014/Nt))

- Sequence includes classifier and common noun

Example:

(NP dẫu/Nn-H quốc_huy/Nn)

(NP Khu/Nn-H rừng/Nn)

(NP Hoa/Nn-H loa_kèn/Nn)

(NP Cá/Nn-H chép/Nn)

If classifier is modified by a verb, it is bracketed flat structure. However, if the verb has modifiers, all modifiers of the classifier are grouped as verbs phrase.

(NP máy/Nn-H nổ/Vv)

(NP (Nn_swsp máy/Nn-H (VP xay/Vv-H (NP sinh_tổ/Nn-H))))

(NP lễ/Nn-H khởi_động/Vv)

*(NP lễ/Nn-H (VP khởi_động/Vv-H (NP chương_trình/Nn-H (VP phòng
chống tội phạm))))*

- An adjunct should be annotated the same level with the words they modify.

Examples:

(NP hoa/Nn-H cúc/Nn (ADJP rất/R đẹp/Aa-H)) {very beautiful flower}

(VP đi/Vv-H (ADJP quá/R nhanh/Aa-H)) {go too fast}

- Demonstrative pronouns should be annotated the same level with the words they modify.

Example:

(NP con/Nc-H mèo/Nn này/Pd)

- Ordinal number should also be annotated the same level with the words they modify.

Example:

tôi đứng (NP hạng/Nn-H 2/An)

(NP cầu/Nn-H Bình_Triệu/Nr 2/An)

In the case of structure: <head> <modifier 1> <modifier 2>, if annotating modifier 2 as modifier for modifier 1 or head word do not change the meaning of the sentence, we choose the first choice to annotate.

Example:

(NP-SBJ (Nn-H Lượng)

(NP (Nn-H du_khách)

(VP (Vv-H đồ) (D về)

(NP-CMP (Nr-H Sa_Pa) (PU ,) (Nr-H Lào_Cai))))))

(VP (Vv-H vay)

(NP (Nn-H vốn)

(PP (Cs-H từ)

(NP các ngân hàng))))))

When phrase has structure: head word + Ph1 + Ph2 + Ph3, where Ph3 is post-modifier of Ph1, we annotate as follows:

(S

(NP-SBJ (Pp-H người_ta))

(VP (Vv-H phải)

(VP-CMP (Vv-H mang)
 (NP-DOB (Nn-H người) (Vv chết))
 (VP (Vv-H lên)
 (NP-CMP (Nn-H vùng) (Nr Bẫy_Núi)
 (UCP-PRN (NONE *T*-1)))
 (PP-PRP (Nr-H để)
 (VP (Vv-H chôn_cất)))
 (UCP-PRN-1
 (ADJP (Aa-H xa_xôi))
 (PU ,)
 (VP (Vv-H tồn_kém))
 (PU ,)
 (S
 (NP-SBJ (R chỉ) (Nn-H nhà) (Aa giàu))
 (VP (R mới) (Vv-H lo) (R nổi)))))))))

when two phrases have Coordination, if head of the second phrase is omitted (because it is the same as the first phrase's), head of the second phrase is bracketed null element.

(NP-TPC-2 (Nn-H chuyện)
 (VP
 (VP (Vv-H-1 tồn_tại) (Cp và) (Vv-H-1 phát_triển))
 (Cp hay)
 (VP (R không) (Vv *H*-1))))

3.4.1.2. Coordination

- Words having the same syntactic function are organized at the word-level and are annotated with flat structure. In this case, the words are separated by commas or conjunctions (*và*, *hoặc*, or *hay*).

Example:

(NP Bố/Nn ,/PU mẹ/Nn và/Cp tôi/Pp) cùng đi xem phim {(NP father, mother and me) go to cinema together}

- CExpressions having the same syntactic function (separated by commas or conjunctions (*và*, *hoặc*, or *hay*)) will have the same father node.

Examples:

Mỗi học sinh giỏi được thưởng (NP (NP một cái cặp) , (NP 2 cuốn vở) và (NP một cây viết mực)) {Each good student is rewarded (NP (NP a briefcase), (NP 2 notebooks) and (NP a pen))}

Cô ấy (ADJP (ADJP dễ bắt chuyện) và/Cp (ADJP nhiệt tình))

Nước (VP chảy (UCP (ADJP rất mạnh) và (VP xiết)))

(NP (NP Bố An), (NP mẹ Hoa) và (NP tôi)) cùng đi xem phim

In the cases that the words/expressions having the same syntactic function are not separated by commas or conjunctions, they are still treated similar above.

Examples:

(NP Công_nhân nông_dân) đều sản_xuất giỏi.

Học_sinh phải (VP (VP học_tập tốt) (VP lao_động tốt)).

Note: In case preposition relate two formulations with the same construction and that structure is not coordinator

Anh có thể chọn (NP 2 môn/Nn-H (PP trong/Cs (NP 3 môn))) để học

There are types of coordination as follows:

3.4.1.2.1. Equal coordination

Construction:

Từ Cp/PU từ

Ngữ Cp/PU ngữ

- Enumerative relation: The words or expressions are linked by comma (,) or parallel conjunctions such as: *và, lẫn, cùng, với, or so on*

Example:

(S (NP-SBJ *Tre*)

(*VP (VP giữ làng)*

,/PU

(*VP giữ nước)*

,/PU

(*VP giữ mái nhà tranh)*

,/PU

(*VP giữ đồng lúa chín*))

(./PU))

- Succeeding relation: The expressions are linked by conjunction *rồi*

Example:

Cô (UCP (ADJP im_bắt) rồi/Cp (VP liếc tôi)).

- Optional relation: The words or expressions are linked by comma (,) or parallel conjunctions such as: *hay, hoặc, hay_là*.

Example:

Cá này (VP nướng, chiên hay hấp (ADJP đều ngon)).

- Contrasted relation: The words or expressions are linked by parallel conjunctions such as: *mà and còn*.

Example:

Cô (VP (VP không nói) mà/Cp (VP liếc tôi)).

Caution:

- Many words share a modifier: If the modifier describes characteristics of people or things which the head words mention about, the head words are annotated with flat structure, the modifiers are bracketed with a suitable phrase tag.

Examples:

Tôi thích (NP điện_thoại , laptop (NP màu đen))

Tôi thích (NP (NP điện thoại , laptop (NP màu đen)) , (NP máy tính bảng (NP màu trắng)))

Nó (VP không những la mà còn đánh (NP thằng nhỏ))

(NP Chén, đĩa (NP sứ/Nn)) bán rất chạy {porcelain bowls and plates enjoy large custom}

(NP Chén, đĩa (PP bằng/Cm (NP sứ/Nn))) bán rất chạy. {(NP Bowls and plates (PP making of porcelain)) enjoy large custom}

(NP Chén, đĩa (PP bằng/Cm-H (NP sứ/Nn-H (ADJP trắng/Aa)))) bán rất chạy {(NP Bowls and plates (PP making of (NP white porcelain)))}

- Many expressions sharing a modifier should be annotated as the following example:

(NP

(QP (R hơn) (Num-H 100))

(NP (NP (Nn-H quốc_gia)

*(PP *-1))*

(Cp và)

(NP (Nn-H vùng)

(NP (Nn-H lãnh_thô))

(PP-1 (Cs-H trên)

(NP (Nn-H thế_giới))))))

*trả (NP (NP tiền điện (NP *-1)) và (NP tiền nước (NP-TMP-1 hàng tháng)))*

*Tôi (VP (VP rất thích (NP *-1)) nhưng/Cp (VP sẽ không mua (NP-1 chiếc xe đó)))*

However, the following structure:

(Phrase 1

(Phrase 2 <A>))

(Phrase 3

(Phrase 4))

(C = A + B)

A and B in the data is not written and replaced by C, omitted component is bracketed **SUM** and marked mark reference indices as follow:

(Phase

(Phrase 1

*(Phrase 2 *SUM*-1))*

(Phrase 3

*(Phrase 4 *SUM*-1))*

(C-PRN-1 ...))

Ví dụ:

(VP (R đã) (V_v-H tổ_chức)

(VP-CMP

(VP (V_v-H bàn_giao)

(NP-DOB (Nn-H thư_viện) (Nn sách) (VP *SUM*-2))

(PP-IOB *-1))

(PU ,)

(VP (V_v-H tặng)

(NP-DOB

(NP (Num 450) (Nn-H áo_ấm) (VP *SUM*-2))

(PU ,)
 (NP (Num 450) (Nn-H mũ) (VP *SUM*-2)))
 (PP-IOB-1 (Cs-H cho)
 (NP (Nn-H học_sinh)
 (NP (Nn-H Trường) (Nn tiểu_học))))
 (VP-PRN-1 (Vv-H trị_giá)
 (NP-CMP
 (QP (Num-H 150) (Nq-H triệu)
 (Nu-H đồng))))))

With this structure:

(Phrase 1 <A>)

(Phrase 2)

(C = A+B)

We bracket as follow:

(Phrase

(Phrase 1 *SUM*-1)

(Phrase 2 *SUM*-1)

(C-1))

Example:

(NP

(NP (Num 1) (Nn-H HCD)

(NP (FW-H ASIAD)) (PP *SUM*-1s))

(Cp và)

(NP (Num 1) (Nn-H HCV)

(NP (FW-H SEA_Games)) (PP *SUM*-1))

(PP-1 (Cs-H bằng)

(NP (Num 1) (Nn-H HCV)))

- If many head words share an adjunct, they should be annotated with flat structure.

Examples:

(ADJP (Aa-H linh_thiên) (Cp và) (Aa-H tôn_nghiêm) (R nhất))

(ADJP (R vô_cùng) (Aa-H xinh_đẹp) (Cp và) (Aa-H thông_minh))

(NP (Nq những) (Nn-H tiêu_cực) (PU ,) (Nn-H những_nhiều))

(NP (ADJP (R rất) (Aa nhiều)) (Nn-H tiêu_cực) (PU ,) (Nn-H những_nhiều))

(ADJP (RP (R cũng) (R không)) (Aa-H xinh_đẹp) (Cp và) (Aa-H thông_minh))

(VP (R đang) (Vv-H sinh_sống) (PU ,) (Vv-H làm_việc)

(PP-CMP-LOC (Cs-H tại)

(NP Tp. Hồ Chí Minh)))

- If phrases share one modifier/ word number/ ..., they are grouped.

Examples:

(S

(NP-SBJ vụ cháy)

(VP đã (VP

(VP cướp đi sinh mạng 2 người)

và/Cp

(VP làm bị thương 4 người))))

- If head words share post-modifiers that are lexical word, they are annotated with flat structure.

Examples:

(NP Tiến_sĩ/Nn-H ,/PU bác_sĩ/Nn-H tâm_thần/Nn Nguyễn_Văn _A/Nr)

(NP Giáo_sư/Nn-H Tiến_sĩ/Nn-H Nguyễn_Văn _A/Nr)

- If phrases share 1 post-modifier that are lexical word, the phrases are grouped.

Examples:

(NP (NP (NP Nhà giáo ưu tú) -/PU (NP Giáo_sư/Nn-H) -/PU (NP Tiến_sĩ/Nn-H)) Nguyễn_Văn _A/Nr)

- If words share one category noun/special category noun/ administrative agency (each word combines with category noun/special category noun to call an phenomenon or thing), we group the words/formulations as phrases.

Example:

(NP (Nn_swsp (Nc-H nền) (NP (Nn kinh_tế) (Cp và) (Nn chính_trị))))

(NP (Nn_swsp (Ncs-H việc) (VP (Vv ăn_ở) (Cp và) (Vv đi_lại))))

(NP (Nun-H xã) (NP (Nr-H Vĩnh_Long) (PU ,) (Nr-H Vĩnh_Chấp))))

(NP (Nun-H xã)

(NP (Nr-H Vĩnh_Long) (PU ,) (Nr-H Vĩnh_Chấp))

(LBRK LBRK)

(NP-PRN (Nun-H H.) (Nr Vĩnh_Linh))

(RBRK RBRK))))

If post-modifier includes Sino-Vietnamese formulation that contains Sino-Vietnamese is annotated with flat structure.

(NP (Nn-H lĩnh_vực) (NP (Nn-H chứng_khoán) (Cp và) (Nn_w (Sv bắt)(Nn-H động_sản))))

- When description (for head word) includes words with coordination relation, If they are different in word type, each of these word is annotated properly and the description is marked UCP.

Example:

(NP (Nn-H quá_trình)
 (UCP (ADJP tiêu_cực)
 (PU ,)
 (VP những_nhiều))))

- When description (for head word) includes words with coordination relation, If they are the same in word type, they are grouped as a phrase with flat structure.

Example:

(NP (Nn-H quá_trình)
 (VP (Vv-H hình_thành) (Cp và) (Vv phát_triển))))

- When description (for head word) includes words with coordination relation: words and phrases are annotated properly. The description is marked UCP if it contains phrases with different types. In contrast, it is marked with the same tag as it's child that contains phrases in the same type.

Example:

(NP (Nn-H người)
 (UCP (ADJP (Aa-H dịu_hiền))
 (Cp và)
 (VP (Ve-H có)
 (NP-CMP (Nn-H văn_hóa))))))

(NP Công_ty/Nn
 (VP phát_triển/Vv
 (UCP (VP xây_dựng/Vv)
 và/Cp
 (NP thương_mại/Nn))))
 Á_Âu/Nr)

- In the cases that we are not sure whether the expressions share a modifier expression or not, we treat them as sharing cases.

Examples:

(NP (Nn-H đũa) (Cp và) (Nn-H muỗng)
(NP (Nn kim_loại)))

In above example, we are not sure whether chopstick is make by metal or not.

- If phrase has structure: head + m1 + m2 + Cp + m3 + m4 (where m2 does not modify m1, m4 does not modify m3), we annotate m1, m2, m3 and m4 coordinatively modify head

(VP (Vv-H cung_cấp)
(NP-DOB (Nn-H nước) (Aa sạch))
(PP-IOB (Cs-H cho)
(NP (Nq 85%) (Nn-H dân_số)))
(PP-TMP (Cs-H vào)
(NP (Nt-H năm) (Num 2010)))
(Cp và)
(PP-IOB (Cs-H cho)
(NP (Nq 100%) (Nn-H dân_số)))
(PP-TMP (Cs-H vào)
(NP (Nt-H năm) (Num 2020))))

3.4.1.2.2. Reciprocal coordination

rds or phrases are combined by pairs of conjunctions, such as: *không_những ... mà còn, không_chỉ ... mà, đâu_chỉ ... mà, nói_chung ... nói_riêng, càng ... càng, vừa ... vừa, nửa ... nửa, ngoài_ra ... còn, ngoài ... còn, vừa/mới/chưa ... đã, có ... mới, tuy ... nhưng, etc.*

Structure:

Cp từ Cp/PU từ
Cp ngữ Cp/PU ngữ

Examples:

(ADJP (Cp vừa) (Aa-H bổ) (Cp vừa) (Aa-H rẻ))

(UCP (Cp không_chỉ) (PP để thoát khỏi cái nghèo) (Cp mà_còn) (VP muốn có một cuộc sống đầy đủ hơn))

Reciprocal coordination can not be subject or predicate of clause. It can be object, modifier, etc.

Examples:

Hắn đổ không chỉ nhà cửa mà cả xe cộ vào cờ bạc.

(S (NP-SBJ Hắn)

(VP (Vv-H đổ)

(NP (Cs không_chỉ)

(NP (Nn-H nhà_cửa))

(Cs mà)

(NP cả xe_cộ))

(PP vào cờ bạc))

(PU .))

Một túi (NP vừa/Cp bánh vừa/Cp kẹo) là 20 ngàn.

Caution:

- Distinguish between mutual combinations and adverbial phrase. In mutual combination, relative word often relates two phrases with the same type.

Example:

(NP-ADV Ngoài tiền), (S nó còn đầu tư nhiều thời gian vào việc học)

- Distinguish between mutual combinations and subordinative compound sentence with omitted components.

- The first relative work at the head of the sentence: compound sentence

Tuy nó nghèo nhưng tốt bụng.

(S (Cp Tuy)

(S (NP-SBJ-1 nó) (ADJP-PRD nghèo))

(Cp nhưng)

(S (NP *-1)) (ADJP-PRD tốt bụng))

(PU .))

- Relative word is behind subject of the first clause: sentence with predicate that is mutual combination.

Nó tuy nghèo nhưng tốt bụng.

(S (NP-SBJ nó)

(ADJP-PRD

(Cp tuy)

(ADJP nghèo)

(Cp nhưng)

(ADJP tốt bụng))

(PU .))

3.4.1.3. Sentence or Phrase

Phrase that has subject-predicate structure is as follows:

1) Nn Vv/Aa/VP/ADJP

2) NP VP/ADJP

Rules to distinguish between sentence and phrase:

- I we can insert *mà/là/rằng/sao* between NP and VP, NP VP is not a sentence. (priority 1)

Example:

Hành_trình_gồm (NP_chuỗi_hoạt_động_an_sinh_xã_hội_hướng_đến_kỷ_niệm_60_năm_chiến_thắng_lịch_sử_Điện_Biên_Phủ)

- If we can not insert sunplement word to change tense of the verb, NP VP is not a sentence.
- If we can insert *mà/là/rằng/sao* before NP, NP VP is a sentence.
- If there is pronoun *này/đó/...* , which modify noun at the head of, at the end, this is phrase.
- If there is pronoun *này/đó/...* at the end of NP, NP VP is sentence.

3.4.2. Noun phrase (NP)

An expression of noun or pronoun (head word) and its modifiers will be annotated with NP

Ex:

(NP_Tất_cả_những_cái_con_mèo_đen_ấy) {All_these_black_cats}

There are 2 forms of NP

3.4.2.1. Form 1

A noun phrase includes a maximum of 3 parts as follows:

Pre-modifier			Head	Post-modifier		
-3	-2	-1	0	1	2	3
whole	quantifier	emphasis	noun or pronoun	categorization	description	demonstrative

3.4.2.1.1. Pre-modifier

NP can contain pre-modifier or not. If any, it includes a maximum of 3 components in order from left to right (-3 to -1) as follows:

Component -3

This component includes a quantifier indicating total (Nw), such as: *tất cả, hết thấy, tất thấy, toàn bộ, cả*. It will be annotated the same level with head words.

Examples:

(NP tất_cả/Nw ba cái con mèo đen ấy) {all these 3 black cats}

(NP toàn_bộ/Nw ba cái con mèo đen ấy) {all these 3 black cats}

(NP cả/Nw ba cái con mèo đen ấy) {all these 3 black cats}

(NP cả/Nw lớp/Nn-H) {all students in a class}

(NP tất_cả/Nw học_sinh) {all students}

The foreign word *top* can also stand at this position.

Example:

(NP Top/FW 10/Num nước/Nn-H)

Component -2:

This component modify head word about quantity. It can be a number (Num), a quantifier that cannot be written by number (Nq), or a quantifier phrase (QP).

Examples:

- Num: *một, ba, bốn, etc.*

*(NP tất_cả **ba/Num** cái con mèo đen ấy) {all these 3 black cats}*

- Nq: *các, những, vài, mọi, từng, etc.*

(NP tất_cả mọi/Nq con mèo đen) {all black cats}

(NP từng/Nq người/Nn-H) phụ trách (NP từng/Nq việc) {each member is in charge of a work}

- QP

(NP tất_cả (QP 4/Num triệu/Nq) người) {all of 4 million people}

(NP (QP (R hơn) (Num-H 4) (Num-H triệu)) (Nn-H người)) {more than 4 million people}

(NP (QP (Num-H 20) (Nq phần_trăm)) (Nn-H sinh_viên)) {20% students}

(NP (Nq-H 20%) (Nn-H sinh_viên))

Component -1:

This component is a modifier (M) to emphasize the head word. It is annotated the same level with head word.

Examples:

(NP tất_cả ba cái/M con/Nc-H mèo đen ấy) {all these 3 black cats}

(NP người/M con/Nn-H trai) {son}

(NP cô/M con/Nn-H dâu) {daughter-in-law}

3.4.2.1.2. Head word (component 0)

Head word of NP can be noun (Nc, Ncs, Nn, Nt, Nr, or NA) or pronoun (Pp or Pd).

- Nc (classifier noun) is always head of NP

Examples:

(NP tất_cả ba cái con/Nc-H mèo đen ấy)

(NP con/Nc-H mèo) {cat}

(NP cái/Nc-H ấy) {that one}

- Ncs is modified by a noun, a verb or an adjective to identify/name a person or thing. This sequence will be bracketed with an internal structure tag, Nn_swsp.

Examples:

(NP một (Nn-swsp nhà/Ncs-H khoa_học/Nn) trẻ) {a young scientist}

(NP (Nn-swsp nhà/Ncs-H báo/Nn) lớn tuổi nhất) {the best senior journalist}

(NP (Nn-swsp nhà/Ncs-H nghiên_cứu/Vv)) {researcher}

(NP những (Nn-swsp nhà/Ncs-H nghiên_cứu/Vv) về tin học) {informatic researchers}

(NP (Nn-swsp cái/Ncs-H nhìn/Vv) hút hồn) {the fascinated look}

(NP (Nn-swsp sự/Ncs-H hy_sinh/Vv) anh dũng) {the heroic sacrifice}

(Nn-swsp nền/Ncs-H hoà_bình/Aa) {peace}

(Nn-swsp cái/Ncs-H đẹp/Aa) {beauty}

- Nn will be the head word of NP if it is not a modifier of Nc or Ncs (do not follow a Nc or Ncs).

Example:

(NP những sinh_viên/Nn-H giỏi) {the good students}

- Nu

Examples:

(NP 9/Num độ/Nu-H vĩ_bắc/Nn) {30 degree north latitude}

(NP 9/Num độ/Nu-H C/Nu) {9 degree Celcius}

(NP km/Nu-H 20/An) {20th km}

(NP km/Nu-H (NP số/Nn 20/An)) {20th km}

(NP

(NP (QP 2,2 tỷ) đồng/Nu-H)

(PU /)

(NP (Nn-swsp căn/Nc-H nhà/Nn)))

(NP

(NP 1 lít/Nu-H (NP sữa))

(PU /)

(NP ngày))

- The expressions of Nun and its modifiers will be annotated with NP. Where, Nun is head.

Examples:

(NP hai thành_phố/Nun-H của Việt Nam) {2 cities of Vietnam}

(NP Tp./Nun-H Hồ_Chí_Minh) {Ho Chi Minh city}

- Nr is head of NP if it does not have pre-modifier.

Examples:

(NP Lan/Nr-H) đi học {Lan go to school}

(NP Lan/Nr (PP của mẹ)) rất ngoan {my daughter (Lan) is very obedient}

(NP Thuỳ_Linh/Nr-H kế_toán) rất giỏi {Thuy Linh which is accountant is very good.}

- Nt

Examples:

Chúng ta sẽ gặp (NP lúc/Nt-H 8 giờ sang).

- NA

Examples:

(NP Viện/Nn (VP Nghiên_cứu/Vv quản_lý/Vv (NP kinh tế)) (NP Trung_ương/NA-H))

- Pp

Examples:

(NP Chúng_tôi/Pp-H) là sinh_viên

(NP Tôi/Pp-H) (VP không/R cần/Vv (NP gì/Pp-H))

Tôi rất quý (NP những/Nq gì/Pp-H (S mẹ cho tôi))

không có (NP một/Num ai/Pp-H (PP ở nhà))

(NP những/Nq ai/Pp-H) làm việc xấu thường bị quả báo

(NP tất_cả/Nq chúng_tôi/Pp-H) là sinh_viên

(NP bao_nhiêu/Pp-H tiền/Nn) là đủ

(NP thế/Pp-H này/Pd) là sao?

Caution:

- o If Nn precede Pp in NP, Nn is head of NP

Example:

(NP hai vợ_chồng/Nn-H (NP chúng_tôi/Pp-H))

- “Ai, gì” belonging to NP is not have questionable meaning. They are used to indicate unknown person/thing. Please refer to QNP for annotation of “ai, gì” with questionable meaning.
- If Pd is a component of sentence (it is usually subject), it will be annotated as NP.

Ex:

(NP Đây/Pd-H) là ông nội tôi. {This is my paternal grandfather.}

However, if Pd follows Pp or Nn in NP, it will be treated as modifier.

Ex:

Làm (NP thế/Pp-H này/Pd) cũng được. {This way is OK.}

(NP anh/Nn-H ấy/Pd) là bạn tôi {He is my friend}

- FW is head of NP if is not follow Nc.

Ex:

(NP virus/FW-H này) {this virus}

(NP album/FW-H November)

(NP virus/FW-H HIV)

3.4.2.1.3. Post-modifier

NP can have post-modifier or not. Post-modifier includes a maximum of 3 components (if any).

Component 1

This component denotes categorization in order to identify/name a person/thing. If it is a word, it is annotated the same level with head word. If it has more than one words, it will be bracketed with suitable constituent tag.

- Noun

Examples:

(NP quyển/Nc-H sách/Nn)

(NP quyển/Nc-H (Nn-swsp sách/Nn toán/Nn))

(NP (Nn-swsp nhà/Ncs-H khoa_học/Nn))

(NP cá/Nn-H rô/Nn)

(NP học_sinh/Nn-H THPT/Nn)

(NP cô/Nn-H Lan/Nr)

(NP thành_phố/Nun-H HCM/Nr)

(NP sự_kiến/Nn-H 11-9/Nt)

(NP độ/Nu-H vĩ_bắc/Nn)

- Pronoun (Pp)

Examples:

Tôi (VP ăn (NP cái/Nc-H gì/Pp) (RP cũng/R được/R-H)) {I can eat everything}

Tôi (VP ăn (NP cái/Nc-H gì/Pp) (ADJP cũng/Ra được/Aa-H)) {I can eat everything well}

Tôi (VP ăn (NP món/Nc-H nào/Pp) (RP cũng/R được/Rv-H))

- Aa

Ex:

(NP sự/Ncs-H anh_dũng/Aa)

(NP màu/Nn-H xanh/Aa)

- An

Ex:

(NP lần/Nn-H đầu_tiên/An)

(NP lần/Nn-H (NP thứ/Nn-H nhất/An))

- Foreign word (FW)

Ex:

(NP quán/Nn-H bar/FW)

(NP vi_rút/Nn-H HIV/ FW)

- Idiom (ID)

Ex:

(NP đoạn/Nn-H tối_như_bưng/ID)

- NP

Ex:

(NP ghé'/Nn-H (NP hạng/Nn-H nhất/An))

(NP sự_kiện/Nn-H (NP ngày/Nt 11-9/Nt))

- Quantifier phrase (QP)

Ex:

(NP lần/Nn-H (NP thứ/Nn-H (QP một/An trăm/An lẻ/An bảy/An))) {the 107th time}

(NP lần/Nn-H (NP thứ/Nn-H 107/An)) {the 107th time}

Component 2

Word or phrase shows characteristics of the person or thing denoted by the combination of head word and the modifier in position 1. This component is bracketed with suitable constituent tags.

Examples:

- NP

Cô ấy là (NP Á_hậu/Nn-H 1/An (NP Miss_world/FW-H 2008/Nt))

(NP cái/Nc-H bàn/Nn (NP thứ/Nn-H tư/An-H))

(NP vụ/Nn-H tai_nạn/Nn (NP hôm_qua/Nt-H))

- QP

(NP bàn/Nn-H (QP 14/An-H))

(NP bàn/Nn-H (QP mười/An-H bốn/An-H))

- VP

(NP (Nc-H người) (Nn thợ) (VP (Vv-H sửa) (NP xe)))

(NP (Ncs-H nhà) (Vv nghiên_cứu) (VP (Vv-H có) (NP (Nn chí))))

(NP (cảnh_sát/Nn-H) (VP (VP phòng/Vv-H (VP cháy/Vv-H)) (VP chữa/Vv-H (VP cháy/Vv-H)))) (NP công_an/Nn-H (NP tỉnh/Nun-H Cà_Mau/Nr)))

(NP bà/Nn-H (VP bán/Vv-H (NP bánh_mì/Nn-H)))

- ADJP

(NP con/Nc-H đường/Nn (ADJP tối_như_bưng/ID))

(NP (Nn-H cô) (Nn gái) (ADJP (Aa-H đẹp)))

(NP (Nn-H cô) (Nn gái) (ADJP (ADJP (R vô_cùng) (Aa-H xinh_đẹp)) (Cp và) (ADJP (Aa-H thông_minh))))

(NP (Nc-H chiếc) (Nn xe) (ADJP (An-H đầu_tiên))) // An thi gan ADJP

- PP

(NP (Nn-H phòng) (PP (Cm-H của) (NP (Pp-H tôi))))

(S (NP-SBJ (Nn-H Xe_đạp) (Nn địa_hình) (PP (Cm-H cho) (NP (Nn-H trẻ_em)))) (VP bán_rất_chạy) (PU .))

(NP (Ncs-H vụ) (Aa-H tiêu_cực) (PP (Cm-H ở) (NP (Nn-H công_trình))))

- Combine

(NP Con/Nc-H mèo/Nn (ADJP (ADJP đen/Aa) (ADJP đáng_yêu)) (PP của_tôi))

- Subordinate clause

(NP Con/Nc-H mèo/Nn (SBAR mà/Cs (S tôi mua)))

(NP Con/Nc-H mèo/Nn (SBAR *0*/Cs (S tôi mua)))

Component 3

It includes a demonstrative pronoun (Pd), such as *ấy, này, nọ, kia, đó, đấy, etc.* to confirm/determine the position/role of thing or person in space or time. It is annotated the same depth with the head word.

Ex:

(NP Con/Nc-H mèo/Nn (ADJP trắng) này/Pd)

(NP Con/Nc-H mèo/Nn (PP của Nam) này/Pd)

Caution:

- The inversion between (3) and (2) can change NP to clause.

Ex:

(S (NP-SBJ Con/Nc-H mèo/Nn này/Pd) (ADJP-PRD trắng))

(S (NP-SBJ Con/Nc-H mèo/Nn trắng/Aa này/Pd) (PP-PRD của Nam) ./PU)

- Distinguishing between NP and ADJP having the same construction, adjective stands before noun.

If the phrase expresses the quantity, it is bracketed with NP. In this case, we can replace the adjective by a quantifier.

Ex:

Anh ấy (VP có (NP (ADJP nhiều) kinh_nghịệm)).

(S

(NP-SBJ

(ADJP (R chỉ) (Aa-H duy_nhất))

(PU ")

(Nn-H của)

(PU ")

(PP (Cs-H ở)

(NP (Nn-H cấp) (Nun tỉnh))))

(VP (Vv-H được)

(VP-CMP (Vv-H lập) (PU ,) (Vv-H chỉnh_lý) (Cp và) (Vv-H quản_lý)

(NP-DOB (Nn-H hồ_sơ) (Nn địa_chính) (Nn gốc))))

Conversely, if the phrase expresses quality, the phrase is bracketed with ADJP label.

Ex:

Anh ấy là (NP người/Nn-H (ADJP nhiều/Aa-H (NP kinh_nghịệm)).

(S (NP-SBJ Cô ấy) (ADJP-PRD đẹp/Aa-H (NP người)) (PU .))

- Distinguishing S and NP having the same structure, noun stands before adjective.

If we can insert the adjunct indicating tense as a pre-modifier of the adjective, the expression should be bracketed with S label.

Examples:

(S (NP-SBJ Cây này) (S-PRD (NP-SBJ lá) (ADJP-PRD vàng)) (PU .))

(S (NP-SBJ Cây này) (S-PRD (NP-SBJ lá) (ADJP-PRD đã vàng)) (PU .))

In contrast, the expression will be treated as a NP.

Ex:

Cây này có (NP lá (ADJP vàng)).

In some cases, we can base on conjunction *thì* to distinguish sentence and phrase. If an expression is a sentence, we can use the conjunction *thì* to link subject and predicate (*thì* plays the same role as *to be* in English). Otherwise, the expression is a phrase.

- Distinguishing S and NP having the same structure, noun stands before verb.

If the noun is not subject of the action stated by the verb, the expression is bracketed with a NP label.

Ex:

(S (NP-SBJ Chim) (VP hát) (PU .))

(NP nguyên_tắc/Nn-H hoạt_động/Vv)

3.4.2.2. Form 2

Construction:

Pre-modifier		Head	Post-modifier		
-2	-1	0	1	2	3
adjunct	modifier	noun or pronoun	categorization	description	demonstrative

This form of NP has structure of pre-modifier which is different from previous form. In this NP type, pre-modifier (if any) includes a maximum of 2 component. Where, adjunct/adjunct phrase come before head word. They are annotated the same level with head word.

Ex:

(NP cũng/R cái/M con mèo đen này) {the same black cat}

(NP cũng/R cô gái xinh đẹp ấy) {the same beautiful girl}

(NP vẫn/R cô gái xinh đẹp ấy) {the same beautiful girl}

(NP (RP cũng vẫn) cô gái xinh đẹp ấy) {the same beautiful girl}

(NP bất_kẻ/R ai) {anyone}

Note:

- Adjunct precedes a number. However , it does not modify the number. It modifies the head word of NP.

(S

(NP-SBJ (Pd-H Bây_giờ))

(NP-PRD (R đã) (Num 7) (Nt-H giờ) (Nt tối)))

3.4.2.3. Time phrase

A multi-words expression of time point (precise or imprecise), time period, or length of time will be bracketed with NP label.

Examples:

- Time point:

Precise: *Tôi sẽ đến đó (NP lúc 3 giờ 30 chiều) {I will come there (NP at 3:30 pm)}*

Imprecise: *(NP cái ngày/Nt-H tôi cưới vợ) {the day that I marry}*

- Time period:

Tôi đã làm việc ở đây (PP từ tháng 2 đến tháng 8 năm 2012) {I worked here (NP from February to August 2012)}

- Length of time:

Tôi làm việc này mất (NP 3/Num ngày/Nu-H).

3.4.2.3.1. Time point

Construction:

Pre-modifier + head + post-modifier

a. Pre-modifier

Can be Num, QP, Nn, Cs, M, Aa, Nq, or Pp depending on the context.

Examples:

- Num

(NP (Num 12) (Nt-H giờ) (NP (Nt-H trưa))) {12 pm}

- QP

(NP (QP (Num mười) (Num hai)) (Nt-H giờ) (NP (Nt trưa))) {12 pm}

- Nn

(NP (Nn cuối) (Nt-H tháng) (Num 12) (Pd này))

(NP (Nn trước) (Nt-H Tết))

- M

*(NP cái/M ngày/Nt-H (SBAR (Cp *0*) (S tôi cưới vợ)))*

- Nq

(NP những/Nq ngày/Nt-H này/Pd)

- Aa

(NP (Aa gần) (Nt-H sáng))

b. Head

Head word is always Nt.

Examples:

(NP (Nt-H Lúc) (Nt 10g))

(NP (Nt-H Tối) (Nt nay))

(NP (Nt-H mùa) (Nt đông) (NP (Nt-H năm_ngoái)))

(NP (Nt-H ngày) (Nt giáng_sinh))

(NP Ngày/Nt-H 2/Num và/Cp 3/10/Nt)

(NP 4/11/2013/Nt)

(NP ngày/Nt-H 30/Num

(NP tháng/Nt-H 4/Num

(NP năm/Nt-H 2013/Num)))

(NP (Nt-H ngày) (FW Noel))

(NP (Nt-H ngày) (FW Valentine))

c. Post-modifier

It can be a word, more than one word, or phrases

Examples:

- Num

(PP vào/Cm (NP năm/Nt-H 2003/Num))

- QP

(PP vào/Cm (NP (Nt-H năm) (QP (Num hai) (Nq ngàn) (Num không) (Nq trăm) (Aa linh) (Num ba))))) {in 2003}

- Nt

(NP (Nt-H mùa) (Nt xuân)) {Spring}

- More than one word

(NP (Nn cuối) (Nt-H tháng) (Num 12) (Pd này)) {later this month}

(NP (Nt-H mùa) (Nt xuân) (Pd này)) {this Spring}

- **Complex structure**

(NP lúc/Nt-H

(NP (Num 3) (Nt-H giờ)

(NP (Num 30) (Nt-H phút)))))

(NP lúc/Nt-H

(NP 1/Num giờ/Nt-H

(NP 30/Num-H)

(NP sáng/Nt-H)

(NP ngày/Nt-H 11/9/2013/Nt))))

(NP (Nt-H mùa) (Nt xuân) (NP năm_ngoài/Nt-H)) {last Spring}

*(NP cái/M ngày/Nt-H (SBAR (Cs *0*)(S tôi đi cưới vợ))))*

3.4.2.3.2. Time period

Construction:

(PP (PP từ/Cs word) (PP đến/Cs word))

(PP (PP từ/Cs phare) (PP đến/Cs phrase))

Examples:

(PP (PP từ/Cs (NP thứ hai)) (PP đến/Cs (NP thứ sáu)))) {from Monday to Friday}

(PP (PP từ/Cs (NP ngày 20/12)) (PP đến/Cs (NP 26/12)))) {from 20/12 to 26/12}

(PP (PP từ/Cs (NP tháng 2)) (PP đến/Cs (NP tháng 8 năm 2012))) {from February to August 2012}

If *đến* is replaced by “–”, we annotate as follows:

(PP (Cs-H từ)

(NP (Num-H 19))

(PU -)

(NP (Num 22) (Nt-H giờ)

(NP (Nt-H ngày) (Nt 29.11))))

3.4.2.3.3. Length of time

Ex:

Tôi mất (NP (QP hơn 2) giờ/Nu-H) để làm việc này.

mất (NP 1 phút/Nu-H (NP 30 giây/Nu-H))

3.4.2.4. Measure phrase

Expressions include the unit noun.

Examples:

(NP (NP 2/Num lít/Nu-H) sữa/Nn-H)

(NP (Num 400) (Nu-H lít/ngày))

Construction

Pre-modifier + head + post-modifier

3.4.2.4.1. Pre-modifier

- Quantifier

Example:

Tôi làm việc này hết (NP ba/Num tiếng/Nu-H)

- Quantifier phrase (QP)

Examples:

(NP (QP (Num-H ba) (Num-H bốn)) (Nu-H tháng)) {three or four months}

(NP (QP (Aa âm) (Num-H 3)) (Nu-H độ)) {minus 3 degrees}

(NP (QP gần 7 tỉ) (Nu-H đồng)) {nearly 7 billion dong}

(NP (QP 2 ngàn) (Nu-H đồng/bó))

(NP (QP từ 6 đến 9) (Nu-H tháng))

(NP (QP từ 6-9) (Nu-H tháng))

(NP (QP 200 triệu đến hàng tỉ) (Nu-H đồng))

3.4.2.4.2. Head of measure phrase

Always is unit noun (Nu)

Examples:

(NP (Num-H 2) (Nu-H đô_la))

(NP (Num-H 10) (Nu-H giờ)) {10 hours}

(NP (Nq 24/24) (Nu-H giờ))

(NP (Num 100) (Nu-H Celsius))

(NP (QP 4,8 triệu) (Nu-H tấn/năm)) {4.8 million tones/year}

(NP (NP (Num 50.000) (Nu-H đồng)) (PU /) (NP (Num 30) (Nu-H ngày)))

(NP (Num 2) (Nu-H centimet))

(NP (Nq Mỗi) (Nu-H lít))

3.4.2.4.3. Post-modifier

- Noun (Nr, Nn, Nu, or Nt)

Examples:

(NP (Num 2) (Nu-H USD) (Nr Mỹ))

(NP (Num 10) (Nu-H h) (Nn đồng_hồ)) {10 hours}

(NP (Num 3) (Nu-H ngày) (Nn sau))

(NP (Num một) (Nu-H buổi) (Nt sáng)) {a morning}

(NP (Num -3) (Nu-H độ) (Nu C))

- PP

Example:

(NP (Num 2) (Nu-H ngày) (PP trước/Cs (NP Tết/Nt-H)))

- Modal (M)

Example:

(NP 10/Num ngày/Nu-H nữa/M)

Caution:

- Example 1 and 2 have the same meaning. However, they are bracketed different structures because example 1 can be written in standard measure units, but example 2 is not.

Example 1:

(NP (NP (Num 400) (Nu-H lít) (NP (Nn-H sữa))) (NP mỗi/Nq ngày/Nu-H))

(NP (Num 400) (Nu-H lít_sữa/ngày)) {400 liter of milk per day}

Example 2:

(NP Mỗi/Nq ngày/Nu-H) (NP 400 lít sữa) {400 liter per day}

(NP Mỗi ngày), (NP tôi) (VP uống (NP 400 lít sữa))

- Distinguish between NP and ADJP: based on the rule: “word that cannot remove is head”. Or we can base on meaning of the phrase, adjective phrase express quality, meanwhile, noun phrase indicate things, objects, quantity.

- o NP

Tôi làm việc này (NP (ADJP (Aa suốt)) (Num 9) (Nu-H tháng))

- o ADJP

Bức tường này (ADJP (Aa-H cao) (NP (Num 2) (Nu-H m)))

3.4.3. Quantifier phrase (QP)

QP is used to group the multi-word numerical expressions that are pre-modifier or post-modifier of NPs. For single quantifier, we tag it the same level with head noun.

Examples:

tôi có (NP 2/Num con mèo)

tôi được (NP (QP trên 5) điểm toán)

Anh ấy ăn (NP (QP những/M sáu) bát cơm)

Anh có thể chọn (NP (QP 2 trong 3) môn) để học

Construction is following:

pre-modifier + head + post-modifier

3.4.3.1. Pre-modifier

- Inexact quantifier (Nq), such as: *mấy, vài, một_vài, vài_ba, đôi_ba, năm_ba, năm_bảy, dăm, dăm_bảy, dăm_mười, mười, mười_lăm, or so on*. It modifies quantifiers that can not write as numbers and makes meaning of QP become inexact.

Examples:

(NP (QP vài/Nq chục/Nq-H) người)

(NP(QP dăm_ba/Nq trăm/Nq-H triệu/Nq-H) tiền mặt)

- Quantifiers: *mỗi, từng*

Examples:

(NP (QP từng/Nq tá/Nq-H))

(NP (QP mỗi/Nq cặp/Nq-H) vợ_chồng))

- Estimative adjunct (R), such as: *trên, dưới, khoảng, chừng, độ, khoảng_chừng, cỡ* precedes Num

Example:

(NP (QP trên/R 10/Num-H) sinh_viên)

- Adjunct (R): *hàng; hơn; gần; xấp_xỉ; ít_nhất, đúng, suốt*

Example:

Nó tiếp (QP hàng/R trăm/Nq-H ngàn/Nq-H) người mỗi ngày

- Adjective (Aa): âm, dương, thiếu, dư {minus, add}, tương_đương

Example:

(QP tương_đương 10) {as many as 10}

- Preposition (Cs): từ

Example:

(QP từ/Cs 2/Num-H -/PU 3/Num-H) người

- Modifier (M)

Example:

(QP đến/M 8%/Nq-H)

3.4.3.2. Head

All types of quantifiers can be head word of QP.

QP can have one head word or more than one head word.

Examples:

(NP (QP khoảng/R 2/Num-H) chén cơm)

(NP (QP hai/Num-H mười/Nq-H lăm/Nq-H ngàn/Nq-H) người)

(NP (QP khoảng/R năm/Num-H chục/Nq-H ngàn/Nq-H) người)

3.4.3.3. Post-modifier

Can be adjunct

Ex:

Nó được bao nhiêu điểm? Chắc (QP bảy/Num-H hơn/R)

Exception

Ex:

Nó được bao nhiêu điểm? Chắc (NP bảy/Num điểm/Nn-H hơn/R)

Caution:

- từ ... đến ... (do not treat Cs as head of a PP)
(NP (QP từ/Cs 2/Num-H đến/Cs 3/Num-H) người)
(NP (QP từ/Cs 2/Num-H -/PU 3/Num-H) người)
(NP (QP từ/Cs (QP 2/Num-H triệu/Nq-H) đến/Cs (QP 3/Num-H triệu/Nq-H)) người)
(NP (QP từ/Cm (QP 2/Num-H triệu/Nq-H) -/PU (QP 3/Num-H triệu/Nq-H)) người)
(NP (QP 1/Num-H đến/Cs (QP 15 triệu)) con khi)
(NP (QP khoảng/R 1/Num-H đến/Cs (QP 15 triệu)) con khi)
(NP (QP khoảng/R (QP từ/Cs 1/Num-H đến/Cs (QP 15 triệu))) con khi)
- giữa ... và ...
(QP giữa 12 và (QP 13 phần_trăm))
(QP giữa 4 và (QP 5 trăm ngàn)) người
- Num trong ...
(NP 1/Num-H (PP trong/Cs-H (NP những/Nq sinh_viên/Nn-H)))
- In case where the head noun is missing (understood from context):
 - Single-word number without modifier is tagged as QP
lãi suất (VP tăng (QP 0,7/Num))
 - Single-word number with modifier is tagged as QP
Nó được bao nhiêu điểm? Nó được (QP khoảng 7)
 - Multi-word number is tagged as QP
Cô mua bao nhiêu trứng? Tôi mua (QP một chục).
Nó ăn (QP hơn một nửa)

3.4.4. Verb phrase (VP)

An expression of verb (including Vv, Ve, Vc, Vcp, and ID which means as verb) and its modifiers will be bracketed with VP label.

Ex:

Nó (VP đang đi rất nhanh về phía cô ta). {He is going fast toward her.}

Construction

Each VP includes 3 parts: pre-modifier + head verb + post-modifier.

Structure of each parts as follows:

3.4.4.1. Pre-modifier

Pre-modifier can be adjunct (R, adjunct phrase (RP), adjective, VP, etc.

- Adjunct and RP are used to denote the tenses (*đã, đang, sẽ, luôn, luôn_ luôn, etc.*), degree/level (*rất, quá, hơi, etc.*), negation (*không, vẫn chưa, etc.*), imperative (*đừng, chớ, hãy*), etc.

Examples:

Ngày mai tôi (VP sẽ/R về/Vv-H). {I (VP will come back) tomorrow.}

Anh ấy (VP rất thương mẹ). {He (VP loves his mother so much).}

Hôm nay, tôi (VP không đi học). {I (VP do not go to school) today.}

Chúng ta (VP hãy hoàn_ thành việc này hôm nay). {Let's finish this work today.}

(VP (VP Ăn trưa đi) rồi/Cp (VP hãy làm tiếp)).

Anh ấy (VP (RP vẫn/R còn/R) ở đây). {he is still here}

- Adjective indicates manner, state, onomatopoeia, pictographic

Example:

Họ (VP (ADJP ráo_riết/Aa) tìm_kiểm/Vv cô ta). {They find her hectically}

- Pre-modifier can be idiom (ID) having meaning as adjective.

Ex:

Nó (VP (ADJP nhanh_như_cắt/ID) lao ra sân) {he rushes out the yard with lightning speed}

- Pre-modifier can include VP which expresses an action occur before the action stated by head word.

Ex:

Tôi (VP mới/R (VP đi chợ) về). {I just come back from the market.}

Tôi (VP (RP vừa mới) (VP đưa con đi học) về).

Caution:

This type of VP is ambiguous with the VP which following verb modify preceding verb.

Ex:

1. *Mẹ (VP đi chợ về). {The mother comes back home from the market.}*

2. *Mẹ (VP đưa con về). {The mother brings a child back home.}*

Base on POS tag and context to distinguish 2 these types of VP. In the first example, both verb, *đi* and *về* are tagged as Vv, the action *về* {come back} occurs after the action *đi* {go}. Therefore, we should annotate this VP type as follows:

Mẹ (VP (VP đi/Vv-H chợ/Nn) về/Vv-H).

In the second one, *về* {back} is a verb modifying *đưa* {bring} about direction (*về* is annotated with POS tag D). So, we should treat *về* as a post-modifier of the VP.

Mẹ (VP đưa/Vv-H (NP con/Nn) về/D).

- Pre-modifier can include PP to modify head verb about direction.

Ex:

Tôi (VP (PP từ quê) lên thành phố học). {I am from countryside go to the city to study}

- Pre-modifier can include words and phrases

Ex:

Tôi (VP vừa/R (PP từ quê) lên). {I just come back from my hometown.}

- Pre-modifier can include NP

Ex:

Nó (VP (NP tự/Pp) vá áo)

3.4.4.2. Head word

All verb types can be head word of the VP, except D. Double-POS words (VA, VN) are also head of VP.

- Ve is used to express the existing of things.

Ex:

(VP Có/Ve-H 2 cuốn sách trên bàn) {There are 2 book on the table}

(VP đã hết/Ve-H phim) {the film ended}

Tôi (VP còn/Ve-H tiền) {I have money left}

- Copula “là” verb (Vc) to express the equivalent between two things or people

Ex:

Tôi (VP là/Vc-H sinh viên). {I am a student}

Hà Nội (VP là/Vc-H thủ đô của Việt Nam).

- Comparative verb (Vcp): *thua, hơn, kém, bằng, ngang, or tương đương* to express the comparison between objects.

Ex:

Tôi cao (VP hơn/Vcp-H anh). {I am taller than you.}

Tôi (VP kém/Vcp-H chị 2 tuổi). {I am younger}

- Other verb types annotated with POS tag Vv are also head of VPs, such as:
 - o Verbs indicating mentality (Vv): *tin, nghi ngờ, yêu, ghét, buồn, chán, nhớ, hiểu, etc.* can be modified by adjuncts indicating degree/level.

Ex:

Tôi (VP rất/R ghét/Vv-H bạn). {I hate you so much.}

Tôi (VP rất buồn ngủ).

Caution:

The verb phrase that its meaning is similar to adjective also has pre-modifier which is adjunct indicating degree.

Ex:

Anh ấy sống (VP rất có/Vv-H tình/Nn).

Anh ấy (VP rất biết/Vv-H (VP nghe lời bố mẹ)).

- Verbs indicating feeling, such as: *cảm_thấy* {feel}, *nhận_thấy* {realize}, *nghe*, and *ect*.

Ex:

Tôi (VP cảm_thấy rất buồn).

Caution:

Constructions of *cho* + *verb*, such as *cho hay* {notify}, *cho biết* {tell}, *cho thấy* {show}, *cho vay*, ... are not the words. The expressions including them are annotated as follows:

Họ cho biết kết quả xét nghiệm.

(S (NP-SBJ Họ)

(VP cho/Vv-H

*(NP-DOB *E*)*

(VP-IOB biết

(NP-CMP kết quả xét nghiệm)))

(PU .))

Họ cho chúng tôi biết kết quả xét nghiệm.

(S (NP-SBJ Họ)

(VP cho/Vv-H

(NP-DOB chúng_tôi)

(VP-IOB biết

(NP-CMP kết quả xét nghiệm)))

(PU .))

Họ cho biết kết quả xét nghiệm rất khả quan.

(S (NP-SBJ Họ)

(VP cho/Vv-H

(NP-DOB *E*)

(VP-IOB biết/Vv-H

(SBAR-CMP (Cs *0*) (S kết quả xét nghiệm rất
khả quan))))

(PU .))

Note:

Cho and *là* in the following structure are prepositions which are different from about structures.

Phong cho nó là F4 vậy

(VP Phong

(PP cho/Cs

(NP nó))

(PP là/Cs F4 vậy))

- Transformative verbs such as: *trở_nên*, *nên* {become}, *trở_thành*, *hoá, thành, etc.*

Ex:

(VP (VP *Nên/Vv* vợ) (VP *nên/Vv* chồng)) {become husband and wife}

(VP *Nói* (VP *không nên/Vv* lời)) {speechless}

Cha mẹ đã nuôi dạy con (VP nên/Vv-H người) {Parent educate me to become a good person}

- Verbs indicate necessary, endurance, and modality verb (Vv): *cần, cần_thiết, nhất_thiết, nên, phải, chịu, mắc, etc.*

Ex:

Tôi (VP nên/Vv (VP hoàn_thành/Vv việc này sớm)). {I should finish this work early.}

Bạn (VP không nên/Vv-H (VP hút/Vv-H thuốc lá))

Tôi (VP cần/Vv tiền). {I need money}

Tôi (VP cần/Vv-H (VP xong/Vv-H việc này sớm)). {I need to finish this work early.}

Tôi (VP phải Vv-H (VP xong/Vv-H việc này sớm)). {I must finish this work early.}

Tôi (VP cần/Vv-H (VP phải/Vv-H (VP xong/Vv-H việc này sớm))) {I need to finish this work early.}

Tôi đến đây (VP phải ngày mưa gió). {I come here meet with the bad weather.}

Nó giãy nảy lên như đĩa (VP phải/Vv-H vôi). {He jump up as the leech catch lime.}

- The passive verb: *bị* and *được*

Ex:

Nó được khen.

(S (NP-SBJ-1 Nó)

(VP được/Vv-H

(S (NP-SBJ *E*) (VP khen/Vv-H (NP-DOB *T*-1)))

(PU .))

- Volition verbs: *muốn, mong_muốn, quyết, dám, toan, định, ước, etc.*

Ex:

Tôi (VP muốn (VP-DOB hoàn thành (NP-DOB việc này) (ADJP-CMP sớm))). {I want to finish this work early.}

- Verbs indicating physical activity: *đi, nhảy, chạy, trèo, bước, đẩy, đập, đọc, viết, thực_hiện, hy_sinh, etc.*

Ex:

Mẹ (VP đã đi/Vv-H chợ).

- Process verbs indicating initiation, continuation and end: *bắt_đầu, tiếp_tục, hết, thôi, etc.*

Ex:

Chúng ta (VP hãy tiếp_tục công việc).

- *Xong* and *được* indicating the result or ability can be head word or post-modifier.

Ex:

Nó (VP được/Vv-H giải thưởng). {get award}

Anh ta (VP nói/Vv-H được/R 5 thứ tiếng). {can speak}

Tôi (VP đã xong/Vv-H việc) {finish}

Tôi (VP đã đọc/Vv-H (VP xong/Vv-H cuốn sách này)).{ready}

- Verb-noun

Ex:

Mọi chi tiết của vụ mưu sát (VP (RP vẫn còn) nghi_vấn/VN-H).

- Verb-adjective

Ex:

Sao em (VP (RP vẫn còn) buồn/VA-H).

3.4.4.3. Post-modifier

Post-modifier annotated as word

- Modifier (M): be annotated the same level with head verb.

Ex:

(VP chờ/Vv-H (NP tôi/Pp) với/M)

- Onomatopoeia (ON): be annotated the same level with head verb.

Ex:

(VP cười/Vv-H hi_hi/ON)

- Adjunct (R): be annotated the same level with head verb.

Ex:

Tôi (VP đã nhận/Vv-H ra/R (NP ý nghĩa của cuộc sống)) {I figured out the meaning of life}

Post-modifier annotated as phrase

- RP: an expression of more than one adjuncts will be bracketed with RP label.

Ex:

Tôi (VP đã tìm/Vv-H (RP ra được) (NP cuốn sách)).

- NP: post-modifier is a noun, a FW means as noun, a pronoun, or an expression of noun and its modifiers with be bracketed with NP label.

Ex:

(VP giải_quyết/Vv-H (NP công_việc))

(VP vừa/Rv đi/Vv-H (NP shopping/FW))

(VP giải_quyết/Vv-H (NP những công_việc))

Màn hình (VP bị (VP chia (VP làm (NP 2/Num-H))))

Màn (VP bị (VP chia (VP làm (NP 2 phần/Nn-H))))

Trời (VP vào (NP hạ))

Nó (VP đánh (NP (Pp tôi)))

(VP (Vv Thông_qua) (NP (Pd đó)))

(PP-PRN (Cs-H tức)

(NP (NP (Num-H năm)

(VP (Vv ăn)))

(PU ,)

(NP (Num-H năm)

(VP (Vv thua))))))

- QNP/NP: Pp (without modifier) used to ask will be bracketed with QNP. Otherwise, it is bracketed with NP label.

Ex:

tôi (VP ăn/Vv-H (NP gì/Pp) (RP cũng được))

anh (VP đang làm/Vv-H (QNP gì/Pp))?

- VP:

Directional verb and its modifier is bracketed as a verb phrase modifying the head verb.

Ex:

(VP đi/Vv-H (VP vô/D-H (NP nhà)))

(VP chuyển/Vv-H

(NP-DOB đồ/Nn-H)

(VP-CMP vô/D-H

(NP-CMP nhà)))

Verbs follow volitional verbs (such as: *muốn, cố, định, toan, dám, chịu, buồn, nỡ, mong, thiết, etc.*), modality verbs (such as: *cần, nên, phải, chịu, etc.*), process verbs (such as: *bắt đầu, tiếp tục, hết, thôi, kết thúc, etc.*) or so on will be bracketed with VP label.

Ex:

Tôi (VP muốn/Vv-H

(VP hoàn thành/Vv-H

(NP việc này)

(ADJP sớm)))

Bạn (VP nên (VP làm/Vv-H (NP việc này) ngay/R)).

Chúng ta (VP hãy tiếp_tục/Vv-H (VP làm việc)).

Tôi (Vp đi/Vv-H (VP học).

Nó (VP ra/Vv-H (VP chào tôi)).

- Adjective or expression of adjective and its modifiers will be bracketed with ADJP label.

Ex:

Nó (VP ăn (ADJP rất ít)).

Nó (VP đã làm (ADJP xong))

Nó (VP bị (ADJP đau))

- Preposition and its modifiers (a phrase) will be bracketed with PP label.

Ex:

(VP đi (PP bằng/Cs-H (NP xe bus)))

(VP chờ (PP cho/Cs-H đông đủ))

Post-modifier includes two phrases

We assume that the head verb is modified by a1 and a2 respectively. Where, a1 is a noun phrase, and a2 can be a verb phrase, adjective phrase, etc.

- Equal coordination:
 - o a1 and a2 do not have main-subordinate or subject-predicate relationship. We can invert a1 and a2 without changing of the meaning.

Ex:

Tôi (VP tặng/Vv-H (NP quà) (PP cho cô ấy))

Tôi (VP tặng/Vv-H (PP cho cô ấy) (NP quà))

Tôi (VP viết (NP thư) (PP cho cô ấy))

Tôi (VP viết (PP cho cô ấy) (NP thư))

Note: For the head verbs that mean take or borrow, such as: *lấy, mượn, xin, vay, dành, cướp, bóc lột, huy động, ...* if a2 indicates thing to take/borrow, and a1 indicates person whose has thing (the word *của {of}* can appear before a1), a1 and a2 should be bracketed separately.

Ex:

Tôi (VP mượn (NP-CMP cô ấy) (NP-DOB sách))

Tôi (VP mượn (PP của/Cm-H (NP cô ấy)) (NP sách))

However, if a2 comes before a1, a1 should be treated as a modifier of a2.

Ex:

Tôi (VP mượn (NP-DOB sách (NP cô ấy))

Tôi (VP mượn (NP sách (PP của (NP cô ấy)))

- a1 and a2 seem to be subject and predicate of a single sentence. However, they should be bracketed separately because we can not insert the adjunct indicating tense before a2.

Ex:

(VP tiến (NP bạn) (VP đi học ở nước ngoài))

(VP dắt (NP các bạn) (VP ra vườn hái trái cây))

(VP làm (NP mẹ) (ADJP buồn))

(VP cho (NP con) (VP về nhà ngoại))

(VP quét (NP nhà) (ADJP sạch))

In some cases, we can insert a conjunction before a2.

Ex:

(VP quét (NP nhà) (PP cho (ADJP sạch)))

- If head verb is comparison verb having a post-modifier which indicates object to compare and another indicates result of the comparison, the modifiers will be bracketed separately.

Ex:

(VP hơn (NP bạn) (NP 3 tuổi))

- If a1 is modified by a2, they should be bracketed with main-subordinate construction.

Ex:

(VP có (NP áo mới))

- If we can insert a conjunction, such as: là, rằng, mà, sao, etc. between head verb and the expression of a1 and a2, the expression will be annotated with S label.

Ex:

*(VP tin (SBAR *0*/Cs (S (NP bạn) (VP không làm việc này))))*

(VP tin (SBAR rằng/Cs (S bạn không làm việc này))))

*(VP mong (SBAR *0*/Cs (S con học giỏi)))*

(VP mong (SBAR sao/Cs (S con học giỏi))))

- The following of passive verb (bị or được) always a sentence.

Ex:

(VP bị (S (NP cô giáo) (VP đánh))))

(VP được (S (NP cô giáo) (VP khen))))

3.4.4.4. Caution

- The VPs having the structure: verb (Vv) + Vv + noun/pronoun, where invert Vv and noun/pronoun without changing of the meaning, will be bracketed as (VP (Vv-H) (VP) (NP)).

Ex:

Gió (VP thổi/Vv-H

(VP tan)

(NP mây)).

Nó (VP đập (VP chết) (NP con rắn)).

- Three or more verbs in succession will be bracketed with the structure:

(VP (Vv-H) (VP (Vv-H) (VP (Vv-H) ...))

Ex:

Nó (VP đang ngồi/Vv-H (VP xem/Vv-H (VP đá/Vv-H (NP banh))) trong phòng).

Nó (VP muốn (VP đi (VP xem (VP hát))))).

- Distinguishing between VP and ADJP having the same POS sequence (VP=Aa+Vv, ADJP=Aa+Vv).

Base on meaning of sentence and phrase to determine head word of the phrase (cannot delete head word). If head word is Vv, the phrase will be tagged as VP. If head word is Aa, the phrase will be tagged as ADJP, or

If inverting adjective and verb without changeable meaning, the phrase will be annotated with VP. Otherwise, it will be bracketed with VP if the first word is verb and bracketed with ADJP if the first word is adjective.

Ex:

(VP (ADJP tận_mắt/Aa) chứng_kiến/Vv-H) {to see with one's own eyes}

(VP chứng_kiến/Vv-H (ADJP tận_mắt/Aa)) {to see with one's own eyes}

(ADJP ít/Aa-H ăn/Vv) {to be a small eater}

(VP ăn/Vv-H ít/Aa) {eat little}

However, if there is an adjunct following Vv and modifying Aa, the expression including Aa Vv R will be annotated as ADJP.

Ex:

(ADJP ít/Aa-H ăn/Vv hơn/R)

3.4.5. Adjective phrase (ADJP)

An adjective without modifier or an adjective and its modifiers will be annotated with ADJP label.

Construction: pre-modifier + head + post-modifier

3.4.5.1. Pre-modifier

- Word

An adjunct (R) indicating time (đã, đang, sẽ, ...), comparison (cũng, đều, vẫn, chỉ, cứ, còn, ...), affirmation (có, ...), negation (không, chẳng, chưa, ...), degree (rất, quá, vô_cùng, ...) or imperative adjunct (hãy, đừng, chớ, ...) is annotated the same level with head adjective.

Ex:

(ADJP đã/R đẹp lắm rồi)

(ADJP vẫn/R vui)

- (ADJP có/R gần đây không)

Tình hình (ADJP có/R khác)

(ADJP chẳng/R đẹp)

(ADJP rất/R đẹp)

(ADJP đừng/R buồn/Aa)

- Phrase

Two or more adjuncts in succession will be annotated with RP label.

Ex:

(ADJP (RP sẽ/R rất/R) khó_khăn/Aa)

(ADJP (RP vốn đã không) đẹp)

3.4.5.2. Head

Common adjectives indicating quality (such as: giỏi, đẹp, hiền) or content (such as: nhiều, ít, đông_đúc) annotated with POS tag Aa.

Ex:

Nó (ADJP giỏi/Aa-H toán).

Nó học (ADJP rất giỏi/Aa-H).

Cô ấy (ADJP rất cao/Aa-H).

Nó nói (ADJP rất có_lý/Aa-H).

Cô ấy (ADJP rất ăn_ảnh/Aa-H).

3.4.5.3. Post-modifier

Indicate scope of quality

- A common noun/classifier noun or an expression of common noun/classifier noun and its modifiers will be annotated with NP label.

Ex:

(ADJP nhiều (NP kinh_nghịệm/Nn))

(ADJP khổ (NP mấy đứa con))

Note: Distinguish ADJP and NP having the same POS sequence (Aa Nn)

ADJP usually modifies noun about quality. Meanwhile, NP modifies verb about quantity. In NP, adjective expresses quantity, and we can replace adjective by another quantifier.

Examples:

Anh ấy là người (ADJP nhiều/Aa kinh_nghịệm/Nn)

Anh ấy có (NP nhiều/Aa kinh_nghịệm/Nn) trong việc quản lý công ty.

Tôi làm việc này (PP trong (NP (ADJP nhiều/Aa) năm/Nt-H))

(S (NP-SBJ Chúng tôi) (NP-PRD đã/M (ADJP nhiều/Aa) lần/Nn-H (VP cảnh báo về nguy cơ cháy nổ)))

However, if there is an adjunct following noun and modifying Aa, the expression including Aa Nn R will be annotated as ADJP.

Ex:

Tôi có (ADJP ít tiền hơn).

- Preposition precede NP will be bracketed with PP label.

Ex:

(ADJP khó/Aa-H (PP vì/Cs-H (NP mấy đứa con)))

(ADJP nghiêm_túc (PP trong/Cs-H (NP suy_nghĩ)))

(ADJP giàu (PP về (NP trí tưởng tượng)))

- A temporal noun or an expression of temporal noun and its modifiers will be bracketed with NP label.

Ex:

(NP (ADJP nhiều/Aa-H) hôm/Nt-H) đôi quá.

Tôi đã xong từ (NP (ADJP nhiều/Aa-H) hôm trước)

- A unit noun and its modifiers will be annotated with NP label.

Ex:

(ADJP cao (NP 5 phân))

- A verb or an expression of a verb and its modifiers will be annotated with VP

Ex:

(ADJP khó (VP ăn))

To express the comparison

- Word

An adjunct will be annotated the same level with head adjective.

Ex:

(ADJP quan_trọng nhất/R)

- Phrase

- Comparative verb (Vcp) or an expression of comparative verbs and its modifiers will be annotated with VP label.

Ex:

Bức tranh này (ADJP đẹp (VP hơn/Vcp))

Tôi (ADJP cao (VP kém/Vcp-H (NP bạn) (NP 1 phân)))

(ADJP giỏi (VP hơn tôi rất nhiều))

- PP

(ADJP đẹp (PP như (NP tranh)))

Description

- Word

- Adjunct (R), such as: *quá, lắm, hết_sức, tuyệt, cực, vô_cùng, cực_kì* will be annotated the same level with head adjective.

Ex:

(ADVP đẹp quá/R)

- Modifier (M), such as: *thật, thế, vậy, nhỉ* is annotated the same level with head adjective.

Ex:

(ADJP nóng thật/M)

- Phrase

An adjective will be bracketed with ADJP label.

Ex:

(ADJP rộng (ADJP mệnh_mông))

3.4.5.4. Caution

In some cases, ordinal number (An) plays the role as head word. These phrases are bracketed as ADJP. We can see that, these ADJPs like the phrases missing head word

Ex:

(NP Công_trình/Nn-H này/Pd (/PU (ADJP đầu_tiên/An (PP ở Việt Nam)))/PU)

We can understand this example as:

(NP Công_trình/Nn-H đầu_tiên/An (PP ở Việt Nam))

3.4.6. Prepositional phrase (PP)

Preposition (Cs) and its modifiers will be annotated with PP label.

Construction: pre-modifier + head + post-modifier

3.4.6.1. *Pre-modifier*

- Word

- Pre-modifier which is an adjunct (R) will be annotated the same level with head preposition.

Ex:

(PP đều/R bằng/Cs-H (NP gỗ/Nn))

(PP không/R để/Cs-H (VP gác chân))

- Pre-modifier which is a modifier (M) will be annotated the same level with head preposition.

Ex:

(PP ngay/M ngoài/Cs-H (NP công/Nn))

- Phrase

- Pre-modifier including 2 or more adjuncts will be bracketed with RP label.

Ex:

(PP (RP cũng/R đều/R) bằng/Cs-H (NP gỗ))

- ADJP

Ex:

đĩa trái cây này (PP (ADJP không/Ra phải/Aa) để/Cs-H (VP ăn)).

3.4.6.2. *Head*

- Prepositions, such as *trong, ngoài, trước, trên, dưới, giữa, ở, tại, and etc.* refer to position, manner, sphere, condition, circumstances, impact, cause, target, and etc. for mentioned things to exist or to take place.

Tôi để máy tính (PP ở/Cs-H (NP trường))

Cuốn sách ở/Vv (PP trên/Cs-H (NP bàn))

- Cs is used to indicate the non-physical directions, e.g to indicate time or to come with other non-motion verbs, such as: *vào, lên, đến*.

Examples:

(PP vào/Cs ngày 25-9-2009) {the date 25-9-2009}

chuyển/Vv (PP vào/Cs ngành toán) {mathematic domain}

dự_thi/Vv khỏi A (PP vào/Cs các trường ĐH)

tác động (PP đến/Cs môi trường)

- The way of doing something (*bằng {by}*)

Ex:

nhà (PP bằng/Cs gạch) {brick house}

- The possession (*của {of}*)

Ex:

sách (PP của/Cs tôi) {my book}

- The purpose (*để {to}*)

Ex:

viện cớ (PP để/Cs từ_chối) {find a reason to deny}

- Từ ... đến ...

- o Expressing a range:

Construction: (PP (PP từ ...) (PP đến ...))

Ex:

(NP các tàu ở vùng biển (PP(PP từ (NP Phú Yên))(PP đến (NP Bình Thuận))))

tôi làm việc này mất (MP khoảng/R (PP(PP từ/Cs (NP 2/Num))(PP đến/Cs (NP 6/Num tháng/Nu)))

mất (MP khoảng/R (PP từ/Cs (MP 2-6/Nq tháng/Nu)))

mất (MP khoảng/R (PP(PP từ/Cs (MP 2/Num tháng/Nu))(PP đến/Cs (MP 6/Num tháng/Nu))))

(PP(PP từ (NP ngày 15/2))(PP đến (NP ngày 27/3/1990))

(PP (PP Từ/Cs (NP ngày 2)) (PP đến/Cs (NP ngày 3)))

- Endpoints to indicate direction:

Construction: (PP từ ...) (PP đến ...)

Ex:

(VP đi (PP từ (NP Phú Yên)) (PP đến (NP Bình Thuận)))

(VP đi (PP từ (NP cái đơn giản)) (PP đến (NP cái phức tạp)))

- **Caution**

If đến is replaced by “-”, we annotate as follows:

Range:

Ex:

(NP các tàu ở vùng biển (PP (PP từ (NP Phú Yên)) -/PU (NP Bình Thuận)))

Indicating direction:

Ex:

(VP đi (PP từ (NP Phú Yên)) -/PU (NP Bình Thuận))

If từ is omitted, we annotate as follows:

Range:

Ex:

(NP (Nn-H đoạn)

(UCP (NP (Nn-H ngã_tr) (Nr Sông))

(PP (Cs-H đến)

(NP (Nn-H cầu) (Nr Sông))))

từ ... đến ... which express both or all is not annotated as PP

(NP-SBJ (Cs từ)

(NP (Nn-H công_an)

(NP (Nun-H xã)))

(Cs đến)

(NP (Nn-H công_an)

(NP (Nun-H huyện) (PU ,) (Nun-H tỉnh)))

- *Giữa ... và/với ...*

Construction 1:

(VP liên_kết

(PP giữa ...)

(PP với ...))

(VP (R đã) (Vv-H liên_kết)

(PP (Cs-H giữa)

(NP

(Nn_swsp (Ncs-H việc)

(VP (Vv-H tiêu_thụ)

(NP-DOB

(ADJP (R quá) (Aa-H nhiều))

(Nn-H rượu))))))

(PP (Cs-H với)

(NP (Nn-H nguy_cơ)

(ADJP (Aa-H cao))

(VP (Vv-H bị)

(NP-DOB (Nn-H ung_thư)

(NP (Nn-H vú))))))

Construction 2:

(NP giữa/Cp word và/Cp word)

(NP giữa/Cp tôi/Pp-H và/Cp bạn/Pp-H)

3.4.6.3. *Post-modifier*

- A noun or expression of noun and its modifiers will be bracketed with NP

Ex:

thuận lợi (PP cho/Cs (NP công_việc))

đi (PP bằng (NP xe bus))

- An adjective or expression of adjective and its modifiers will be bracketed with ADJP label

Ex:

học (PP cho (ADJP giỏi))

- Verb or expression of verb and its modifiers will be bracketed with VP label.

Ex:

không phải (PP để/Cs-H (VP ăn))

- A prepositional phrase can have the post-modifier which is a sentence. In this case, PP is the predicate of a sentence.

Việc này không phải tại nó gây ra.

(S (NP-SBJ Việc này)

(PP-PRD không phải tại

(S-CMP nó gây ra))

(PU .))

Caution:

When a Cs has post-modifier which is a sentence, we can be ambiguous that it this structure can be a SBAR. However, SBAR does not have pre-modifier.

(PP

(ADJP (Aa-H có_thể))

(Cs-H do)

(S

(NP-SBJ (Nn-H bạn_bè))

(VP (Vv-H lôi_kéo))))

3.4.7. Adjunct phrase (RP)

- Two or more adjuncts in succession will be bracketed with RP label if they modify the same head word. This type of phrase can be pre-modifier or post-modifier of NP, VP, ADJP, or so on.

Ex:

Chúng ta (VP (RP sẽ/R không/R-H) đi chơi vào tuần tới)

Ai (ADJP (RP rồi/R cũng/R-H sẽ/R) già đi).

(ADJP (RP (R Cũng) (R sẽ) (R-H không) (R còn)) (Aa-H nghèo))

(ADJP Nóng/Aa quá/R) (ADJP (RP cũng/R sẽ/R không/R-H) ngon/Aa)

(VP (Vv-H Trùng_tu) (R mãi)) (VP (RP (R vẫn) (R-H chưa)) (Vv-H được))

Two or more adjuncts in succession. However, if they modify different head word, they are not bracketed with a RP.

Ex:

(NP-PRD (R cũng)

(QP (R trên) (Num-H 1) (Nq-H triệu))

(Nu-H đồng_[]_tháng))

- When adjuncts in succession are linked by comma or conjunction và, all adjuncts are head.

Example:

Các dự án (RP đã/R-H và/Cp đang/R-H) được đầu tư

- When adjunct is modified by other words, these words are also bracketed by an RP

Các dự án (RP đã/R-H và/Cp đang/R-H) được đầu tư

- When an adjunct coordinate with a phrase, the adjunct is bracketed by an RP:

(VP

(UCP

(ADJP (R rất) (Aa-H ít))

(Cp hoặc)

(RP (R-H không)))

(Vv-H khai_báo)

(NP-DOB (Nq những) (Nn-H vụ_việc) (Vv liên_quan)))

3.4.8. Conjunction phrase (CONJP)

Two or more conjunctions in succession will be bracketed with CONJP label.

Ex:

(CONJP (Cp-H cùng) (Cp-H với))

(CONJP (Cp-H và) (Cp-H cùng) (Cp-H với))

(CONJP (Cp-H Ví_dụ) (Cp-H như))

(CONJP (Cp-H tại) (Cp-H vì))

Caution:

- Combinations of Pp and Cp, Cp and Pp, and Cp, Pp and Cp are not CONJP

Ex:

Vậy/Thế nên/thì/mà/nhưng

Vì/Do/Bởi/Tại vậy/thế

Vì/Do/Bởi/Tại thế/vậy nên/mà

- Conjunctions with different meanings standing side by side do not create CONJP

Ex:

Mà trái_lại

Nhưng rồi thay_vì/với/do/nếu

Nhưng nhìn_chung thì

Nhưng dù_sao thì

Nhưng phải_chăng trước_hết

Nhưng nói_chung nếu

Nhưng trước_hết

Thì nói_chung như

- When an adjunct plays the same role as a conjunction, we do not bracket it as a CONJP

Ex:

(S Đồng_thời/R

,/PU

(PP-ADV với việc cục hàng không Việt Nam đề nghị giảm giá cước)

,/PU

(NP-SBJ một số hãng hàng không)

(VP giảm giá vé máy bay)

./PU)

3.4.9. Unlike coordinated phrase (UCP)

- Equal coordinations of unlike phrases or phrase and clause are bracketed as UCP

Ex:

Cô ấy đẹp và nấu ăn ngon.

(S (NP-SBJ Cô ấy)

(UCP-PRD (ADJP (Aa-H đẹp)) (Cp và) (VP (Vv-H nấu_ăn) (Aa ngon)))
(PU .))

Ngọn đèn không tắt và Sông là 2 quyển sách nổi tiếng của Nguyễn Ngọc Tư.

(S (UCP-SBJ (S-TTL Ngọn đèn không tắt) (Cp và) (NP-TTL Sông))
(VP là 2 quyển sách nổi tiếng của Nguyễn Ngọc Tư)
(PU .))

- Equal coordinations of phrases having the same constituent tag will be bracketed with that constituent tag.

Ex:

Cô ấy (ADJP không_những/Cp (ADJP rất đẹp) (CONJP mà còn) (ADJP rất giỏi))

(S (NP-SBJ Hấn)

(VP (Vv-H đổ) (NP (Cp không_chỉ) (NP (Nn-H nhà_cửa)) (Cp mà) (NP cả xe_cộ)) (PP vào cờ bạc))

(PU .))

3.4.10. Questioning phrase

Add Q to the expression of words including interrogative pronoun (such as *ai, gì, sao, nào, mấy, thế_nào, bao_nhiêu, bao_giờ, bao_lâu, đâu, or so on*) which indicates the object to ask.

Ex:

- What

Hôm nay chúng ta ăn (QNP gì/Pp-H)?

đây là (QNP cái gì)?

Bạn đang làm (QNP gì)?

- Possession

cuốn sách này (QPP của (NP ai))?

- đưa bé này con (QPP của (NP ai))?*
- đưa bé này con (QNP ai)?*
- Which
(QNP cuốn sách (NP nào)) của bạn?
 - Who
(QNP Những ai/Pp) sẽ đi chơi?
 - When
(QNP Khi/Nt (NP nào)) bạn về quê?
 - Where
Nhà bạn (QPP ở (NP đâu))?
Bạn từ (QNP đâu) đến?
 - How
(SQ (NP anh)
(VP sẽ giải
(NP bài toán này)
(QPP bằng
(NP cách nào)))
(. ?))
 - Why
Vì sao chưa ai đến?
(SQ (Cp Vì)
(QNP-PRP sao)
(NP-SBJ chưa ai)
(VP đến)
(PU ?))

Caution:

For other type of sentences, such as: declarative sentences, imperative sentences, or so on, the pronouns mentioned above are bracketed with NP label.

Ex:

Tôi ăn (NP gì/Pp) cũng được.

3.4.11. Modal phrase (MDP)

This tag is used to mark the **expressions of exclamation word and modifier**

Ex:

(MDP Trời ạ), tôi làm sao biết được!

When exclamation word or modifier is separated from main components by a comma/... (PU), if is also bracketed with MDP.

Ex:

(S (MDP A/E) (PU !) (NP-SBJ mẹ) (VP đã về))

3.4.11. Unknow phrase (XP)

In the cases that we cannot determine phrase type, the phrases will be bracketed as XP.

Ex:

(S (XP âm âm âm) (PU ,) (NP-SBJ những tiếng nổ mìn) (VP vang vọng) (PU .))

3.5. Function tags

This tag type is to classify grammatical role of words, phrases and clauses in the sentence. Function tags of phrase and clause are tagged following their constituent tags. These two tags are linked together by hyphen (-)

3.5.1. Head of phrase (H)

Each phrase has at least a head word that determines its type. For instance, the head word of a noun phrase is a noun. Thus, the head word of each phrase is an impossible removed constituent. The head word carries general meaning and all grammatical properties of the phrase. Function tag H is used to mark this unit.

In the expressions to name objects or things (A B), A usually fixes, meanwhile, B can be changed.

Examples:

(NP Cá/Nn-H rô/Nn)

(NP Cá/Nn-H chép/Nn)

(NP Cá/Nn-H lia_thia/Nn)

(NP Số/Nn-H tiền/Nn)

(NP Số/Nn-H tài_sản/Nn)

- One head phrase:

Examples:

(NP 3 con/Nc-H mèo đen)

(VP (R sẽ) (Vv-H đi))

(ADJP (R rất) (Aa-H đẹp))

- Multi-headed phrase:

The words in equal coordination are head words .

Ex:

(NP (Nn-H đưa) (PU ,) (Nn-H muốn) (NP (Nn-H kim_loại)))

Note: In adjunct phrases, if the adjuncts parallelly modify a head word, we tag H for all adjunct. Otherwise, if an adjunct modifies other adjuncts, we don't tag H for modifier adjunct.

3.5.2. Subject (SBJ)

This label is used to mark the subject of the sentences or clauses, including null subjects.

Ex:

- **Subject is noun phrase**

Tôi có cái bàn gỗ.

(S (NP-SBJ (Pp-H Tôi))

(VP có cái bàn gỗ)

(PU .))

- **Subject is questioning noun phrase**

Ai chịu trách nhiệm việc này?

(SQ (QNP-SBJ (Pp-H Ai))

(VP chịu trách nhiệm việc này)

(PU ?))

- **Subject is verb phrase**

Có những điều anh hỏi nghe rất ngô nghê.

(S (VP-SBJ (Ve-H Có)

(NP (Nq những) (Ncs-H điều)

(SBAR (Cp *0*) (S (NP-SBJ anh) (VP hỏi))))))

(VP nghe rất ngô nghê)

(PU .))

- **Subject is adjective phrase**

Nóng dễ chịu hơn lạnh.

(S (ADJP-SBJ Nóng)

(ADJP-PRD dễ_chịu hơn lạnh)

(PU .))

- **Subject is idiom**

Ba bề bốn bên đều kín người.

(S (NP-SBJ (ID Ba_bề_bốn_bên))

(ADJP-PRD đều kín người))

(PU .))

Tre già măng mọc là quy luật tự nhiên rồi.

(S (XP-SBJ (ID Tre_già_măng_mọc))

(VP là quy luật tự nhiên rồi)

(PU .))

- **Subject is clause**

Chuột chạy làm vỡ đèn.

(S (S-SBJ Chuột chạy)

(VP làm vỡ đèn)

(PU .))

3.5.3. Logical subject (LGS)

There are two types of passive sentence.

- The first one is complex sentence whose complement is a clause. Subject of this clause is active subject. Subject of complex sentence is passive subject.

Ex:

Nó bị bạn bè chê cười.

(S (NP-SBJ-1 Nó)

(VP (Vv-H bị)

(S-CMP (NP-SBJ bạn bè)

(VP chê cười

(NP-DOB *T*-1))))

(PU .))

- The other form of passive sentence (different from above examples but same to English passive sentence) has active subject following **bởi** {by}. This active subject is annotated with LGS.

Ex:

Nó bị chê cười bởi bạn bè.

(S (NP-SBJ-1 Nó)
 (VP (Vv-H bị)
 (VP-CMP (Vv-H chê cười)
 (NP-DOB *T*-1))
 (PP-IOB (Cs-H bởi)
 (NP-LGS bạn bè)))
 (PU .))

Other examples:

*Người phụ nữ hiện đại bị ràng buộc bởi **những công việc nội trợ***

*Con gái của thanh tra McClane đang bị đe dọa bởi **bọn tin tặc***

*Tôi luôn bị hút hồn bởi **chiếc váy của các cô gái***

Caution:

Distinguish between above structures and the following examples which are not passive sentences. The words *bởi* in these cases means *because*.

Máy bay bị vỡ bởi vỏ máy bay không thể chịu được chấn động mạnh

Nhà tôi bị giải toả bởi dự án Tân Sơn Nhất – Bình Lợi

3.5.4. Predicate that is not VP (PRD)

We mark the predicates that are not VP with PRD label.

Ex:

- Predicate is noun phrase

Tôi cũng sinh viên giỏi.

(S (NP-SBJ Tôi)
 (NP-PRD (R cũng) (Nn-H sinh_viên) (ADJP (Aa giỏi)))
 (PU .))

Tivi này màn hình phẳng.

(S (NP-SBJ Tivi này)

(NP-PRD màn hình phẳng)

(PU .))

- **Predicate is quantifier phrase**

Nhà này 2 tỷ.

(S (NP-SBJ Nhà này)

(QP-PRD 2 tỷ)

(PU .))

- **Predicate is measure phrase**

Nhà này 2 tỷ đồng.

(S (NP-SBJ Nhà này)

(NP-PRD (QP 2 tỷ) đồng/Nu-H)

(PU .))

- **Predicate is adjective phrase**

Cô ấy đẹp.

(S (NP-SBJ Cô ấy)

(ADJP-PRD đẹp)

(PU .))

Hắn giàu nứt đố đổ vách.

(S (NP-SBJ Hắn)

(ADJP-PRD (ID giàu _nứt _đố _đổ _vách))

PU .))

- **Predicate is prepositional phrase**

Bàn này bằng gỗ.

(S (NP-SBJ Bàn này)

(PP-PRD bằng (NP gỗ))

(PU .))

Chiến dịch này từ ngày 3/9 đến 4/10.

(S (NP-SBJ Chiến dịch này)

(PP-PRD từ ngày 3/9 đến 4/10)

(PU .))

- **Predicate is unlike coordinated phrase**

Cô ấy đẹp và nấu ăn ngon.

(S (NP-SBJ Cô ấy)

(UCP-PRD (ADJP (Aa-H đẹp)) (Cp và) (VP (Vv-H nấu_ăn) (ADJP (Aa ngon))))

(PU .))

- **Predicate is clause**

Nhà này mái đã hỏng.

(S (NP-SBJ Nhà này)

(S-PRD mái đã hỏng)

(PU .))

3.5.5. Object and Complement (CMP)

There are 2 types of objects, direct object and indirect object.

3.5.5.1 Direct object (DOB)

- If the transitive verb (Vv) is modified by an object (NP) which indicates person/thing/... is directly impacted by the action (stating at verb) which is implemented by subject, this object will be tagged as DOB.

Direct object answer the questions including “**verb + ai/cái gì**”. It can be the subject of the passive sentence when the sentence is converted into passive sentence.

Ex:

Hắn hút thuốc.

(S (NP-SBJ Hấn)
 (VP (Vv-H hút)
 (NP-DOB thuốc))
 (PU .))

Above sentence can be converted into the passive form as follows:

Thuốc bị hấn hút.

(S (NP-SBJ Thuốc)
 (VP (Vv-H bị)
 (S-CMP hấn hút))
 (PU .))

- If the passive verb is modified by VP, DOB is tagged for the object effected by the head verb of VP

Nó bị chê cười bởi bạn bè.

(S (NP-SBJ-1 Nó)
 (VP (Vv-H bị)
 (VP-CMP chê cười
 (NP-DOB *T*-1))
 (PP-IOB bởi
 (NP bạn bè)))
 (PU .))

- If the verb (Vv) is modified by two NP, the second NP will be tagged as DOB. In this case, we can add the preposition *cho* before the first NP.

(S (NP Tôi)
 (VP tặng
 (NP-IOB cô ấy)
 (NP-DOB hoa))

(PU .))

- For the verb (Vv) is modified by two NP, if we can add a preposition before the second NP to create a prepositional phrase, the first NP is a DOB.

(S (NP Tôi)

(VP nhận

(NP-DOB tiền)

(NP-TMP hôm_qua))

(PU .))

Add a preposition before the second NP:

S (NP Tôi)

(VP nhận

(NP-DOB tiền)

(PP-TMP vào

(NP hôm_qua)))

(PU .))

3.5.5.2. Indirect object (IOB) (giống nhãn BNF của PTB)

Indirect object that could be noun phrase indicates entity/person receiving something or benefiting. There is (or can add) preposition *cho* before indirect object.

If a verb is modified by DOB and IOB, we can invert IOB and DOB.

Ex:

Tôi tặng cô ấy hoa.

(S (NP-SBJ Tôi)

(VP tặng

(NP-IOB cô ấy)

(NP-DOB hoa))

(PU .))

Tôi tặng hoa cho cô ấy.

(S (NP-SBJ Tôi)

(VP tặng

(NP-DOB hoa)

(PP-IOB cho cô ấy))

(PU .))

Exception:

The object having following roles is also annotated with IOB.

1. Victim (nạn nhân)

Ex:

Chúng nó (VP (Vv-H xé) (PP-IOB cho đứa bé) (ADJP-DOB rách áo))

In above sentence, although *rách áo* is an ADJP, we annotated it with DOB. This is because *áo* is a direct object of the verb *xé*.

2. Possessor

In the following order:

(Vv-H) (PP-IOB (Cs-H của) (NP)) (NP-DOB)

Ex:

Nó (VP (Vv-H mượn) (PP-IOB của ngân hàng) (NP-DOB tiền))

3. If *cho* can be replaced by *để*

We do not annotate the phrase following *cho* in this case as IOB

Ex:

Nó (VP (Vv-H dành) (NP-DOB tiền (PP-PRP cho đầu tư)))

3.5.5.3. Complement (CMP)

Complement can be a phrase or a clause which is required by the head verb. We cannot understand the sentence if the complement is deleted. However, complement is different from DOB and IOB.

Examples:

(S (NP-SBJ Tôi)

(VP là/Vc-H

(NP-CMP sinh viên))

(PU .))

In above sentence, the verb *là* is used to express the equivalent between two object. Therefore *là* has to be followed by the complement *sinh viên*.

(S (NP-SBJ Tôi)

(VP có/Ve-H

(NP-CMP rất nhiều tiền))

(PU .))

In above sentence, the verb *có* has to be followed by a complement.

In passive sentence, if the passive verb is modified by a sentence, CMP is tagged for the modifier sentence

Nó bị bạn bè chê cười.

(S (NP-SBJ-1 Nó)

(VP (Vv-H bị)

(S-CMP (NP-SBJ bạn bè)

(VP chê cười

(NP-DOB *T*-1))))

(PU .))

Complement can modify main verb about:

- **Spatial complement**

Mẹ đi chợ.

(S (NP-SBJ Mẹ)

(VP (V_v-H đi)

(NP-CMP chợ))

(PU .))

- **Identity**

Ông xem nó như con.

(S (NP-SBJ Ông)

(VP xem

(NP-DOB nó)

(PP-CMP như con))

(PU .))

Tôi gọi ông ấy bằng bác.

(S (NP-SBJ Tôi)

(VP gọi

(NP-DOB ông ấy)

(PP-CMP bằng bác))

(PU .))

- **Property (đặc tính, tính chất, sở hữu)**

Noun phrases (or quantifier phrases) in relational dependent predicate correspond to property

Tôi là giáo viên.

(S (NP-SBJ Tôi)

(VP (V_c-H là)

(NP-CMP giáo viên))

(PU .))

Ly này không phải là bằng thuỷ tinh

(S (NP-SBJ Ly này)

(VP không phải (Vc-H là)

(PP-CMP bằng thuỷ tinh))

(PU .))

Tôi có xe máy.

(S (NP-SBJ Tôi)

(VP (Ve-H có)

(NP-CMP xe máy))

(PU .))

Bút này của tôi.

(S (NP-SBJ Bút này)

(PP-PRD của

(NP-CMP tôi))

(PU .))

Moreover, some head verbs or idioms such as: *đi, làm, hoá, hoá ra, té ra* require property complement as post-modifier.

Nó đi (NP-CMP lính)

Tôi làm (NP-CMP dẫu)

Nhộng hoá (NP-CMP ngại)

Tưởng thiếu hoá ra (ADJP-CMP đủ)

Tưởng ai lạ, té ra (NP-CMP người quen)

- **Comparative**

Sức nó hơn tôi.

(S (NP-SBJ Sức nó)

(VP (Vcp-H hơn)

(NP-CMP tôi))

(PU .))

3.5.5.3. *Caution*

- DOB, IOB and CMP is tagged for arguments of the verb in following cases:

- o VP playing the role as predicate.

Tôi xem đá bóng ở sân vận động Thống Nhất

(S (NP Tôi)

(VP xem

(VP-CMP đá

(NP-DOB bóng))

(PP-LOC ở sân vận động Thống Nhất))))

- o VP explaining an other word (in this case, VP likes the predicate of a ellipsis clause).

Xi măng sản xuất theo phương pháp truyền thống này bán rất chạy.

(S

(NP-SBJ (Nn-H-1 Xi_măng)

(VP *P* (VP (Vv-H sản_xuất)

(NP-DOB *D*-1)

(VP-MNR (Vv-H theo)

(NP (Nn-H phương_pháp)

(ADJP (Aa-H truyền_thống))))))

(Pd này))

(VP bán rất chạy)

(PU .))

However, when a verb plays the role as a noun, we do not tag the argument tags.

Ex:

Việc nuôi tôm

- Object tags are also in the structures of cho ... biết, cho ... thấy and cho ... hay

Họ cho biết kết quả xét nghiệm rất khả quan.

(S (NP-SBJ Họ)

(VP cho/V_v-H

(NP-IOB *E*)

(VP-CMP biết/V_v-H

(SBAR-CMP (Cs *0*) (S kết quả xét nghiệm rất khả quan))))

(PU .))

- For a predicate including more than one verbs in succession, function tags are tagged for arguments of each verb.

(VP (R đã) (V_v-H tổ chức)

(VP-CMP (V_v-H tặng)

(NP-DOB

(NP (Num 450) (Nn-H áo_ấm))

(PU ,)

(NP (Num 450) (Nn-H mũ)))

(PP-IOB (Cs-H cho)

(NP (Nn-H học_sinh))))

- If coordination verb phrases plays the role as an argument of a head verb, the coordination verb phrases are bracketed with a father VP and we tag argument tag for the father.

(SQ (NP-SBJ Anh)

(VP (Vv-H muốn)

(VP-CMP (VP đi (VP-CMP xem (NP-DOB phim)))

(Cp hay)

(VP đi (VP-CMP ăn))))

(PU ?))

- Distinguishing CMP and modifier: A head word can be modified by the modifiers. However, different from CMP, modifiers are not required by the head word. In the sentence, the modifiers are used to give additional information. We can understand the sentence when the modifiers are deleted. The modifiers are annotated the function tags according to their meanings (if any).

Examples:

Temporal modifier

Tôi mới nhận bằng tốt nghiệp tuần trước.

(S (NP-SBJ Tôi)

(VP mới nhận

(NP-DOB bằng tốt nghiệp)

(NP-TMP tuần trước))

(PU .))

Manner (means or method to do something)

Nó đi học bằng xe đạp.

(S (NP-SBJ Nó)

(VP đi

(VP-CMP học)

(PP bằng xe đạp))

(PU .))

Nó thức rất khuya.

(S (NP-SBJ Nó)

(VP thức

(ADJP rất khuya)

(PU .))

Scope or frequency of activity (phạm vi hoặc tần suất của hoạt động)

Tôi đã chăm sóc cho em ấy về mặt đời sống tinh thần.

(S (NP-SBJ Tôi)

(VP đã chăm sóc

(NP-IOB cho em ấy)

(PP-CMP về mặt đời sống tinh thần))

(PU .))

Nó đi (NP-CMP 2, 3 ngày) rồi

3.5.6. Topicalized (TPC)

This label is used to mark the main component (subject, predicate, object, or complement) which is moved to the position preceding the subject or between subject and predicate in order to emphasize or notice.

Ex:

- ***Completely moved***

Khuyết điểm này tôi đang cố khắc phục. (Tôi đang cố khắc phục *khuyết điểm này*)

Tất cả những khó khăn, chúng tôi sẽ nghĩ cách giải quyết. (Chúng tôi sẽ nghĩ cách giải quyết *tất cả những khó khăn*)

Hành động của hấn *một phần* vì thiếu hiểu biết. (Hành động của hấn vì thiếu hiểu biết *một phần*)

- ***Reappeared***

Con người, đó là cái vốn quý nhất.

Sống, chúng ta mong được *sống* làm người.

- **Group-unit relation**

Gà, tôi mua 2 *con*. (simple sentence)

Gà còn 2 *con*. (existing sentence)

Topicalized could be:

- **Noun phrase**

In simple sentence

Ex:

Khuyết điểm này tôi đang cố khắc phục.

(S (NP-TPC-1 *Khuyết điểm này*)

(NP-SBJ *tôi*)

(VP đang cố khắc phục

(NP *T*-1))

(PU .))

Lĩnh vực này tôi phụ trách.

(S (NP-TPC-1 *Lĩnh vực này*)

(NP-SBJ *tôi*)

(VP phụ trách

(NP *T*-1))

(PU .))

Hàng này công ty tôi phân phối.

(S (NP-TPC-1 *Hàng này*)

(NP-SBJ công ty tôi)

(VP phân phối

(NP *T*-1))

(PU .))

*Tôi thì tôi không ngại.*⁴

(S (NP-TPC Tôi)

(Cp thì)

(NP-SBJ tôi)

(VP không ngại)

(PU .))

In existing sentence

Ex:

Việt còn 2 con.

(SPL (NP-TPC Việt)

(VP (Ve-H còn) (NP-CMP 2 con))

(PU .))

Gạo còn trong kho.

(SPL (NP-TPC-1 Gạo)

(VP (Ve-H còn) (NP-CMP *T*-1))

(PP-LOC trong kho)

(PU .))

In elliptical sentence (missing subject)

Ex:

Còn mình, mới từ Việt Nam qua.

(S (Cp Còn)

(NP-TPC-1 mình)

(PU ,)

4

quote [3;]

(NP-SBJ *-1)

(VP mới từ Việt Nam qua)

(PU .))

In coordination

Ex:

Ông giáo ấy, *thuốc* không hút, *rượu* không uống.⁵

(S (NP-SBJ Ông giáo ấy)

(VP (NP-TPC-1 thuốc)

(VP không hút

NP-DOB *-1)

(PU ,)

(NP-TPC-2 rượu)

(VP không uống

NP-DOB *-2))

(PU .))

In compound sentence

Ex:

Ông giáo ấy, thuốc ông ấy không hút, *rượu* ông ấy không uống.

(S (NP-TPC Ông giáo ấy)

(S (NP-TPC-1 thuốc)

(NP-SBJ ông ấy)

(VP không hút

NP-DOB *-1))

(PU ,)

5

quote [7; 169]

(S (NP-TPC-2 rượu)
 (NP-SBJ ông ấy)
 (VP không uống
 NP-DOB *-2))

(PU .))

An tuy đẹp nhưng cô rất dốt.

(S (NP-TPC-1 An)
 (Cs tuy)
 (S-CNC (NP-SBJ *-1) (ADJP-PRD đẹp))
 (Cs nhưng)
 (S cô rất dốt)
 (PU .))

In complex sentence

Ex:

Tôi cứ *nhà tôi* tôi ở, *việc tôi* tôi làm, *cơm gạo của tôi* tôi ăn.⁶

(S (NP-SBJ Tôi)
 (VP (V_V-H cứ)
 (S-DOB (S (NP-TPC-1 nhà tôi)
 (NP-SBJ tôi)
 (VP ở
 NP-CMP *T*-1))
 (PU ,)
 (S (NP-TPC-2 việc tôi)
 (NP-SBJ tôi)

6

quote [7; 170]

(VP làm
NP-CMP *T*-2))

(PU ,)

(S (NP-TPC-3 cơm gạo của tôi)

(NP-SBJ tôi)

(VP ăn

NP-CMP *T*-3))))

(PU .))

Nơi mình gặp nhau lần đầu anh có nhớ.

(S (NP-TPC-1 (Nn-H Nơi) (SBAR (Cs *0*) (S mình gặp nhau lần đầu)))

(NP-SBJ anh)

(VP có nhớ

NP-DOB *T*-1)

(PU .))

- **Adjective phrase**

Ex:

Trẻ thì hấn trẻ thiệt.

(S (ADJP-TPC Trẻ)

(Cp thì)

(NP-SBJ hấn)

(ADJP-PRD trẻ

(M thiệt))

(PU .))

- **Clause**

Ex:

(*S-TPC-1 Họ đi đâu*), tôi làm sao biết *-1.

3.5.7. Modal phrase (MDP)

The word types express modal meaning in Vietnamese such as exclamation word, modifier, idiom and adjunct.

Examples: (the italics)

- **Modifier**

Tôi nhầm (*M thật*).

- **Adjunct**

Hình_như (*R là*) tôi có gặp anh rồi.

Thằng bé *coi_bộ* thông minh đấy.

- **Modifier**

Đã bảo *mà/mà_lại/mà_lị*

Em không rõ *ạ!*

Nhớ gửi mail *nhé!*

- **Exclamation word**

Chao_ôì, cảnh đẹp quá!

- **Idiom**

Ác_cái_là không ai cho nó hay!

Còn_khuya tôi mới tin!

Té_ra/thì_ra/hoá_ra ai cũng biết!

May_sao bão đi chệch hướng!

- **Exclamation word + Modifier**

(**MDP** *Trời ạ*), tôi làm sao biết được!

MDP marks phrase indicating emotion attitude of speakers or writers beyond above phrases. It is ADJP, VP, NP, or so on having modal meaning.

- **MDP is objective modal phrase:**

Denoting the relation between objective reality and the content of sentence
(positive or negative)

Ex:

(ADJP-MDP *Đúng là*) hoa sữa thơm.

Hoa sữa (ADJP-MDP *đúng là*) thơm thật.

(ADJP-MDP *Đúng thế*) (NP-VOC con ạ), không ai cho không ai cái gì.

(ADJP-MDP *Không phải là*) mỗi câu hỏi đều có câu trả lời.

Mỗi câu hỏi (ADJP-MDP *không phải là*) đều có câu trả lời.

- **MDP is subjective modal phrase:**

Denoting the relation between the emotion or attitude of the speaker or
writer and the content of sentence

Ex:

Noun phrase

Đồ quý!

Đồ hèn!

Quý sứ!

Adjective phrase

Anh ta (ADJP-MDP *chắc_chắn là/M*) biết việc này.

Việc đó (ADJP-MDP *hẳn là*) làm được.

Việc đó (ADJP-MDP *chắc là*) làm được.

Việc đó (ADJP-MDP *chắc_hẳn/Aa là*) làm được.

(ADJP-MDP *Đúng*), tôi làm đấy

Verb Phrase

(VP-MDP *Những_tương*) cuộc sống vợ chồng sẽ bình yên.

Nhà này (VP-MDP *bị cái là*) mái đã hỏng.

Xe của tôi (VP-MDP *được cái là*) máy rất tốt.

3.5.8. Vocative (VOC)

This tag is used to mark phrases indicating addresses.

Ex:

Đóng cửa lại đi con!

(S (NP-SBJ *-1)

(VP đóng cửa lại đi))

(NP-VOC-1 con)

(PU !))

Vocative can be:

- **Exclamation word**

Dạ, em hiểu rồi.

Này, lại đây!

Thôi, đừng nói nữa!

Nào, nói đi!

Ê, đi đâu đấy!

- **Noun phrase**

Em, ngài mai mình đi nhé!

Em kia, lại đây bảo

Phó giám đốc, việc này là sao?

- **Verb phrase**

Thưa chị, tình hình đã thay đổi

Không dám, mời chị!

Xin_lỗi, có cô Lan ở đây không?

Cảm ơn, tôi tự làm được

- **Exclamation word + Exclamation word**

Ô này có mưa không?

À này, ra ngoài nhớ đóng cửa

Alô Alô ai gọi đây?

- **Exclamation word + Modifier**

Ừ hén

Ờ nhỉ

Này nhé

Thôi nhé

Ờ thì

Ờ nhỉ

- **Exclamation word + Noun phrase**

Nè anh,

Thưa bác sĩ,

Dạ thưa chị, tình hình đã thay đổi

- **Noun phrase + Exclamation word**

Mẹ ơi,

Cô Lan này,

- **Exclamation word + Noun phrase**

Nhớ nghe con

3.5.9. Parenthetical (PRN)

Marks phrases or clauses giving extra information (origin, attitude, manner, order, or so on) or clarifying the meaning of whole sentence or its components.

Parenthesis is splitted from the rest by using commas, hyphens, brackets or colons. It can be phrase, coordinate, clause or sequence of clauses.

Ex:

Wang Yin – một phụ nữ 8x đã kết hôn - chia sẻ.

(S (NP-SBJ (Nr-H Wang_Yin)

(PU -)

(S-PRN một phụ_nữ 8x đã kết_hôn)

(PU -))

(VP chia_sẻ)

(PU .))

(NP-SUB (Nn-H Anh)

(Nr Thái)

(LBRK LBRK)

(NP-PRN (Nun-H xã) (Nr Đăk_Búk_So)

(PU ,)

(NP-PRN (Nun-H huyện) (Nr Tuy_Đức)

(PU ,)

(NP-PRN (Nun-H tỉnh) (Nr Đăk_Nông))))

(RBRK RBRK))

(NP (Nn-H Trường) (Nn tiểu_học) (Nr Lý_Tự_Trọng) (PU ,)

(NP-PRN (Nun-H xã) (Nr Đăk_Búk_So) (PU ,)

(NP-PRN (Nun-H huyện) (Nr Tuy_Đức) (PU ,)

(NP-PRN (Nun-H tỉnh) (Nr Đăk_Nông))))

Hai bị_can đã lợi_dụng chức_vụ, quyền_hạn, gây thiệt_hại cho Nhà_nước 80.000 USD (VP-PRN tương_đương 1,2 tỉ đồng).

Quê tôi, **thành phố Huế**, thật là đẹp

Thứ nhất: không có cơ sở để xác định mục tiêu đào tạo và thiết kế nội dung chương trình hợp lý. **Thứ hai:** không có cơ sở khoa học để xây dựng chương trình liên thông giữa các cấp đào tạo. **Ba:** không có cơ sở đánh giá chất lượng đào tạo.

Caution:

- Distinguishing between PRN and SBAR.

They have the same function, giving extra information. However, PRN has the construction as above which is different from SBAR. SBAR has the construction of Cs + S.

Example:

Người phụ nữ (SBAR mà/Cs (S tôi gặp hôm qua)) đã chia sẻ.

In the case that the conjunction is omitted (we can add the conjunction *mà/là/rằng/...* before the sentence):

- If the sentence is placed between pairs of commas, hyphens, brackets or colons, we annotate the sentence as PRN.

In the case that the sentence giving additional information is splitted from main components by one comma (comma stands in front of the sentence giving additional information), if we can add another comma at the end of the sentence giving additional information, we annotate the sentence giving additional information as PRN.

- If the sentence giving additional information is splitted from main component by a comma or a colon,

- If the sentence giving additional information modifies a verb or an adjective, it is a SBAR

*Ông ta nói (SBAR (CS **0) (Sông ta không tham dự cuộc họp hôm nay)).*

- Otherwise, it is a PRN

- If the sentence giving additional information is not splitted from main components by a pair of punctuation, we annotate the sentence giving additional information as SBAR.

given additional information, we will annotate the sentence given additional information as SBAR

(S

(NP-SBJ (Nn-H Điều) (Pd này))
 (VP (Vv-H dựa)
 (PP-CMP-MNR (Cs-H trên)
 (NP (Nn-H kết_quả)
 (VP (Vv-H phân_tích)
 (NP (Num hai) (Nn-H khu_vực)
 (SBAR (Cs *0*) (PU ,)
 (S
 (PP-ADV (Cs-H trong)
 (NP (Pd-H đó)))
 (NP-SBJ (Nr-H Cydonia))
 (PU ,)
 (NP-PRD (Nn-H nơi)
 (SBAR (Cs *0*)
 (S
 (NP-SBJ (NONE *E*))
 (VP (Vv-H phát_hiện)
 (NP-DOB (PU ") (Nn-H khuôn_mặt)
 (PP-LOC (Cs-H trên)
 (NP (Nn-H sao) (Nn Hoả)))
 (PU ")))))))))))))
 (PU .))

- Comma split sentence components (adverb phrase, subject, predicate) or commas to mark PRN:

PRN have to put between two commas. In this case, we do not have the comma splitting sentence components.

Wang Yin – một phụ nữ 8x đã kết hôn - chia sẻ.

(S

(NP-SBJ (Nr-H Wang_Yin)

(PU ,)

(S-PRN một phụ_nữ 8x đã kết_hôn)

(PU ,))

(VP chia_sẻ)

(PU .))

PRN can be a modifier of the word that is given additional information

Ex:

(NP-SBJ (Nr-H Wang_Yin)

(PU ,)

(S-PRN một phụ_nữ 8x đã kết_hôn)

(PU ,))

Or explains a sentence

(S

(S

(NP-SBJ (Num Một) (Nn-H khó_khăn)

(ADJP (Aa-H hi_hữu)))

(VP (Vv-H xảy) (R ra)))

(PU :)

(S-PRN

(S

(NP-SBJ (Nn-H dịch) (Nr SARS))

(VP (Vv-H hoành_hành)

(PP-LOC (Cs-H ở)

(NP (Nr-H Trung_Quốc))))))
 (Cp nên)
 (S
 (NP-SBJ (Nq các)
 (Nn_w (Aa-H chuyên) (Sv gia)))
 (VP (R không_thể) (Vv-H sang)
 (ADJP-TMP (Aa-H đúng)
 (NP (Nn-H thời_hạn))))))
 (PU .))

3.5.10.Subordinate components

Subordinate component is not main component of sentence. It modifies sentence about time, location, purpose, etc.

Annotation guidance for each type of circumstantial constituent as follows:

3.5.10.1. Temporal (TMP)

The subordinate component indicating time will be annotated with TMP label.

Ex:

- Noun phrase

(NP-TMP Sau đó), mọi chuyện bắt đầu rắc rối.

(NP-TMP (Nt-H Mai)) tôi sẽ đến.

(NP-TMP (Nt-H Mùa) (Nt xuân)), trăm hoa đua nở.

(NP-TMP (Nt 10) (Nt-H giờ) (Nt khuya)), hăn lao ra đường.

(NP-TMP Mọi ngày) (NP-TMP giờ này) tôi đều ghé cửa hàng.

*(NP-TMP (Nt-H Lúc) (SBAR (Cp *0*) (S (NP-SBJ (Nn-H anh)) (VP (Vv-H đến)))))) chúng tôi chưa về.*

*(NP-TMP (Nt-H Khi) (SBAR (Cp *0*) (S (NP-SBJ đưa bé) (VP lên sáu)))) thì cha mẹ nó qua đời.*

(NP-TMP Mỗi ngày) uống thuốc 3 lần.

- **Verb Phrase**

(VP-TMP Tiếp_theo đó), y bỏ trốn.

- **Adjective Phrase**

(ADJP-TMP Gần đây), nó có biểu hiện lạ.

- **Prepositional phrase**

(PP-TMP (Cs-H Vào) (TP (Nt-H mùa) (Nt xuân))), trăm hoa đua nở.

*Tôi hay ăn vặt (PP-TMP (Cs-H trong) (NP (Nt-H lúc) (SBAR (Cp *0*) (S (NP-SBJ *) (làm việc))))).*

*(PP-TMP (Cs-H Trong) (NP (Nt-H lúc) (SBAR (Cp *0*) (S (anh) (ngủ)))) thì tôi làm việc.*

(PP-TMP Từ đầu năm đến nay), tôi đã kiếm được kha khá.

TMP is also tagged for the other expression indicating time.

Đại_hội đại_biểu toàn_quốc Hội LHTN VN lần thứ 7 , (NP-TMP nhiệm_kỳ 2014 - 2019)

Caution:

1. If there is no “lúc or khi”, time phrase changes to clause that is not temporal subordinate component.

Ex:

- Single sentence with temporal subordinate component:

*(S (NP-TMP Khi/Nt (SBAR *0*/Cs (S đưa bé lên sáu))), cha mẹ nó qua đời.)*

- Compound sentence:

(S (S Đưa bé lên sáu) (Cp thì) (S cha mẹ nó qua đời).)

2. Subject vs temporal subordinate component

There are the sentences that begin with temporal subordinate components. However, subject is omitted. In these cases, we add null element to subject.

Ex:

(S (NP-TMP *Lúc khó khăn*) (NP-SBJ *E*) (VP *chẳng thấy mặt mũi hấn đâu*).)

(S (NP-TMP *Mỗi ngày*) (NP-SBJ *E*) (VP *uống thuốc 3 lần*).)

3.5.10.2. Locative (LOC)

The subordinate component indicating location will be annotated with LOC label.

Ex:

- Noun phrase

(NP-LOC *Làng bên*), đàn ông bỏ đi hết.

Adjective Phrase

(ADJP-LOC *Xa_xa*), lớp lánh ánh đèn.

- Prepositional phrase

(PP-LOC *Ở đây*), ngày đêm không phân định.

(PP-LOC *Từ/Cm nam ra/Cm bắc*), cảnh quê hương Việt Nam đẹp vô vàn.

Caution:

1. Subordinate components of the sentence (temporal/locative) vs subordinate components of the phrase

In both cases, the subordinate components are annotated with TMP/LOC in proportion with their meaning.

Subordinate components of the phrase	Subordinate components of the sentence
It is impossible to move to a position preceding the subject or other positions (the meaning will change due to a movement)	It is possible to move to a position preceding the subject Modify head verb of predicate
Anh kể về vụ tai nạn (PP <i>trong lớp</i>) (vụ tai nạn xảy ra trong lớp) Anh kể về vụ tai nạn (NP <i>hôm qua</i>) (vụ tai nạn xảy ra hôm qua)	Anh kể về vụ tai nạn (PP-LOC <i>trong lớp</i>) (kể trong lớp) = <i>Trong lớp</i> , anh kể về vụ tai nạn Anh kể về vụ tai nạn (NP-TMP <i>hôm qua</i>) (kể hôm qua) = <i>Hôm qua</i> , anh kể về vụ tai

	nạn
--	-----

2. Distinguish locative subordinate component and other NP.

The locations usually begin with a preposition, such as: *ở, trong, or so on*. In the case that preposition is pruned, if we can insert preposition indicating location before expression (meaning of sentence does not change), the expression will be annotated with LOC.

E.g.

Quán nước mà anh và em gặp nhau lần đầu không còn nữa.

(S (NP-SBJ (Nn-H Quán) (Nn nước) (SBAR (Cp mà) (S anh và em gặp nhau lần đầu))))
(VP không còn nữa)
(PU .))

Nơi mà anh và em gặp nhau lần đầu không còn nữa.

(S (NP-SBJ (Nn-H Nơi) (SBAR (Cp mà) (S anh và em gặp nhau lần đầu))))
(VP không còn nữa)
(PU .))

Anh và em gặp nhau lần đầu (PP-LOC ở quán nước).

Anh nhớ nơi mình gặp nhau lần đầu.

(S (NP-SBJ Anh)
(VP có nhớ
(NP-DOB (Nn-H nơi) (SBAR (Cp *0*) (S mình gặp nhau lần đầu))))
(PU .))

Nơi mình gặp nhau lần đầu anh có nhớ?

(SQ (NP-TPC-1 (Nn-H Nơi) (SBAR (Cp *0*) (S mình gặp nhau lần đầu))))

(NP-SBJ anh)

(VP có nhớ)

(NP-DOB *T*-1)

(PU ?))

Chúng ta đi (NP-DOB **đâu tổ quốc cần**)

(NP-LOC Làng bên), đàn ông bỏ đi hết

(PP-LOC Ở làng bên), đàn ông bỏ đi hết

Đàn ông (NP-LOC **làng bên**) bỏ đi hết

Đàn ông (PP-LOC ở **làng bên**) bỏ đi hết

2. Distinguish locative subordinate component and other subordinate component.

Prepositions, such as: *ở, trong, tại, or so on*. usually begin a phrase indicating location. In this case, prepositional phrase that its head word is prepositions *ở, trong, tại, or so on* will be tagged as location. However, in some cases, preposition *ở, trong, tại, or so on* begins a phrase which does not indicate location. In this case, prepositional phrase will be tagged with functional tag ADV.

Examples:

Tôi gặp anh (PP-LOC tại/Cs-H (NP quán nước)).

Tôi gặp anh (PP-ADV tại/Cs-H (NP lễ trao giải Cánh Diều Vàng)).

3.5.10.3. Manner (MNR)

Indicate the circumstance/method/means which a work/action/... is implemented. It usually begins with *bằng, theo, qua, với, như, or so on*.

Ex:

- Noun phrase

(NP-MNR Với một cây súng và ba viên đạn), Kow Long bám gót giặc.

- Verb Phrase

(S (VP-MNR Theo/Vv ông A) thì/Cp (NP công ty) (VP đã không đủ khả năng tồn tại qua sang năm) ./PU)

(S “/PU (S Công ty đã không đủ khả năng tồn tại qua sang năm) ”/PU ,/PU
(VP-MNR theo ông A) ./PU)

- **Prepositional phrase**

(PP-MNR Qua thử thách), tôi mới biết năng lực của mình

(PP-MNR Bằng giọng hát ngọt ngào), cô đã chiếm được nhiều cảm tình của giám khảo

(PP-MNR Dưới sự điều hành của ông), công ty ngày một ăn nên làm ra.

(PP-MNR Trong điều kiện kinh tế khắc nghiệt), doanh nhân phản ứng dữ dội với quyết định tăng thuế thu nhập cá nhân.

(PP-MNR Nhờ có anh giúp đỡ), công việc mới tốt đẹp

- **Clause**

Subordinate clause's subject and main clause's subject have the relation of whole-partitive

Ex:

(S-MNR **Chân** nọ đá chân kia), **chị** trở về ngôi ở đầu phản.

(S-MNR **Mắt** đeo kính trắng), **người** đàn ông nhìn về phía nhà ga

Note that the following example is still circumstantial constituent

Ex:

Người đàn ông (SBAR-MNR *0* (S mắt đeo kính trắng)) nhìn về phía nhà ga

- **Clause following conjunction “như”**

Ex:

(Cp Như) (S-MNR ai nấy đều biết) (Cp thì) tháng này thường có bão

- **Phrase of elliptical clause**

Still identify as elliptical subordinate clause of compound sentence.
Nevertheless, just annotate this function tag for existing unit (phrase).

Ex:

Trông thấy bà, hắn tái mặt.

(S (S (NP-SBJ *-1) (VP-MNR Trông thấy bà))

(PU ,)

(NP-SBJ-1 hắn)

(ADJP-PRD tái mặt)

(PU .))

Vội vàng, hắn chạy thục mạng về nhà.

(S (S (NP-SBJ *-1) (ADJP-PRD-MNR Vội vàng)

(PU ,)

(NP-SBJ hắn)

(VP chạy thục mạng về nhà)

(PU .))

Muốn tăng thu nhập phải thường xuyên tăng ca.

(S (S (NP-SBJ *E*) (VP-MNR Muốn tăng thu nhập))

(NP-SBJ *E*)

(VP phải thường xuyên tăng ca)

(PU .))

3.5.10.4. Purpose or reason (PRP)

A subordinate component indicating purpose or reason will be annotated with functional label PRP. This type of phrase usually follows a linking word (Cs), such as: *vì, do, tại, bởi, để, để_cho, or so on*. In some cases, the linking word is pruned or replaced with the verb *muốn*.

Ex:

- Noun phrase**

(NP-PRP Cũng chính nguyên nhân này), (ADJP-TMP gần đây), ông thường xuyên cầu gắt

- **Verb phrase**

Đã có vốn (VP-PRP (Vv-H mở) (NP-DOB (Nn-H cơ_sở)))

Mắt dán chặt lên màn hình (VP-PRP theo dõi diễn biến trận đấu)

Một số người chọn đây (VP-PRP làm nơi cư ngụ)

- **Adjective phrase**

Tất cả đồng loạt nổi dậy (ADJP-PRP cùng chống lại cuộc chiến tranh phi nghĩa)

Noun phrase, verb phrase, or clause following causal conjunctions, causal conjunction phrases (such as: *do, vì, tại, bởi, bởi vì, tại vì, nhờ,...*) or purpose (such as: *mà, ...*)

Ex:

Tìm việc (PP-PRP (Cs để) (VP làm))

Nó bỏ học (SBAR-PRP (Cs vì) (S-PRP nhà nó nghèo))

Prepositional phrase beginning with causal prepositions such as: *tại, nhờ* and purpose, such as: *để, cho, để_cho*.

Ex:

Cần đánh dấu lại (PP-PRP cho dễ xác định)

Tôi đến (PP-PRP để anh khỏi lo)

Caution:

- In the cases that *để, cho, etc.* is replaced by the verb *muốn*, the subordinate components will mean manner and should be annotated with MNR

Ex:

(PP-PRP Để đạt kết quả tốt), bạn cần chăm chỉ.

(VP-MNR Muốn đạt kết quả tốt), bạn cần chăm chỉ.

- Subordinate clause of compound sentence

In this case, we can add the conjunction *vì* before causal subordinate clause

Ex:

(S-PRP Đường xá bị ngập) nên chẳng ai muốn ra đường

(NP-TMP Năm nay), (S-PRP mưa nhiều) nên lúa tốt

(S-PRP Đường xá như thế) nên chẳng ai muốn ra đường

- Distinguish between PRP and a sequence of consecutive actions

1. Consecutive actions

(VP ngồi (VP xem đá bóng trong phòng))

This VP is different from:

2. Purpose

(VP ngồi (PP-PRP để xem đá bóng))

Caution:

Distinguish between compound sentence and single sentence including circumstantial constituent.

1. Compound sentence including subordinate clause and main clause, where:

- Each clause has subject and predicate.

Ex:

Vì nó ốm, nó không đi làm được.

- Subordinate clause includes subject and predicate and main clause have only predicate or is a special sentence.

Ex:

Vì nó ốm nên không đi làm được. (main clause has only predicate)

Vì anh nóng vội mà hỏng việc. (main clause is a special sentence)

- Each clause just has predicate or is a special sentence (compound sentence)

Ex:

Vì ốm nên không đi làm. (each clause has only predicate)

Vì đông người nên tắc đường. (each clause is the special sentence)

2. Single sentence including circumstantial constituent.

- If subordinate component which is a word or phrase stands in front of main component which includes subject and predicate, sentence is treated as compound sentence.

Ex:

Vì ốm, nó không đi làm được.

- If subordinate component follows main component which includes subject and predicate, sentence is treated as single sentence.

Nó không đi làm được vì ốm.

Con gà chết vì đói.

Distinguish between special sentence and elliptical clause.

In main-subordinate compound sentence, if main clause has only predicate, we have to recognize main clause is a special sentence or elliptical clause.

Special sentence	Elliptical clause
It is not necessary to add the main constituents to special sentence because the meaning of special sentence is itself complete. If subject of subordinate clause is different from main clause's, subordinate clause the is special sentence.	We can add subject to elliptical clause because 2 clauses have the same subject.
<i>Mưa nên <u>tôi nghỉ học</u></i> <i>Vì đông người nên <u>kẹt xe</u></i> <i>Bão nên <u>lụt</u></i> <i>Có tai nạn nên <u>tắc đường</u></i> <i>Tài chính eo hẹp, <u>công ty phải phá sản</u></i>	<i>Vì ốm, nó không đi làm được.</i>

3.5.10.5. *Adverbial (ADV)*

The circumstantial constituent that does not belong above circumstantial constituents will be annotated with ADV.

Ex:

Với tôi, sự nghiệp và gia đình đều quan trọng.

(S (PP-ADV Với tôi)

(PU ,)

(NP-SBJ sự nghiệp và gia đình)

(ADJP-PRD đều quan trọng)

(PU .))

3.5.10.6. *Cautions of annotating subordinate component*

We base on the context to determine role of an expression in the sentence. However,

- In single sentence, if we cannot determine an expression which is subordinate component of sentence or of the phrase, we annotate it with all possible structures.

Example:

Anh kể về vụ tai nạn hôm qua.

(S (NP-SBJ Anh)

(VP kể

(PP về

(NP-DOB vụ tai nạn (NP hôm qua))))

(./PU))

(S (NP-SBJ Anh)

(VP kể
 (PP về
 (NP-DOB vụ tai nạn))
 (NP-CMP-TMP hôm qua))
 (./PU))

- In compound sentence, if we can not determine an expression is the subordinate component of the clause or of the sentence, we will annotate it as component of the clause which it go with.

Example:

Nay tôi đọc báo, tôi rất mừng thấy anh được pháp luật bảo vệ.

(S (S (NP-TMP Nay) (NP-SBJ tôi) (VP đọc báo))

(./PU)

(S (NP-SBJ tôi) (VP rất mừng thấy anh được pháp luật bảo vệ))

(./PU))

3.5.11. Condition (CND)

Condition can be noun phrase, verb phrase, clause, or elliptical clause indicating condition or supposition. Condition usually follows conjunction or conjunction phrase such as: *nếu, giá mà, giá, miễn là, hễ, nếu, miễn, miễn là, giả_sử, lỡ, nhờ, nếu_mà, nếu_như, giá_như*. Conjunction is sometimes pruned.

E.g.

Nếu (S-CND trời lạnh), chúng ta không đi picnic nữa.

*(NP-TPC-1 Bài này) nếu (S (NP-SBJ *-I) (VP-CND hát nhanh)) thì hay.⁷*

3.5.12. Concession (CNC)

Noun phrase, verb phrase, clause, or elliptical clause following conjunction or conjunction phrase indicating concession such as: *tuy, mặc_dẫu, mặc_dù, dù,*

cho_dù, dẫu, dẫu_cho, dù_cho, tuy_rằng, thà, or so on will be annotated with CNC.

Ex:

Dù (S-CNC mưa to) chúng ta vẫn cứ đi.

*Tuy (S (NP-SBJ *) (ADJP-PRD-CNC giàu)) nhưng bà ta vô cùng keo kiệt.*

3.5.13. Headline (HLN)

This label is used to mark headlines, datetimes, headings and authors (of newspapers). These parts are separate from text. Headline appearing inside text is also annotated with HLN.

Ex:

(S-HLN (ADJP Choáng với cát xê tiền tỷ của sao nhí Hàn))

(S-HLN (TP Thứ tư, 14/01/2013 19:00))

(S-HLN (NP TP.HCM) (PU ,) (TP Thứ tư, 14/01/2013 19:00) (PU ,) (Nguyễn Văn A))

(S-HLN 5 nguyên nhân khiến bạn luôn cảm thấy đói)

(NP-TMP Vừa qua) (PU ,) (S (NP-SBJ (Nc-H bài) (Nn-H báo) (PU “) (NP-HLN Con gái Sài Gòn và con gái Hà Nội) (PU ”)) (VP đã gây sốt cộng đồng mạng))

3.5.14. Title (TTL)

Titles (names) of *films, plays, songs, books, paints, competitions, programs*, and so on which introduced inside texts will be annotated with TTL.

Ex:

Chương trình “ (VP-TTL Tưởng niệm NSND Bảy Nam - một năm ngày mất) ”

Bạn đã đọc cuốn (NP-TTL Người tình Sputnik) chưa?

Bộ phim hoạt hình “ (NP-TTL Frozen) ” sắp khởi chiếu

Cuộc thi (NP-TTL The Voice Việt 2013) sắp đến hồi kết

(S-TTL (NP-SBJ Chương trình “ (VP-TTL Là người con đất Việt) ” thu được hơn 1 tỷ đồng))

Cường con gió bắc là cuốn sách tôi thích nhất.

(S (NP-TTL-SBJ Cường con gió bắc)

(VP là cuốn sách tôi thích nhất)

(PU .))

Tôi thích Cường con gió bắc.

(S (NP-SBJ Tôi)

(VP thích (NP-TTL-DOB Cường con gió bắc))

(PU .))

Ngọn đèn không tắt và Sông là 2 quyển sách nổi tiếng của Nguyễn Ngọc Tư.

(S (UCP-SBJ (S-TTL Ngọn đèn không tắt) (Cp và) (NP-TTL Sông))

(VP là 2 quyển sách nổi tiếng của Nguyễn Ngọc Tư)

(PU .))

Names of brands (games, phones, etc.), organizations (institutions, companies, unions, etc.) should not be bracketed TTL.

3.5.15. Exclamative sentence (EXC)

Exclamative sentence express emotion of the speaker. This type of sentence end with exclamation mark and can include:

- **Exclamation word:** *ôi, trời_ơi, ô_hay, etc.*

Ex:

Ôi, cái bụng tôi đau quá!

(S-EXC (XP-MDP (E Ôi))

(PU ,)

(NP-SBJ cái bụng tôi)

(ADJP đau quá)

(PU !))

Trời ơi, cháy!

(SPL-EXC (XPV (E Trời_oi)))

(PU ,)

(VP cháy)

(PU !))

Đẹp thay!

(SPL-EXC (ADJP Đẹp)

(E thay)

(PU !))

Thương thay hoàn cảnh của em!

(SPL-EXC (VP Thương)

(E thay)

(NP hoàn cảnh của em)

(PU !))

- **Modifier:** *nhỉ, thật, sao (mà), thế, sao (mà)...thế*

Ex:

Mày giỏi nhỉ!

Nó ghê gớm thật!

Chán thế!

- **Adjunct:** *lạ, quá, ghê, biết bao, vô cùng, chết (đi được), etc.*

Ex:

Chán chết đi được!

Nó đẹp ghê!

- **Adjunct and modifier:** *thật là*

Ex:

Thật là đẹp!

(SPL-EXC (Ra Thật)

(M là)

(ADJP (Aa-H đẹp))

(PU !))

- **Adjunct, modifier and exclamation word:** *thật là...thay*

Ex:

Thật là mĩ mai thay!

(SPL-EXC (Ra Thật)

(M là)

(ADJP mĩ mai)

(E thay)

(PU !))

- **Lexical word and exclamation word** *ơi*

Ex:

Con ơi là con!

(SPL-EXC (NP-VOC Con ơi)

(M là)

(NP con)

(PU !))

Buồn ơi là buồn!

Đẹp ơi là đẹp!

3.5.16. Imperative sentence (CMD)

This label is to mark the sentences expressing an order. This type of sentence also end with exclamation mark. They can include an imperative adjunct, such as: *hãy, đừng, chớ, đi, thôi, không được, nào, or so on.*

Ex:

Đừng nói!

(S-CMD (NP-SBJ *E*))

(VP Đừng nói)

(PU !))

Đóng cửa lại đi con!

(S-CMD (NP-SBJ *-1))

(VP đóng cửa lại đi)

(NP-VOC-1 con)

(PU !))

Non-imperative adjuncts

E.x:

Im lặng!

(SPL-CMD (ADJP Im_lặng) (PU !))

Đóng cửa lại!

Anh cứ nói tiếp.

Tiến lên!

REFERENCES

- [1] Thai P. Nguyen, Luong X. Vu, Huyen T.M. Nguyen. 2010. *VTB part-of-speech tagging guidelines*.
- [2] Vietnam Social Sciences Commission. 1983. *Vietnamese Grammar*. Social Sciences Publishers.
- [3] Phe Hoang. 2007. *Vietnamese Dictionary*. Scientific & Technical Publishing.
- [4] MT Group. Institute for Inforcomm Research. 2012. *Guidelines for word segmentation in Vietnamese corpus*.
- [5] Fei Xia. 2000. *The segmentation guidelines for the Penn Chinese Treebank*.
- [6] Ban Q. Diep. 2011. *Vietnamese Grammar*. Vietnam Education Publishing House.
- [7] Ann Bies, Mark Ferguson, Karen Katz, and Robert MacIntyre. 1995. *Bracketing Guidelines for Treebank*.
- [8] Dinh T. Vo. *Từ ngữ tỉnh lược trong tiếng Hán và tiếng Việt hiện đại*.
- [9] Trong D. Le. 1993. *Từ điển giải thích thuật ngữ ngôn ngữ học*. Ho Chi Minh City Publishers
- [10] Hao X. Cao. 1999. *Từ điển giải thích thuật ngữ ngôn ngữ học*. Social Sciences Publishers
- [11] Hanh V. Hoang. *Vietnam Social Sciences Commission*. 2008. *Vietnamese Idioms*. Social Sciences Publishers