

The Part-Of-Speech Tagging Guidelines for the Vietnamese Treebank (0.1)

Quy T. Nguyen and Yusuke Miyao
June, 2016

Table of content

1. Introduction	4
2. Sub-word (syllable)	4
2.1. Brief description of sub-word tag.....	4
2.2. Principles of POS annotation	4
3. Word.....	6
3.1. Brief description of word tags	6
3.2. Principles of POS annotation	8
3.2.1. Noun.....	8
3.2.1.1. Classifier noun (Nc)	9
3.2.1.2. Special classifier noun (Ncs).....	12
3.2.1.3. Unit noun (Nu)	13
3.2.1.4. Special unit noun (Nun)	15
3.2.1.5. Nouns of time (Nt)	16
3.2.1.6. Quantifier nouns	17
3.2.1.7. Proper nouns (Nr).....	19
3.2.1.8. Common nouns (Nn).....	23
3.2.2. Verb.....	25
3.2.2.1. Existing verb (Ve)	25
3.2.2.2. Copula “là” verb (Vc).....	26
3.2.2.3. Directional verbs (D).....	27
3.2.2.4. Comparative verb (Vcp) ([6], page 85)	29
3.2.2.5. Other verbs (Vv).....	31
3.2.3. Adjectives.....	32
3.2.3.1. Ordinal number (An).....	32
3.2.3.2. Other adjectives (Aa).....	33
3.2.4. Pronouns.....	34
3.2.4.1. Demonstrative pronouns (Pd).....	34
3.2.4.1.1. Other pronouns (Pp)	35
3.2.5. Adjuncts.....	35
3.2.6. Conjunction	38
3.2.6.1. Prepositions and conjunctions introducing relative clauses (Cs)	38
3.2.6.2. Other conjunction (Cp).....	41
3.2.7. Double POS.....	43
3.2.7.1. Verb-adjective (VA).....	43
3.2.7.2. Verb-noun (VN)	43
3.2.7.3. Noun-adjective (NA).....	43
3.2.8. Abbreviations/acronym	44
3.2.9. Onomatopoeia (ON).....	44
3.2.10. Punctuations (PU).....	45
3.2.11. Idioms (ID)	45
3.2.12. Exclamation words (E).....	45
3.2.13. Modifier (M).....	46
3.2.14. Foreign words (FW)	47
3.2.15. Unidentified POS (X).....	47
4. Internal structure tags	47
4.1. Brief description of W (word) and SWSP (supra-word sub-phrase).....	47
4.2. Principles of W and SWSP annotation.....	48
4.2.1. Principles of W annotation	48
4.2.1.1. Common nouns (Nn_w)	48

4.2.1.2.	Verb (Vv_w).....	49
4.2.1.3.	Adjective (Aa_w)	50
4.2.1.4.	Adjunct (R_w).....	51
4.2.2.	Principles of SWSP annotation	51
4.2.2.1.	Noun sub-phrase (Nn_swsp)	51
4.2.3.	Note	53
4.2.3.1.	Repetition form.....	53
5.	Appendices	54
5.1.	Combination rules of Sino-Vietnamese elements and other words	54
5.2.	List of classifier nouns	56
5.3.	List of special classifier nouns	58
5.4.	List of unit nouns.....	59
5.5.	Unit nouns of administrative location	67
5.6.	List of nouns of time	68
5.7.	Distinguishing between Cs and other word types	72
REFERENCES		74

1. Introduction

This document is designed for Vietnamese corpus. It shows how a word should be tagged part-of-speech (POS) in a particular context. For each word type or syllable, we show the way to recognize it, the ambiguous word types, ect.

The rest of this document is recognized as follows. Section 2 presents a special type of syllable in Vietnamese and the way to tag them. Section 3 will show POS tag set and how the words are tagged POS based on the context. Set of the internal structure tags and the way to annotate them will be presented in section 4. Finally, in section 5, we will present lists of closet words of each word type.

If you have questions that you do not find in the guidelines, please let us know so that we can discuss and update the guidelines.

2. Sub-word (syllable)

2.1. Brief description of sub-word tag

Groups	POS	Tag
Affix	Sino-Vietnamese	SV

2.2. Principles of POS annotation

A Sino-Vietnamese element, such as: *siêu* {super}, *bất* {not}, *đồng* {same}, *đa* {many}, *tiểu* {little}, *tân* {new}, *cựu* {old}, *nguyên* {former}, *cố* {late}, *vi* {micro}, *gia* {-er}, *hoá* {-ize}, *trưởng* {head}, *phó* {deputy, vice}, *phi* {non}, *đại* {great}, and etc. combines with a Vietnamese noun, verb or adjective to create a new word. Meaning of the new word is different from original noun, verb or adjective. The Sino-Vietnamese elements can come before or follow the words they modify.

Examples:

phương trình {equation} → *bất* phương trình {inequation}

công nghiệp {industry} → công nghiệp *hóa* {industrialize}

List of the Sino-Vietnamese elements and the rules to create new words are presented in appendices 5.1.

In this document, the Sino-Vietnamese elements are splitted¹ from the words they modify and tagged as SV.

¹We use the internal structure tags to annotate the sequences of Sino-Vietnamese elements and noun/verb/adjective, see bracketing guidelines for more details.

Examples:

bất/SV phương_trình/Nn²

bất/SV lịch_sự/Aa {impolite}

bất/SV bình_đẳng/Aa {injustice}

bất/SV chính/Aa {unrighteous}

đa/SV phương_tiện/Nn

huấn_luyện/Vv viên/SV {coach}

diễn/Vv viên/SV {actor, actress}

giáo/Nn viên/SV {teacher}

vô/SV điều_kiện/Nn {unconditional}

siêu/SV lợi_nhuận/Nn {super profit}

tiểu/SV sản_xuất/Vv {small production}

tiền/SV Seagames/FW {pre-Sea Games}

hạ/SV Sahara/Nr {sub-Saharan or south of the Sahara}

chi_cục/Nn trưởng/SV

viện/Nn trưởng/SV

đại_đội/Nn trưởng/SV {captain}

trưởng/SV khoa/Nn {head of department}

phó/SV trưởng/SV công_an/Nn

nguyên/SV phó/SV chủ_tịch

phó/SV tổng/SV giám_độc/Nn {deputy general director}

tân/SV thủ_tướng/Nn {prime minister}

ngọt/Aa hoá/SV

công_nghiệp/Nn hóa/SV

cựu/SV siêu/SV xe/Nn

hậu/SV kỳ/Nn

Caution:

²The tags, such as: Nn, Aa, etc. are explained in the below sections of this document.

Distinguishing between Sino-Vietnamese element and Vietnamese word.

Ex:

Bộ/Nn Xây_dựng: *Bộ* is a national organization vs *Đảng bộ, huyện bộ, ...*

If Vietnamese element has post-modifier, we treat Sino-Vietnamese element as Nn. This means that the expression including the Vietnamese element and its post-modifier is a phrase modifying Nn.

Ex:

Trường/Nn (NP ban văn nghệ) vs *bộ trưởng/SV, viện trưởng/SV, đại đội trưởng/SV, ...*

Exceptions:

There are expressions including Sino-Vietnamese elements, however, they are treated as the compound words³. These words will be tagged POS according to their meanings in the context.

Examples:

bất_di_bất_dịch/ID {immutable}

bán_cầu_não/Nn

đa_dạng_sinh_học/Aa

địa_hoá_học/Nn

đa_thần_giáo/Nn

hiệu_trưởng/Nn

sinh_viên/Nn

thứ_trưởng/Nn

đại_đoàn/Nn

List of these words will be presented in our word segmentation guidelines.

3. Word

3.1. Brief description of word tags

Groups	POS	Tag	Description
Noun	Classifier noun	Nc	Indicating that an entity belongs to animal or inanimate class.

³Details are presented in our word segmentation guidelines.

	Special classifier noun	Ncs	Classifier nouns that have functions similar to Sino-Vietnamese elements in creation of new words. They cause the changing of the meaning and POS of the words they modify.
	Unit noun	Nu	Units used to measure, calculate
	Special unit noun	Nun	Unit nouns of administrative location and unit nouns of organization
	Noun of time	Nt	Nouns imply the time
	Quantifier noun	Nq	Nouns imply the quantity but not numbers
		Num	Quantifiers can write by numbers
		Nw	Quantifiers indicating the whole
	Proper noun	Nr	Proper names of people, things
	Other nouns	Nn	Other nouns, including single nouns, collective nouns, abstract nouns, etc.
Verb	Existing verb	Ve	Verbs indicate the state of the entities
	Copula “là” verb	Vc	Verb has a special quality, expressing A = B (A and B are subject and object which have the same phrase type)
	Directional co-verb	D	Verbs modifying directions for other verbs
	Comparative verb	Vcp	Verbs to compare the objects
	Other verbs	Vv	Transitive, intransitive, transformative, volition, acceptation, etc.
Adjective	Ordinal number	An	Adjectives indicate orders, positions of entities
	Other adjectives	Aa	Other adjectives
Pronoun	Demonstrative pronoun	Pd	Pronouns used as suffixes for other nouns
	Other pronouns	Pp	Other pronouns
Adjunct		R	Adjuncts modifying nouns, verbs, adjectives about tenses, degrees, etc.
Conjunction		Cs	Prepositions and conjunctions introducing a SBAR
		Cp	Other conjunctions
Others	Noun-Adjective	NA	
	Noun-Verb	VN	
	Verb-Adjective	VA	
	Onomatopoeia	ON	Words describe sounds
	Punctuation	PU	Including all punctuations
	Idioms	ID	Fixed combinations of words whose meanings often cannot be interpreted simply by the words forming them.
	Exclamation word	E	Words used to express emotional reactions, call

			and response
	Modifier word	M	Words used to express the speaker's attitudes (surprised, suspicious, ironic, happy...)
	Foreign word	FW	Words adopted from foreign languages
	Unidentified words	X	Words whose POS cannot be identified

3.2. Principles of POS annotation

We base on following criteria to determine POS tag of a word:

- *Combining ability* of the word.

Examples:

khó_khăn can be understood as *difficulty* or *difficult*. However, if it is a noun, it cannot combine with the adjunct *rất* {very}. If it is an adjective, it cannot combine with the quantifier *những* {-s/-es}.

- Syntactic function of the word.

Examples:

If the quantifier indicating the whole modifies a noun, it will be annotated with an Nw tag. The quantifier indicating the whole will be annotated with a Pp tag if it is the head word of the noun phrase.

- Meaning of the words in the sentence.

Examples:

The combination ability of the verb *đi* {go} and the adjective *đẹp* {beautiful} mentioned above is the same, they are modified by the adjunct *đã*. They also have the same syntactic function which are the head word of the predicate. However, their meanings are different: the adjective expresses the quality, and the verb expresses the action.

cô/P mua một chiếc xe *cổ/Aa* {she has bought an ancient car}

3.2.1. Noun

Nouns are classified into 10 categories, classifier noun (Nc), special classifier noun (Ncs), unit noun (Nu), special unit noun (Nun), nouns of time (Nt), numbers (Num), quantifiers

indicating the whole (Nw), other quantifiers (Nq), proper nouns (Nr), other nouns (Nn). Annotation instruction of each category is given below.

3.2.1.1. Classifier noun (Nc)

Classifier nouns are used to classify single entities into the units of animal or units of inanimate thing. A classifier noun does not point to a specific thing/object if it is not modified by a common noun (categorization noun) or refer to a thing/object. Therefore, in the sentence, the classifier noun is always followed by a categorization noun. If classifier noun is not followed by a categorization noun, the expression of classifier noun and categorization noun has to occur before.

Example:

Tôi có 2 *con*_{classifier noun} mèo_{categorization noun}, một *con*_{classifier noun} có đuôi, *con*_{classifier noun} còn lại cụt đuôi.

Categorization noun cannot follow a number. Therefore, to count thing/object, categorization noun has to follow classifier noun in the form: quantifier + classifier noun + categorization noun.

The classifier nouns are tagged as Nc, their modifiers are always tagged as Nn (common nouns).

Two kinds of single entities mentioned here:

- Animated object: Single entity follows “*con*” denoting a unit of animal or animal-like thing.

Examples:

con/Nc bò/*Nn* {cow}

con/Nc kiến/*Nn* {ant}

con/Nc sông/*Nn*{river}

The classifier nouns, such as: *cậu*, *ông*, *cái*, *cô*, *chú*, *bác*, *người*, and *etc.* also denote units of animal or animal-like thing. Therefore, they are also tagged as Nc.

Examples:

chú/Nc voi/*Nn*

chú/Nc chim/*Nn*

người/Nc lính/*Nn*

người/Nc nông_dân/*Nn*

- Inanimate object: Single entity follows “*cái*” denoting a unit of inanimate thing, a small or personified animal

Examples:

cái/Nc nhà/Nn {house}

cái/Nc áo/Nn {shirt}

cái/Nc xe/Nn {car}

cái/Nc kiến {ant}

The classifier nouns, such as: *chiếc, bức, quyển, and etc.* also denote units of inanimate thing. Therefore, they are also tagged as Nc.

Examples:

chiếc/Nc xe {car}

bức/Nc tranh {picture}

áng/Nc văn/Nn {literary work}

bài/Nc thơ/Nn {poem}

bài/Nc diễn_văn/Nn {speech}

bản/Nc tuyên_ngôn/Nn {declaration}

A list of classifier noun is described in appendix 5.2.

To distinguish Nc and other types of noun, we base on characteristics of Nc as follow:

- When the classifier nouns are used to reference, they are not followed by their modifiers, they are tagged as Nc.

Example:

Tôi có 2 *con/Nc* mèo/*Nn*, một *con/Nc* có đuôi, *con/Nc* còn lại cụt đuôi.

- If the modifiers is the nouns which generally mention human, the classifier nouns *ông, bà, chú, bác, anh, chị, cô, gã, chàng, đồng chí, thằng, thầy, and etc.* are tagged as Nc

Example:

Đồng_chí/Nc chủ_tịch/*Nn* {Comrade Chairman}

Anh/Nc nông_dân/*Nn* {A young peasant}

- If the classifier nouns are used to express the family relationship, they are tagged as Nn

Example:

Chị/Nn gái/*Nn* {elder sister}

Em/Nn trai/*Nn* {younger brother}

Con/Nn gái/*Nn* {daughter}

Con/Nn trai/*Nn* {son}

Con/Nn dâu/*Nn* {daughter-in-law}

Con/Nn rể/*Nn* {son-in-law}

Chị/Nn dâu/*Nn* {sister-in-law (wife of ones' elder brother)}

Anh/Nn rể/*Nn* {brother-in-law (husband of one's elder sister)}

Anh/Nn chồng/*Nn* {brother-in-law (elder brother of one's husband)}

Chị/Nn vợ/*Nn* {sister-in-law (elder sister of one's wife)}

Bà/Nn nội/*Nn* {paternal grandmother}

Ông/Nn ngoại/*Nn* {maternal grandfather}

Tôi có 2 *con/Nn*.

Anh/Nn đi làm còn *chị/Nn* ở *nhà/Nn* chăm *con/Nn*.

Note: The words “*anh, chị, cô, bà, ông, and etc.*” are tagged as M when they denote genders, in these cases, they are used to denote the attitudes of the speaker or emphasis. They stand in front of head nouns.

Examples:

Con/Nn gái/*Nn* {daughter}, *Cô/M* *con/Nn* gái/*Nn* {Ms. daughter}

Dâu/Nn {daughter-in-law}, *cô/M* *dâu* {Mrs. daughter-in-law}, *con/M* *dâu* {Mrs. daughter-in-law}, *chị/M* *dâu* {Mrs. daughter-in-law}

- If modifiers are the proper names, the classifier nouns are tagged as Nn

Examples:

Ông/Nn Dũng

Lão/Nn Trương

Bà/Nn Hoa

- If modifiers are a demonstrative pronoun which replace a personal name, the classifier nouns are tagged as Nn

Examples:

Ông/Nn này

- The classifier nouns do not come before verbs or adjectives.

Examples:

Bài/Nn thi/Vv

Bài/Nn kiểm_tra/Vv

- *Cái* will be labeled as Nn when it expresses things. In this case, *cái* is not modified by a categorization noun.

Examples:

Cái/Nn cần phản đối

3.2.1.2. Special classifier noun (Ncs)

There are the classifier nouns that their modifiers can be verbs or adjectives, such as: *sự, việc, cuộc, nhà, nỗi, niềm, điều, vụ, cái, and etc.* We call them as special classifier nouns. The special classifier nouns don't have clearly meaning, therefore, they cannot stand alone as a component of sentence (subject or object). They always go with their modifiers. In addition, special classifier nouns cause changing of meaning (or POS) of their modifiers. A sequence of a special classifier noun and their modifier has meaning as a noun which denotes a person, a thing, an event, or etc.

Examples:

Special classifier noun	Modifier is verb	Sequence of special classifier noun and their modifier
cái	nhìn {look}	cái nhìn {view}
sự	chú ý {pay attention}	sự chú ý {attention}
niềm	thương nhớ {miss}	niềm thương nhớ {nostalgia}
cuộc	thi {take an exam}	cuộc thi {examination/ competition}
nhà	nghiên cứu {research}	nhà nghiên cứu {researcher}
điều	tâm đắc {understand thoroughly}	điều tâm đắc {favourite thing}
niềm	tin {believe}	niềm tin {belief}
nỗi	buồn {sad}	nỗi buồn {melancholy}
vụ	tham ô {embezzle}	vụ tham ô {embezzlement}
việc	đi lại {travel}	việc đi lại {traffic}

Special classifier noun	Modifier is adjective	Sequence of special classifier noun and their modifier
cái	đẹp {beautiful}	cái đẹp {beauty}

sự	uyên bác {erudite}	sự uyên bác {scholarship}
niềm	hạnh phúc {happy}	niềm hạnh phúc {happiness}
cuộc	vui {fun}	cuộc vui {treat/spree}
nhà	hảo tâm {kind}	nhà hảo tâm {benefactor}
điều	kì diệu {magical}	điều kì diệu {miracle}
niềm	tin {believe}	niềm tin {belief}
nỗi	vất vả {hard}	nỗi vất vả {hardship}
vụ	rắc rối {complicated}	vụ rắc rối {complication}

The special classifier nouns are tagged as Ncs.

Examples:

cuộc/Ncs gặp_gỡ/Vv {meeting}

nhà/Ncs hảo_tâm/Aa {benefactor}

Caution:

The words *việc*, *điều*, or so on will be tagged as Nn in the context that it cannot be replaced by the word *sự*. In this case, it is usually modified by a sentence or a phrase.

Examples:

Việc/Nn tôi đi học trễ đã làm ảnh hưởng đến cả lớp.

Điều/Nn khiến tôi hài lòng nhất là những kỳ nghỉ cùng gia đình.

The words *việc*, *điều*, or so on will be tagged as Nn if it is followed by a SBAR

Examples:

Việc/Nn cục hàng không Việt Nam đề nghị giảm giá cước

3.2.1.3. Unit noun (Nu)

The materials, such as “*nước*” {water}, “*đất*” {land}, “*rượu*” {wine}, “*thịt*” {meat}, “*sắt*” {iron}, “*thép*” {steel}, currency, etc. cannot exist as single entities. The quantifier nouns cannot precede the common nouns which show the materials or currency to express the measure and calculation. However, we can measure or calculate the materials and currency through the unit nouns. For example, we cannot say “*một đất*” or “*2 thịt*”. However, we can say “*một sào đất*” or “*2 kg thịt*”. The words “*sào*” and “*kg*” in these examples are unit nouns. The unit nouns are not single entities. They cannot be modifiers of the classifier nouns.

- Accurate unit nouns of materials:
 - Volume: *lít* {liter}, *mililit* {mililitre}, *m³*, and etc.
 - Length: *cm*, *m*, *dm*, *thước* {metre}, and etc.

- Area: m^2 , km^2 , sào {acre}, mẫu {piece}, and etc.
- Weight: gam, kg, cân {pound/kilo}, tấn {tonne}, and etc.
- Velocity: km/h, and ect.
- Electric capacity: wh, kwh, and etc.
- Inaccurate unit nouns of materials (popular unit): liều (thuốc), sải (tay), bó, nắm {handful of}, miếng {piece}, chông, cục, thỏi, viên, mảnh, ngụm {sip of}, hóp, vốc, and etc.
- Unit nouns including slash: người/thang, km/h, lít/người, ...
- Unit nouns of currency/money: đô_la {dollar}, dollar, Việt Nam đồng {Vietnam dong}, VNĐ, yên {yen}, £, USD, and etc.

A list of unit nouns is described in appendix 5.4.

The unit nouns are tagged as Nu. They play the role as head words in measure phrases.

Examples:

2 lít/Nu nước {water} (2 litres of water)

10 tấn/Nu sắt {10 tonnes of iron}

30 sào/Nu đất {30 acres of land}

một nắm/Nu gạo {a handful of rice}

100,000 đồng/Nu {100.000 vnd}

Note:

Nc, Nn and Nu are preceded by Num or Nq to indicate the measurement or calculation. However, the unit nouns indicate quantity. Meanwhile, Nc denotes the classification, and Nn indicates general meaning. In addition, we can base on the following principles to distinguish these word types.

- Distinguishing between Nu and Nn: There are some things which are also used to measure the materials. However, we do not tag them as Nu because these things are single entities (can be modifiers of Nc). They are tagged as Nn.

Examples:

Hai vạ/Nn lúa

Hai bao/Nn lúa

Hai mớ/Nu lúa

- Distinguishing between Nu and Nc:
 - o Modifiers of Nc or Nc themselves denote the single entities. Meanwhile, Nu and its modifiers are not single entities.

Examples:

quyển/Nc sách/Nn

bó/Nu rau/Nn

- o Post-modifiers of Nc are usually countable nouns, meanwhile, post-modifiers of Nu are usually uncountable nouns.

Examples:

Con/Nc gà

Cái/Nc áo

Cân/Nu thịt

Lít/Nu nước

- o Post-modifiers of Nc are objects, meanwhile, post-modifiers of Nu are substances.

Unit noun including expression is tagged as follows:

Example:

2,2/Num tỷ/Nq đồng/Nu //PU căn/Nc nhà/Nn

1/Num lít/Nu sữaNn //PU ngày/Nt

3.2.1.4. Special unit noun (Nun)

The special unit nouns are tagged as Nun. They include unit nouns of administrative location, such as: *làng* {village}, *xã* {commune}, *tổ, tổ dân phố, lô* {block/neighbourhood}, *tỉnh* {province}, *thành phố* {city}, *khu, khu vực, khu phố* {quarter}, *châu* {Vietnam's mountain district}, *huyện* {suburban district}, *thị trấn, lệ* {Vietnam's township}, *nước, quốc gia* {country}, *khu vực* {region}, *châu lục* {continent}, *lục địa* {mainland}, and etc.

We also add the quantifier nouns before the special unit nouns (pre-modifiers) in order to count. However, the special unit nouns are different from unit nouns as follows: if post-modifiers of special unit nouns are proper nouns, pre-modifier cannot be quantifier nouns. Otherwise, pre-modifiers are quantifier nouns, post-modifier cannot be proper nouns.

Examples:

hai/Nq thành phố/Nun {2 cities} (correct)

hai/Nq thành_phố/Nun xinh_đẹp/Aa {2 beautiful cities} (correct)

thành_phố/Nun Hồ_Chí_Minh/Nr {Ho Chi Minh city} (correct)

hai/Nq thành_phố/Nun Hồ_Chí_Minh/Nr {2 Ho Chi Minh city} (incorrect meaning)

In addition, the special unit nouns have characteristics which are similar common noun, such as: thành_phố/Nun biển/Nn, thành_phố/Nun núi/Nn, làng/Nun chài/Nn, làng/Nun góm/Nn, làng/Nun chiến_đấu/Vv and etc. The special unit nouns are head words of noun phrases.

3.2.1.5. Nouns of time (Nt)

The nouns indicating time (*durations and time points*), such as: giờ/tiếng {hour}, phút {minute}, giây {second}, buổi {half a day}, thứ {day of week}, ngày {date}, tuần {week}, tháng {month}, quý {quarter}, mùa {season}, {năm {year}, thế_kỉ {century}, thiên_niên_kỉ {millennium}, kỉ {period}, thời {(một) lát {(a) while}, (một) chốc {(a) while}, khi {when}, chủ_nhật, xuân, sáng, Tết, xưa, nay, hôm_qua, hiện_nay, ngày_xưa, sáng, trưa, chiều, tối, bình_minh, hoàng_hôn... are tagged Nt.

The numbers indicating time are tagged as Num.

Examples:

Thứ/Nt 2/Num ngày {date}/Nt 10/Num tháng {month}/Nt 5/Num năm {year}/Nt 2009/Num

Thứ/Nt hai/Num ngày/Nt ba/Num mươi/Nq tháng/Nt tư/Nq năm/Nt một/Num nghìn/Nq chín/Num trăm/Nq bảy/Num mươi/Nq lăm/Nq

If the phrase of time including giờ {hour}, phút {minute}, giây {second} is written in full form: 2 giờ sáng, 5 giờ 5 phút 5 giây, ... we tag each of its components individually according to above principle.

Examples:

2/Num giờ/Nt sáng/Nt

5/Num giờ/Nt 5/Num phút/Nt 5/Num giây/Nt

If the phrase of time is written in in short form, such as 10g, 10g30, 10h30m45s, 10h30'45'', and etc., we treat them as a word and tag as Nt

Examples:

30/ 04/1975/Nt, 30/_04/_1975/Nt, 30-04-1975/Nt, 30_-04_-1975/Nt

10g/Nt, 10_g/Nt {10h}

10g30/Nt, 10_g_30/Nt, ... {10h30}

10h30 m 45 s/Nt, 10h_30 m _45 s/Nt, ... {10h30m45s}

10h30'45''/Nt, ... {10h30m45s}

A list of noun of time is described in appendice 5.6

Note:

- Distinguishing between Nu and Nt: Nu is tagged for a unit of time which appears in measure phrase. Nt is tagged for noun of time to indicate time point, time period, etc.

Examples:

Mất 2/Num giờ/Nu để làm việc này {It take 2 hours to do this work}

Cuộc họp diễn ra từ/Cs 2/Num giờ/Nt chiều/Nt đến/Cs 3/Num giờ/Nt chiều/Nt
(The meeting is from 2 pm to 3 pm)

Cuộc họp bắt đầu lúc/Cs 2/Num giờ/Nt chiều/Nt (The meeting begin at 2 pm)

3.2.1.6. Quantifier nouns

- Quantifiers indicating total will be tagged as Nw

Examples:

tất_cả/Nw, cả/Nw, hết_thấy/Nw, toàn_bộ/Nw, etc.

- Numbers indicating quantities or times are tagged as Num

Examples:

30/Num, 30.5/Num, 30,5/Num

một {one}/Num, mười {ten}/Num

- Quantifier which cannot write by number are tagged as Nq

Examples:

đôi/Nq, cặp/Nq, chục/Nq, mười/Nq, tá/Nq, trăm/Nq, nghìn/Nq, vạn/Nq, triệu/Nq, mỗi/Nq, những {some}/Nq, mấy/Nq, các/Nq, vài/Nq, một_đôi/Nq, vài_ba/Nq, từng/Nq

3/Num triệu/Nq, hai/Num chục/Nq, 5/Num trăm/Nq

ba/Num mười/Nq, ba/Num mười/Nq phẩy/Nn năm/Num

- Quantifier with unit:
 - o The short forms are tagged as Nq.

Examples:

2%/Nq, 200.000/người/tháng/Nq, etc.

- For the full form, numbers are tagged as Num, the others are tagged according to their meaning.

Examples:

2/Num phần_trăm/Nq, 2/Num đô-la/Nu, 2/Num độ/Nu C/Nu

hai/Num phần_trăm/Nq, hai/Num đô-la/Nu, hai/Num độ/Nu C/Nu

200.000/người/tháng/Nq {VND200,000/person/month}

200/Num ngàn/người/tháng/Nu {VND200,000/person/month}

200/Num ngàn/Nq một/Num người/Nc một/Num tháng/Nt {VND200,000 per person per month}

20/Num lít/Nu mỗi/Nq ngày/Nt {20 liters a day}

- Quantity nouns with tmesis factors

For the quantity phrases indicating: portion, ratio, score, degree, interval, and fraction (the tmesis marks are usually a colon “:”, a hyphen “-”, plus “+”, division “÷”, multiplication “x or *”, equality “=”, slash “/”, ...), the shorts form are tagged as Nq, the full form are tagged individually, depending on word meaning.

Examples:

2:3/Nq, 2_: _3/Nq

2-3/Nq, 2_- _3/Nq

4_x_6/Nq, 4x6/Nq

11:2x3/Nq

11_: _2_x_3/Nq

3/4/Nq

ba/Num phần/Nn tư/Nq

mười/Num một/Num nhân/Nn hai/Num chia/Nn ba/Num

Note:

- The words *nhiều* and *ít* are tagged as Aa

Examples:

nhiều/Aa ý_kiến/Nn

nói/Vv nhiều/Aa

ít/Aa quà/Nn

làm/Vv ít/Aa

- The word *những* will be tagged as Nq when it precede a classifier noun, common noun or unit noun. However, it will be tagged as M when it precede a quantifier.

Examples:

những/Nq con/Nc cá/Nn

những/Nq cánh_đồng/Nn

những/Nq bó/Nu rau/Nn

những/M năm/Num con/Nc cá/Nn

- Distinguishing between Num and An:

Numbers indicating time points, periods of time, and quantities will be tagged as Num. However, ordinal number will be tagged as An.

Examples:

lúc 2/Num giờ/Nt {2 am}

tôi là người thứ 2/An nhận quà (I am the second person received gift)

hôm nay là thứ 2/Num {Today is Monday}

3.2.1.7. Proper nouns (Nr)

Nouns indicating proper names of people, animal, organizations, geographic places, things, and etc. are tagged as Nr.

Examples:

Smith/Nr, Thanh/Nr, John_Smith/Nr, Hồ_Chí_Minh/Nr, Vích-to_Huy-gô/Nr, Mickey/Nr, Jerry/Nr, Hà_Nội/Nr, Việt_Nam/Nr, Tokyo/Nr, Italia/Nr, Mê-hi-cô/Nr, Mát-xơ-va/Nr, Mao_Trạch_Đông/Nr, etc.

The numbers playing the role of names for geographic places/address are tagged as Nr

Examples:

Quận/Nun 3/Nr {District 3}

Q./Nun 3/Nr, Tp./Nun Hồ_Chí_Minh/Nr {District 3}

Phường 14/Nr {Ward 14}

P./Nun 14/Nr {Ward 14}

K55/12/49/Nr

488/12A/Nr

10-12/Nr đường/Nn Đình_Tiên_Hoàng/Nr

The numbers in proper names that indicate the orders are tagged as An

Examples:

Bệnh_viện/Nn Nhi_Đồng/Nn 1/An

Cầu/Nn Bình_Triệu/Nr 2/An

Công_ty/Nn Cienco/Nr 6/An

Quân_khu/ 7/An

The abbreviations of proper name are tagged Nr.

Examples:

U18/Nr, U50/Nr, G7/Nr, G20/Nr

Trường/Nn ĐH/Nn CNTT/Nr

Class names are tagged as Nr

Examples:

lớp/Nn 6A/Nr

lớp/Nn 12/4/Nr

All the telephone numbers, mobile phone, fax are tagged Nr.

Examples:

0902631209/Nr

083_7262368/Nr

E-mail address, website, nick name (chat) are tagged as Nr

Examples:

<http://uit.edu.vn>/Nr

xyz@yahoo.com/Nr

abc@gmail.com/Nr

cong_chua_pikachu/Nr

Names of the decrees, resolutions, laws, resolutions, agreements, document will be maintained and tagged Nr. (including the decree's name with the mark “/” or “-“)

Example:

Nghị_định/Nn 43/2009ND-CP/Nr

Chemical symbols/elements are tagged as Nr

Examples:

CH4/Nr, CO/Nr, HCl/Nr, Fe/Nr, Kali. Uranium, ...

A mathematical formula is a word and tagged as Nr

Examples:

2x/Nr

2x²/Nr

2x+3y/Nr

2x_+_3y/Nr

For the sequences of proper nouns and Sino factors, Sino factors are tagged according to their meaning, the proper nouns are tagged as Nr

Examples:

Hà/Nr phu_nhân/Nn

Hồ/Nr Chủ_tịch/Nn

The numbers, titles or descriptive factors that combines with proper names to form a popular names or nicknames (not a nickname to chat) will be annotated as common in part-of-speech tagging.

Examples:

- Titles:

Trạng/Nn Quỳnh/Nr

Nghè/Nn Tân/Nr

Tú/Nn Xương/Nr

Đồ/Nn Chiêu/Nr

Thủ_khoa/Nn Huân/Nr {First class Huan}

Bà/Nn Trung/Nr {Mrs Trung}

Cụ/Nn Phan/Nr {Mr Phan}

Bác/Nn Hồ/Nr {Uncle Ho}

Hồ/Nr Chủ_tịch/Nn {President Ho}

Minh_Tri/Nr Thiên_hoàng/Nn {Mikado Minh Tri}

Hoa_Nhi/Nr phu_nhân/Nn {Madame Hoa Nhi}

Nam_Phương/Nr Hoàng_hậu/Nn

Ngọc_Hân/Nr Công_chúa/Nn

Alexander/Nr Đại_đế/Nn

Nhụy_Kiều/Nr Tướng_quân/Nn

Từ_Hy/Nr Thái_hậu/Nn

Bát_Dị/Nr Hoà_thượng/Nn

Hà/Nr thủ_quỹ/Nn

Phương/Nr giám_đốc/Nn

- Numbers combining with proper names to form a popular name or nickname will be tagged as Nr

Ba_Giai/Nr ("Ba" is a number)

Tám_Danh/Nr ("Tám" is a number)

Hai_Nguyên/Nr ("Hai" is a number)

Một_Hiên/Nr {second lieutenant Hien} (In the pre-revolutionary period, when the numerals indicate the rank in the army (một {one} = thiếu úy, năm {five} = đại tá {colonel}, ...)

Năm_Tôn/Nr {Colonel Ton}

- Descriptive factors:

Năm/Nr Sài_Gòn/Nr {Nam Saigon}

Minh/Nr sẹo/Nn {Minh scar}

Xuân/Nr tóc/Nn đỏ/Aa {Xuan red hair}

Ba/Nr Cầu/Nn Bông/Nr {Ba Bong Bridge}

Ba/Nr chữ_kí/Nn {Ba signature}

Ốc/Nn Thanh_Vân/Nr {Thanh Van Snail}

Công_chúa/Nn bong_bóng/Nn Bảo_Thy/Nr

Caution:

For the expressions composed by common nouns used to name an organization, an object, ect., we tag the words according to their meaning.

Example:

báo/Nn Thanh_Niên/Nn

sao/Nn Hỏa/Nn

An expression of descriptive factors and proper noun used to name a place, a street, etc. will be treated as a word and tagged as Nr.

Example:

Đường/Nn Hai_Bà_Trung/Nr

3.2.1.8. Common nouns (Nn)

Nouns that do not belong to the above categories are tagged as Nn, including:

- Abstract and concrete objects/things, such as: *nhà {house}*, *trâu {buffalo}*, *muối {salt}*, *gia_đình {family}*, *lập_trường {stance}*, *thiên_đường {paradise}*, and etc.
- Single entities and collective objects/things (as for meaning), such as: *xe {bike/car}*, *vườn {garden}*, *đu_đu {payaya}*, *nhân_dân {people}*, *bàn_ghế {furniture}*, *giấy_tờ {papers/documents}*, and etc.
- Countable and uncountable objects/things (as for format), such as: *ghế {chair}*, *mũ {hat}*, *chim {bird}*, *tình_yêu {love}*, *hạnh_phúc {happiness}*, and etc.
- Chemical substances written in Vietnamese, for instance *các-bon/Nn {carbon}*, *ô-xi/Nn {oxygen}*, *ôxy/Nn {oxygen}*, *lưu_huỳnh/Nn {brimstone}*, and etc.
- Directional and positional words in context, such as: *đông/Nn {east}*, *tây/Nn {west}*, *nam/Nn {south}*, *bắc/Nn {north}*, *trung/Nn {central}*, *trên {above}*, *dưới {downside}*, *trong {inside}*, *ngoài {outside}*, *trước {front}*, *sau {behind}*, and etc.
- Nouns indicating objects, plants, animals and human, such as: *cá*, *hoa*, *xe*, *máy*, and etc. These nouns usually accompany with classifier nouns, such as: *cái*, *cây*, *con*, *người*, and etc.
- Nouns indicating solid, liquid and gas states, such as: *đá*, *nước*, *khói* which usually accompany with unit nouns, such as: *cục*, *ngụm*, *luồng*, and etc.
- Nouns indicating the weathers, such as: *bão*, *chớp*, *lũ*, *lụt*, *gió*, *mưa*, *nắng*, *sấm*, *chớp*, *sét*, *sấm_sét*, *sét*, *lũ_lụt*. These nouns usually accompany with unit nouns, such as: *ánh*, *con*, *trận*, and etc.

- Unit nouns of organization, such as: *khu {area}, lớp {class}, đoàn, đoàn_thể {union}, ban {department}, ngành {branch}, khoa {faculty}, trường {school}, viện {institute}, bộ {ministry}, công_ty {company}, bệnh_viện {hospital}, tập_đoàn {corporate group}, and etc.*
- Collective nouns: *bầy, bọn, đám, đàn {flock}, lũ, mớ, đội {team}, nhóm {group}, tổ {team/group}, tốp, đoàn {caravan/convoy/group/ team/troupe}, and etc.*

Note:

- A verb or adjective follows quantifier nouns to mention a single entity will be tagged as common noun.

Examples:

những/Nq trao_đổi/Nn {exchanges}

các/Nq bổ_nhiệm/Nn {appointments}

những/Nq đau_buồn/Nn

các/Nq khó_khăn/Nn

- A verb or adjective follows nothing:
 - o If we can add a special classifier noun before verb or adjective, the verb/adjective still maintains its POS.

Example:

(Việc) Báo_cáo/Vv của cô ấy rất tốt.

- o If we can add a classifier noun before verb or adjective, the verb/adjective will be tagged as common noun.

Example:

(Cuốn) Báo_cáo/Nn của khoa đã được nộp cho phòng Tổ chức – Hành chính.

- o If we can add a quantifier noun before verb or adjective, the verb/adjective will be tagged as common noun.

Example:

(Những) Nghiên_cứu/Nn này rất hấp dẫn {this(these) research(es) is very interesting}

(Những) Khảo_sát/Nn của Ciem cho thấy rằng

- If we can add either quantifier noun or special classifier noun before verb or adjective. If verb/adjective plays the role as object, we tag it as common noun. Otherwise, the verb/adjective will be tagged with a double POS tag.
- If we can add the preposition *của_{of}* before the verb, the verb will be tagged as Nn.

Example:

Kỷ_niệm/Nn 60 năm chiến_thắng/Nn lịch_sử Điện_Biên_Phủ

Kỷ_niệm/Nn 60 năm của chiến_thắng/Nn lịch_sử Điện_Biên_Phủ

- Nouns indicating kinship relation such as *cô*{*aunt*}, *dì*{*aunt*}, *chú*{*uncle*}, *bác*{*uncle*} will be tagged Nn when they come before proper name of person.

Examples:

cô/Nn Hoà/Nr

dì/Nn Nguyệt/Nr

- Distinguish between Nn and Nc: based on role of identification of words:
 - The words that be tagged as Nn be used to name the objects, phenomena, etc.

Examples: *xe, ghế, gà, kiến, etc.*

- The words that be tagged as Nc do not have the role of identification. They are not used to name the objects, phenomena, etc.

Examples: *cái, con, quyển, etc.*

3.2.2. Verb

Verbs are classified into 5 categories, exist verb (Ve), copula “là” verb (Vc), co-verb (D), comparative verb (Vcp), and other verb (Vv). Annotation instruction of each category is given below.

3.2.2.1. Existing verb (Ve)

The existing sentences which indicate existing state of things are begun with verb phrases. Head words of these verb phrases are the existing verbs such as: *có* {*there are/is*}, *còn* {*remain*}, *còn_lại* {*remain*}, *hết* {*end*}, *kết_thúc* {*end*}, etc. These verbs are tagged as Ve.

Examples:

Có/Ve 2 cuốn sách trên bàn.

Còn/Ve hai con cá trong chậu.

Còn_lại/Ve hai con cá trong chậu.

Hết/Ve phim.

Kết_thúc/Ve phim.

Note:

- There are the existing sentences that are also begun with the verb phrases. However, the head words of these verb phrases are not the existing verbs. We also tag these verbs as *Ve*.

Example:

Đến/Ve hè rồi

- There are the existing sentences that the verb phrases are the predicates, and the subjects are noun phrases. In these cases, the existing verbs that are head words of the verb phrases are also tagged as *Ve*.

Examples:

Tiền *còn/Ve* trong túi.

Túi *còn/Ve* tiền.

- Besides indicating state of things, the words *có* {have}, *còn* {remain}, *hết* {end}, and etc. can be used in other contexts as a conjunction, adjunct, and etc.

Examples:

ngoại_tê *có/Rv* biến_động/Vv

Em *có/Rv* biết/Vv

Ngoài_ra/Cp *còn/Cp*

Không lường *hết/R*

3.2.2.2. Copula “là” verb (*Vc*)

The special verb “là” {to be} comes before object of a sentence to express that subject and object of the sentence are equal. Where, the subject and the object have the same phrase type.

là will be tagged as *Vc*.

Examples:

(NP Hà_Nội) *là/Vc* (NP thủ_đô của nước Việt_Nam). {Hanoi is the capital of Vietnam}

Note:

- When “là” is not the main verb and follows another verb, it is used as a conjunction and should be tagged as Cs. In this case, *là* can be replaced by *rằng*. However, repetition form of “là” which is modifier of main verb will be tagged as Aa.

Examples:

Tôi biết/Vv *là*/Cs bạn sẽ đến. {I know *that* you will come.}

Ai cũng khen/Vv *là*/Cs giỏi.

Chim bay *là* *là*/Aa trên cách đồng

- Distinguishing between Vc and Vv: The word “là” which means “to iron” or “moving from higher level to a lower plane” will be tagged as Vv. The repetition form of “là” which main verb be also tagged as Vv.

Examples:

Tôi đã *là*/Vv quần áo rồi.

Cành liễu *là*/Vv xuống mặt nước.

Đàn chim *là* *là*/Vv trên cánh đồng.

- Distinguishing between Vc and M:

Examples:

(NP Cô ấy) (ADJP rất *là*/M đẹp). => we can understand the sentence if the *là* is deleted. *là* is used to emphasize.

- The equal relationship that copula “là” expresses can be the relationship between noun phrase and sentence.

Examples:

(NP Điều đáng lưu ý) *là*/Vc (S những chương trình này đang hoạt động một cách tự động).

- *Là* is tagged as Cp in the structure *hễ ... là*

Examples:

Hễ/Cp có lệnh *là*/Cp đi

3.2.2.3. Directional verbs (D)

When the directional verbs, such as: *bằng* {across}, *đến* {arrive/forward}, *lại* {back}, *lên* {up}, *lui* {back}, *lùi* {back}, *qua* {across}, *ra* {out}, *ra đi* {leave}, *ra vào*, *sang* {across}, *tới* {arrive/forward}, *trở* {turn}, *vào* {in}, *về* {back}, *vô* {in}, *xuống* {down}, follow transitive

motion verbs such as: *mang* {carry}, *cắt* {cut}, *vác* {carry}, *bẻ* {fold}, and etc. or intransitive motion verbs such as: *đi* {go}, *nhảy* {jump}, *chạy* {run}, *trèo*, and etc. which this combination shows movement direction (the motion verbs themselves do not indicate the direction), they will be tagged as D.

Examples :

Tôi đi/Vv vào/D nhà {I enter the house}

đi qua/D đi lại/D

đi tới/D đi lui/D

There are some cases that we can remove the intransitive motion verb and use the directional verb as a replacement. When the directional verbs replace the intransitive verb, they will be tagged as Vv.

Examples:

Tôi đi/Vv lên/D lớp → Tôi lên/Vv lớp

Tôi đi/Vv ra/D vườn → Tôi ra/Vv vườn

Hôm_nay giá vàng lên/Vv {Today gold price up}

There are the motion verbs (head verbs) that indicate the direction by themselves. However, these motion verbs are always modified by the directional verbs. The directions that the motion verb and its modifier show are similar. However, if we remove the directional verbs, the sentence will not mean. In these cases, the directional verbs are also tagged as D.

Examples:

ngắm/Vv vào/D đa

xuất_khẩu/Vv phần_mềm của Việt_Nam sang/D Nhật

vứt/Vv rác xuống/D sông

Note:

- If the directional verbs or non-motion verbs have the abstract sense, the directional verbs will be tagged as adjunct R.

Examples:

Cuộc sống mang/Vv lại/R cho tôi nhiều niềm vui {Life brings back to me many joys}

Tôi đã nhận/Vv ra/R ý nghĩa của cuộc sống {I figured out the meaning of life}

Sản_xuất/Vv ra/R năng_lượng

Nghĩ/Vv ra/R hướng giải_quyết

- The words *vào, lên, đến, ...* can also be used in non-physic directional contexts, e.g to indicate time or to come with other non-motion verbs, in these cases, it is tagged as Cs

Examples:

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

chuyên/Vv vào/Cs ngành toán {mathematic domain}

dự_thi/Vv khối A vào/Cs các trường ĐH

3.2.2.4. *Comparative verb (Vcp) ([6], page 85)*

The comparative sentences usually express the comparison between 2 objects or compare an object with others. The comparative indicators such as *hơn, thua, kém, bằng, ngang, tương_đương, như, etc.* are used in these sentences to express the comparison.

- Comparative indicators are tagged Vcp if it is head of predicate to describe the comparison between 2 objects (2 objects have the same type).

Examples:

Tôi bằng/Vcp anh

Tôi hơn/Vcp anh

Tôi thua/Vcp anh

Tôi ngang/Vcp anh

Điểm của tôi như/Vcp anh.

- Comparative indicators are tagged Vcp if it modifies an adjective to describe the comparison between 2 objects.

Examples:

Tôi cao/Aa hơn/Vcp anh 2 cm

Tôi cao/Aa hơn/Vcp tất_cả/Pp

Tôi cao/Aa thua/Vcp anh 2 cm

Tôi cao/Aa kém/Vcp anh 2 cm

Tôi cao/Aa bằng/Vcp anh

Tôi và anh cao/Aa bằng/Vcp nhau

Tôi cao/Aa *ngang/Vcp* anh

- In cases that object follows the comparative indicator is omitted, however, we can insert it into the sentence, the comparative indicator is also tagged as Vcp.

Examples:

Có sức khỏe vẫn *hơn/Vcp*

Bức tranh này đẹp *hơn/Vcp*

Note:

- The word “*như*” and “*nhất*” are also used to indicate the comparison. However, “*như*” is tagged as Vcp when it is head of predicate. Otherwise, *như* is tagged as Cs. The word “*nhất*” will be is tagged as R if it indicate the most quality (modify an adjective, verb). To indicate the first order, the word “*nhất*” ia tagged as An.

Examples:

Tôi cao/Aa *như/Cs* anh

Tôi cao *nhất/R*

Tôi chạy nhanh *nhất/R*

Tôi đứng hạng *nhất/An*

Cần *nhất/R* là sức khỏe

Tôi chạy về *nhất/An*

- The comparative words are tagged as R when they are preceded the quantifier nouns.

Example:

hơn/R 200 triệu đồng { more than 200 million Dong }

- The comparative words are tagged as Aa when they do not express the comparison.

Example:

Tôi học *kém/Aa*

Sức khỏe tôi *kém/Aa*

Bây giờ là bảy giờ *kém/Aa*

Quốc gia *kém/Aa* phát triển

Mặt gương rất *bằng/Aa*

Máy bay cánh *bằng/Aa*

“anh”, “chàng” là những tiếng *bằng/Aa*

- *hơn* is tagged as R when it modifies degree for adjective/verb. In this case, we can not insert the comparative object after *hơn*

Example:

Tiến độ *ì_ạch/Aa* *hơn/R*

Xe đạp giúp tôi đi nhanh *hơn/R* và đỡ mệt *hơn/R*

Hiểu nhau *hơn/R*

3.2.2.5. Other verbs (Vv)

Verbs do not belong to the above categories will be tagged Vv, including:

- Transitive, such as: *ăn {eat}*, *đọc {read}*, *xây_dựng {construct}*, *cải_tiến*, *cuốc*, *đánh*, *viết*, and *etc.*
- Intransitive, such as: *ngủ*, *tắm*, *cười*, *chạy*, *bay*, *làm_lụng*, *ngủ_ngoi*, and *etc.*
- Verbs indicating thinking, feeling, emotion, etc. also belongs to this type, such as: *cảm_thấy {feel}*, *nhận_thấy {realize}*, *nghe*, *biết*, *tin*, *yêu*, *nhớ*, *ngghi_ngờ*, *hiểu*, and *ect.*
- Transformative verbs such as: *trở_nên*, *nên {become}*, *trở_thành*, *hoá*, *thành*, and *etc.*

Examples:

Nên/Vv vợ *nên* chồng

Nói không *nên/Vv* lời

Trở_nên/Vv thân_thiết

- Volition verbs: *muốn*, *quyết*, *dám*, *toan*, *định*, and *etc.*
- Modality verb: *cần*, *nên {should/had better}*, *phải*

Examples:

Bạn *cần/Vv* có kiên thức

Tôi *cần/Vv* anh

Nên/Vv khiêm_tốn

Bạn *phải/Vv* hoàn_thành việc này đúng thời_hạn

- The word *gây*

Examples:

Gây/Vv cười

Trồng cây gây/Vv rừng

Gây/Vv khó khăn cho công việc

Note:

- When a verb is a head of the subject of a clause/sentence, we base on (or insert) its modifier to determine its POS
 - o Noun can be modified by a quantifier, a demonstrative pronoun or a possessive prepositional phrase

Examples:

những lựa_chọn/Nn

các hỗ_trợ/Nn

lựa_chọn/Nn này

lựa_chọn/Nn của tôi

Note: If a verb is modified by a demonstrative pronoun or a possessive prepositional phrase (verb is not a modifier of Ncs), we tag it as Nn. Do not use the rule of inserting the quantifier, the classifier, etc.

- o Verbs following special classifier nouns (Ncs)

Examples:

sự hi_sinh/Vv

- A word in possessive form should be tagged as common noun:

Examples:

nguyên_tắc cơ_bản của giáo_dục/Nn

- When POS of a verb can not be determined on the basis of its modifier or surrounding text, if it is head of subject, we label it with Nn

Examples:

liên_hoan/Nn đã trải_qua chín kì

3.2.3. Adjectives

Adjectives are classified into 2 categories, ordinal number (An) and other adjectives (Aa). Annotation instruction of each category is given below.

3.2.3.1. Ordinal number (An)

Ordinal numbers that indicate the order of things such as: *nhất {first}, nhì {second}, đầu_tiên {first}*, are tagged as An. For sequences of the word “*thứ*” and numbers that indicate the order of things, such as *thứ hai, thứ nhì, thứ ba {third}, thứ tư, thứ bốn {fourth}, and etc.*, the words *thứ* are tagged as Nn, the numbers are tagged as An.

Examples:

Tôi là người *đầu_tiên*/An nộp bài

Người mua hàng *thứ/Nn một/An ngàn/An (thứ/Nn 1000/An)* sẽ được tặng quà

In specific cases, such as: *lớp 1 {grade 1}, lớp 7 {grade 7}, cấp 2 {secondary level}, cấp 3 {high school level}, hạng 1, hạng 2, ...* the numbers are also tagged An.

Note:

- Distinguishing between Nr and An: a proper name which is number will be tagged as Nr. However, ordinal number will be tagged as An.

Examples:

Tôi học lớp 7A/Nr (proper name is number)

Tôi học lớp 7/An (ordinal number)

- In a proper name including ordinal number, the ordinal number is tagged as An.

Examples:

Vua/Nn Henry/Nr II/An

Cầu/Nn Bình_Triệu/Nr II/An

- Distinguishing between Num and An: for weekdays (*thứ hai, thứ 3, and etc.*), the word *thứ* will be tagged as Nt, numbers tagged as Num.

Examples:

Thứ/Nt hai/Num

Thứ/Nt 2/Num

3.2.3.2. Other adjectives (Aa)

Adjectives do not indicate the order will be tagged as Aa, including:

- Quality adjectives such as: *tốt {good}, đẹp {nice}, giỏi {good}, xấu, ngoan, thông_minh, khôn, tích_cực, trong_sạch, bản, ngu_xuẩn, and etc.*
- Quantity adjectives such as: *cao, thấp, ngắn, dài, rộng, hẹp, xa, gần, nông, sâu, and etc.*

Note:

Noted about the transition of POS:

- Any words (regardless their original POS) standing behind “*rất*”, which refer to theirs property, are transferred to adjectives.

Examples:

rất Việt_Nam/Aa

rất Huế/Aa

- Some verbs following *rất, quá, hơi, hết_sức, vô_cùng, thật, khá, cực_kì, tương_đối, một_cách* or preceding *hơn, cực_kì, nhất, quá, lắm, hết_sức, vô_cùng, thật* are possible to be transferred to adjectives.

Examples:

thật hài_hước/Aa

khá kiên_trì/Aa

ổn_định/Aa hơn

có_chấp/Aa quá

Exception:

In the sequences of *rất, quá, lắm, etc.* and the verb indicating thinking and feeling (emotion verb), these verbs are not tagged as adjectives.

Examples:

Rất nghe/Vv lời

Cực_kì hiểu/Vv chuyện

Yêu/Vv lắm

rất thương/Vv (ghét, kính, nể, vui, buồn,...)

3.2.4. Pronouns

Pronouns are classified in two types: demonstrative pronouns (Pd) and others (Pp).

3.2.4.1. Demonstrative pronouns (Pd)

Demonstrative pronouns are pronouns modifying nouns (Nn) to point things and persons that nouns refer. They often indicate temporal and spatial meaning such as: *ấy, đây, đấy, đó, kia, này, nọ, nay, etc.*

Examples:

quyền sách *này/Pd*

đây/Pd là

3.2.4.1.1. Other pronouns (Pp)

The following pronouns that are head of NP are tagged as Pp:

- Personal pronouns (không phân biệt giới tính) denoting objects/things, such as: *tôi {I}, tao, ta, chúng_tôi, chúng_ta {we}, mày {you}, chúng_mày, nó {it}, chúng_nó, hắn {he/she}, y, and etc.*
- Pronouns denoting quantity, such as: *bấy_nhiều*
- Pronoun denoting quality and action, such as: *thế, vậy*
- Interrogative pronouns, such as: *ai, gì, bao_giờ, đâu, bao_nhiều, mấy, thế_nào, and etc.*
- *Này, kia, đó, etc.* play the role as head word of noun phrases.
- Words denoting whole, for examples *tất_cả, hết_thấy, toàn_bộ, etc.* will be tagged as Pp when they are head of NP.

Examples:

Tôi ăn được *tất_cả/Pp*.

Note: Pp is not modified by a Pd.

3.2.5. Adjuncts

Adjuncts are used to modify verbs, adjectives, nouns, pronouns and etc. They can come before or follow the words they modify to denote the capacity, ability, passive, increase, decrease, degree, manner or abstract motion. Lists of adjuncts is following:

- Adjuncts denoting time (or tense): *đã (for past tense), đang (for present tense), sẽ (for future tense), vừa, mới, sắp, từng, liên, bèn, rồi and etc.*

Examples: *đã đi, sắp đi*

- Adjuncts denoting degree: *rất {very}, quá {much}, khá, khi, hơi, thật, hoàn_toàn, hết_sức, tuyệt_đối, cực, vô_cùng, lắm, and etc.*

Examples: *quá tốt, hơi xấu*

- Comparative adjuncts denoting the homogeneity of actions, states, quality and etc.: *cũng, đều, vẫn, cứ, còn, mãi, mãi_mãi, luôn, luôn_luôn, hoài, liên_tục, liên_tiếp, and etc.*

Examples:

Tôi học giỏi, bạn cũng học giỏi

- Adjuncts with negative denotations: *chưa* {yet}, *chẳng* {not}, *không* {not}, *không_thể* {can not}, *chưa_thể* {can not/ not...yet}, *chẳng_thể* {can not}, and ect.

Examples:

Tôi chưa/R {did} nghe/Vv {hear} qua

- Adjuncts with positive denotations: *có* {do/did}

Examples:

Tôi có/R nghe/Vv qua

Nó có/R nói_dối/Vv

Cô ấy có/R yêu/Vv John thật đấy

- Adjuncts denoting the commands, advices, invitations, forbiddances: *hãy*, *chớ*, *đừng*.

Examples:

Bạn đừng đi, hãy ở lại cùng tôi

Adjuncts are tagged as R.

- Adjuncts modify adjectives.

Examples: *Rất/R đẹp/Aa*, *khá/R đẹp/Aa*, *đẹp/Aa lắm/R*, *đẹp/Aa quá/R*, *đang/R rảnh/Aa*, *vẫn/R minh_mẫn/Aa*, etc.

- Adjuncts modify verbs.

Examples:

đã/R đi/Vv, *sẽ/R đi/Vv*, *rất/R lo_lắng/Vv*, *không/R làm/Vv*, *vẫn/R thức/Vv*, *nói/Vv được/R bốn thứ tiếng*, *sẽ/R được/R trình_bày/Vv*, etc.

Năm nào Mai và Lan cũng/R đều/R học/Vv giỏi/Aa.

- Adjuncts modify nouns or pronouns.

Examples: *không/R một/Nq bóng/Nn cây/Nn*, *bất_kì/R sản_phẩm/Nn*, *vẫn/R cứ/R chúng/Nn nào/Pp tất/Nn nấy/Pd*, etc.

Note:

The adjectives modifying quantifier nouns will be tagged as R or Aa based on the principles as follows:

- If deleting adjective do not affect understanding the sentence, the adjective will be tagged as R (Nếu tính từ bỏ nghĩa cho số thì nó được gán nhãn R. Ngược lại, ngữ danh từ đứng phía sau tính từ bỏ nghĩa cho tính từ thì tính từ được gán nhãn Aa)

Examples:

Tôi có mặt *đúng*/R 6 giờ tối

Tối_thiếu/R 25 ngàn \$

- If we cannot delete adjective, the adjective will be tagged as Aa.

Examples:

Hàng rào *cao*/Aa

Hàng rào *cao*/Aa 2 mét

➔ if we delete the word *cao*, we will not know that the fence is 2 meters high or wide.

- Other adjuncts.

Examples:

Cũng/R *bằng*/Cs 2 bàn tay này

Chỉ/R vì/Cs trời mưa mà/Cs tôi đi học trễ

Note:

- Distinguishing between adjuncts and M:
 - M are used to indicate emotion of speakers or writers and to emphasize meaning of the words that they modify.

Examples:

Bức tranh đẹp/Aa *thật*/M

Tôi đã nhận được/M *tiền*

- Adjuncts modify adjectives, verbs, etc. about time, tenses, passive voice, and etc. They can increase or decrease meaning of adjectives.

Examples:

Bức tranh thật/R *đẹp*/Aa {very beautiful}

Tôi được/R *nhận*/Vv *tiền* {am received}

Bác sĩ trị/Vv *được*/R *bệnh*

Bài toán này tôi làm được/R

- *đã* is tagged as M if it is not used to denote pass tense. In this case, *đã* is used to emphasize something.

Examples:

Nhà ấy đã/M giàu lắm rồi

However, *đã* is tagged as R in pair of conjunction *đã ... lại*

Examples:

Cô ấy đã/R đẹp lại còn thông minh

- *đã* is tagged as M if it modifies a noun.

Examples:

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

- Distinguishing between adjuncts and D.

- Adjuncts are used to emphasize directional meaning, the words which they modify have directional meaning.

Examples: *tăng lên/R*, *giảm xuống/R*, *bán ra/R*, *mua vào/R*, *gây/Vv ra/R* {to cause}, *nhận/Vv ra/R* {to recognize}, *trả/Vv lại/R* {to give back}, *nhận/Vv được/R*, and etc.

- Directional co-verbs are the modifiers of the verbs which do not have directional meaning.

Examples: *đi lên/D lâu*, *đi xuống/D đất*, and etc.

- Distinguishing between adjuncts and Cs

Adjuncts are not modified by a lexical modifier. Meanwhile, Cs is always modified by a noun phrase, a verb phrase, an adjective phrase, etc.

3.2.6. Conjunction

Conjunctions are classified into 2 types. Prepositions and conjunctions introducing relative clauses are annotated with Cs. Other conjunctions are annotated with Cp.

3.2.6.1. Prepositions and conjunctions introducing relative clauses (Cs)

The following word types are tagged as Cs:

- Words, such as *ở*, *tại* mentioning things to exist or to take place are usually modified by a noun or noun phrase.

Examples:

Tôi để cuốn sách *ở*/Cs trên bàn

Chúng ta sẽ gặp *tại*/Cs công viên

- *Trên, dưới, trong, ngoài, giữa, trước, or so on:*

Tagged as Cs if it follows a directional word and precede a noun indicating the target of action. Otherwise, tagged as Nn if it precedes nothing.

Examples:

Trèo lên *trên*/Nn

Trèo lên *trên*/Cs núi

Tagged as R if it precede quantifier

Examples:

Trên/R 500 người

Tagged as Cs if it precede a noun in circumstant constituent

Ngoài/Cs vườn, hoa nở

Hoa nở *ngoài*/Cs vườn

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

Example:

nhà *bằng*/Cs gạch {brick house}

- The possession (*của {of}*)

Example:

sách *của*/Cs tôi {my book}

- The purpose (*để {to}*)

Example:

viện có *để*/Cs từ chối {find a reason to deny}

- The interval (in the construct *from...to*), words like *từ {from}*, *đến*, *tới*, *lên {to}* are tagged Cs.

Example:

Từ/Cs ngày 2 đến/Cs ngày 3

Từ/Cs Hà Nội đến/Cs Tp. Hồ Chí Minh

- Enumeration (như {as})

Example:

Các tài sản cá nhân *như/Cs* vàng, bạc, đất đai

- The word *với* {with}

Example:

Với/Cs một cây súng và 2 viên đạn

Tôi làm việc này *với/Cs* cả tấm lòng

- The word *mà, là, rằng, ...* beginning a subordinate clause (SBAR)

Example:

Cuốn sách *mà/Cs* tôi mua hôm qua rất hay

- The word *qua* indicating position of a thing

Example:

Đoạn đường *qua/Cs* thị trấn Đông Hà

- The word *cho* which is head of a post-modifier of a verb.

Example:

Tặng hoa *cho/Cs* người yêu

Mẹ mong *cho/Cs* con học giỏi

Note:

- The words such as: *trong, ngoài, trên, dưới and ect.* which follow the words such as: *bên, mặt, phía, đằng, vòng* are tagged as Nn

Examples:

phía/Nn trên/Nn {above}

bên dưới/Nn {downside}

mặt/Nn trên/Nn {inside}

đằng/Nn ngoài/Nn {outside}

vòng/Nn trong/Nn

- The words such as: *trong, ngoài, trên, dưới, ect.* which indicate the places are tagged as Nn. In these cases, we can add a preposition before them.

Examples:

trong/Nn này

ngoài/Nn kia

bài hát *trên/Nn*

3.2.6.2. *Other conjunction (Cp)*

The conjunctions which are not subordinating conjunctions will be tagged as Cp. There are word types as follows:

- The words, such as: *và, với, hay, hay là, hoặc, hoặc là, còn* are used to link the words, phrases, or clauses which have equal relationship and the same function in the sentence.

Examples:

Tôi *và/Cp* bạn đều giỏi

Tôi *với/Cp* anh cùng làm

Tôi *hay/Cp* anh làm việc này

Tôi *hoặc/Cp* bạn sẽ làm việc này

Tôi, bạn *và/Cp* An được phân công làm việc này

Bố đi làm *còn/Cp* mẹ đi chợ

Cô ấy *vừa/Cp* đẹp *vừa/Cp* giỏi *vừa/Cp* hát hay

- The words or pairs of words, such as: *không những ... mà còn, không chỉ ... mà còn, vừa ... vừa, càng ... càng, and etc.*, are used to link the words, phrases, or sentences. This linking express increased relationship.

Example:

Nó đẹp *còn/Cp* giỏi.

Gừng *càng/Cp* già *càng/Cp* cay

- The words, such as: *rồi, mới, and etc.* are used to express serial relationships, an incident follows other incidents

Examples:

Mây tan *rồi/Cp* mưa tạnh.

Phải tốt nghiệp *mới/Cp* đi làm.

Tôi vừa về đến nhà *thì/Cp* trời mưa.

Tôi đến nhà *thì/Cp* anh ấy đi vắng.

- The locutions having the linking function (thành phần chuyển tiếp), such as: *hơn_nữa {besides}*, *tóm_lại {in short}*, *tựu_trung {in general}*, *trái_lại*, and *etc.* are also tagged Cp.

Examples:

Cô ấy đẹp. *Hơn_nữa/Cp*, cô ấy lại thông minh.

Công việc này quan trọng, *hơn_nữa/Cp* lại là việc gấp.

Vấn đề này, *trái_lại/Cp*, có tầm quan trọng lớn.

- The words or word pairs to express the main-subordinate relationship: *nhờ*, *miễn*, *nên*, *cho nên*, *do ... mà*, *vì ... nên (cho nên)*, *vì ... mà, nếu ... thì*, *sở dĩ ... là vì*, *miễn ... thì*, *hễ ... thì*, *nhờ ... mà*, and *etc.*

Nhờ/Cp bác sĩ tận tình cứu chữa *mà/Cp* tôi mới khỏe mạnh

Tôi khỏe *nhờ/Cp* tập thể dục thường xuyên

Có_thể *vì/Cp* mưa lớn, *vì/Cp* mất người *nên/Cp* Thể_Công thua trận. (Maybe because it was raining heavily and the team lost its player, The Cong failed the game.)

Vi/Cp trời mưa, tôi đi học trễ

Tôi đi học trễ *vì/Cp* trời mưa

Thầy giáo bệnh *nên/Cp* tôi nghỉ học

Hễ/Cp bị mắng *là/Cp* nó bỏ nhà đi.

- The words or word pairs to express the contrary relationship: *song*, *nhưng*, *mà*, *chứ*, *thà ... chứ*, *tuy ... mà*, *tuy ... nhưng*, and *etc.*

Tuổi nhỏ, *song/Cp* chí lớn.

Tôi muốn đi làm *nhưng/Cp* tôi chưa tìm được việc

Tôi muốn trả công cho bạn *nhưng/Cp* có lẽ không có gì xứng đáng với việc làm của bạn.

- The word *thì* to indicate the homogeneous/comparative relation

Phụ nữ Việt Nam *thì/Cp* trung hậu đảm đang.

Tôi đến nơi *thì/Cp* mọi việc đã được giải quyết.

Note: Distinguish Cp and R for càng ... càng

When the pair càng ... càng is used to link 2 clauses of a compound sentence, we annotate them as R.

Ông ấy càng/R nói, người ta càng/R làm ồn.

3.2.7. Double POS

There are the words that have little consensus in community for classifying them. For example, some people treat buồn as an adjective, some other people treat buồn as a verb, and the others accept both, verb and adjective. The words like that are annotated with double-POS tag.

3.2.7.1. Verb-adjective (VA)

The words *buồn/VA*, *vui/VA*, *chán/VA*, etc. that express state of mind will be tagged as VA.

Examples:

Tôi rất vui/VA vì bạn đã hoàn thành công việc sớm hơn dự kiến.

Caution:

buồn which means want will be annotated as Vv.

Examples:

Tôi buồn/Vv ngủ.

Tôi chẳng buồn/Vv làm việc.

3.2.7.2. Verb-noun (VN)

The words, such as *nghĩ_vấn* in following example that are treated as noun or verb by linguists will be annotated with VN label.

Example:

Mọi chi tiết của vụ mưu sát còn nghĩ_vấn/VN.

However, if a word follows a quantifier or preceded a possessive prepositional phrase (của ...), it will be treated as a noun (Nn).

Example:

Mọi chi tiết của vụ mưu sát còn nhiều nghĩ_vấn/Nn.

3.2.7.3. Noun-adjective (NA)

The words, such as *an_ninh* in following example that are treated as noun or adjective by linguists will be annotated with NA label.

Example:

Cơ quan an ninh/NA.

3.2.8. Abbreviations/acronym

Abbreviation/acronym is tagged as follows.

- Abbreviation of an expression including proper name is tagged separately as its full form.
- Abbreviation of an expression including common nouns will be tagged as Nn.
- Abbreviation of an expression including more than one word type will be tagged as X.

Examples:

TP./Nun HCM/Nr {Ho Chi Minh city}

TP/Nun HCM/Nr {Ho Chi Minh city}

ASEAN/Nr (Hiệp hội các nước Đông Nam Á) {Association of Southeast Asian Nations}

WTO/Nr (Tổ chức Thương mại Thế giới) {World Trade Organization}

Trường/Nn ĐH/Nn CNTT/Nr (Công nghệ Thông tin) {IT - Information Technology}

TNHH/X (trách nhiệm hữu hạn) {Ltd - limited liability}

GS./Nn TS./Nn (Giáo sư Tiến sĩ) {Prof. Dr. - Professor.Doctor}

ĐHQG/Nn TP./Nun HCM/Nr (Đại học Quốc gia thành phố Hồ Chí Minh) {VNU-HCM - Vietnam National University in Ho Chi Minh City}

ĐH/Nn KHXH&NV/Nr (Đại học Khoa học Xã hội và Nhân văn) {USSH - University of Social Sciences and Humanities}

TTXVN/Nr (Thông tấn xã Việt Nam) {VNA - Vietnamese News Agency}

UBND/Nn TP./Nun HCM/Nr (Ủy ban Nhân dân Thành phố Hồ Chí Minh) {HCMCPC - Ho Chi Minh City People's Committee}

Ông/Nn N.V.N/Nr

Bà/Nn L.T.M_Lệ/Nr

3.2.9. Onomatopoeia (ON)

Words describing real sound (onomatopoeia) are tagged as ON

Examples: *tin/ON, toét/ON, bôm_bốp/ON, ùng/ON ùng/ON, etc.*

3.2.10. Punctuations (PU)

- All punctuations in the sentence, which appear alone, will be annotated as PU.

Examples:

?/PU, ./PU, ./PU, .../PU, //PU, {/PU, }/PU, etc.

Cô ấy đẹp ./PU giỏi ./PU thông_minh ./PU .../PU

“/PU Em giỏi lắm ”/PU :/PU thầy giáo khen tôi ./PU

- When the punctuation is written in letters, not symbols, it is tagged as Nn.

Examples:

dấu/Nn chấmNn, dấu_than/Nn, dấu/Nn chấm_than/Nn, dấu/Nn hỏi/Nn,
dấu/Nn chấm_hỏi/Nn, dấu_lửng/Nn, dấu/Nn chấm_lửng/Nn, dấu/Nn
huyền/Nn, dấu/Nn nặng/Nn, dấu/Nn sắc/Nn, dấu/Nn hai_chấm/Nn, dấu/Nn
ba_chấm/Nn, etc.

3.2.11. Idioms (ID)

Fixed combination of words have the meaning which often cannot be interpreted simply by the words forming them. Only when the idiom is put in appropriate contexts, it will express the clearest meaning. We consider this combination as a word and tag as ID.

Examples:

Hiền lành *rán_sành_ra_mỡ*/ID {to skin a flea for its hide}

Chữ nghĩa đã *đội_nón_ra_đi*/ID

Idioms are listed in our word segmentation guidelines.

3.2.12. Exclamation words (E)

The words that show emotional reactions (joy, complaint, curse, insult) of the speaker/writer or be used to call and respond, such as: *a, à, ái, á, ơ, ô_hay, này, ơi, trời, ồ* will be tagged as E.

Examples:

A/E , mẹ đã về!

À/E , tôi quên nói với bạn

Ái/E , đau quá!

Ơ/E , bạn cũng ở đây à!

Ô_hay/E , sao lại làm thế!

Này/E , sao lại làm thế?

Trời/E , chuyện gì vậy!

Ồ/E , tuyệt thật!

Anh ơi/E !

Con ở nhà nhé! *Dạ/E* !

3.2.13. Modifier (M)

The words, such as: *a, à, ư, nhỉ, hử, hề, không hề, chính, chính là, ngay, ngay_cả, nhé, vậy, đó, đấy, đâu, nào, đây, cả, thôi, nha, ạ, cái, sao, and etc.*, are tagged as M when they are used as follows:

- Insert to declarative sentences to create interrogative sentences, causative sentences, exclamation sentences.

Examples:

Mới đó mà quên rồi *à/M* ?

Con trông em *nhé/M* .

Chờ em tí *nha/M* !

Sao có chuyện lạ *vậy/M* !

Việc này nhờ anh *vậy/M* !

- Express the speaker's emotional nuances, including surprised, suspicious, ironic, etc.

Examples:

Quê ông ấy ở đâu *ta/M* ?

Điệu nhạc *sao/M* nghe buồn quá!

- Emphasize something và thể hiện thái độ của người nói.

Examples:

Cái/M con mèo trắng là của tôi.

Suốt mấy ngày *trời/M*.

Nó tàn nhẫn *ngay_cả/M* với mẹ nó.

Tiếng trống nghe điếc *cả/M* tai.

Note: *là* in the cases that are similar to the following example is annotated as M.

Kết quả là/M, các yếu tố ấy có thể có hại cho sức khỏe.

3.2.14. Foreign words (FW)

- Words adopted from foreign languages like *web, email, internet, and etc.* are tagged FW. But for foreign proper names like *Mary, Tom....*, they are still tagged Nr (proper name).
- Foreign word terminology or abbreviation are tagged FW: *FTSE/FW, TOEIC/FW, TOEFLT/FW, and etc.*

3.2.15. Unidentified POS (X)

Words which do not belong to any of the above categories will be tagged as X.

Examples: *hic/X, @/X, hehe/X, and etc.*

4. Internal structure tags

4.1. Brief description of W (word) and SWSP (supra-word sub-phrase)

Group	Tag	Meaning	Description
W	Sino-Vietnamese	Nn_w	Common noun A combination of a Sino-Vietnamese and the word it modifies to create a noun
		Vv_w	Verb A combination of a Sino-Vietnamese and the word it modifies to create a verb
		Aa_w	Adjective A combination of a Sino-Vietnamese and the word it modifies to create a adjective
		R_w	Adjuncts A combination of a Sino-Vietnamese and the word it modifies to create a adjunct
SWSP	Classifier noun	Nn_swsp	Noun sub-phrase A sequence of a classifier noun (Nc) and its modifier (noun)
	Special classifier noun	Nn_swsp	Noun sub-phrase A sequence of a special classifier noun (Ncs) and its modifier (verb/adjective)
	Categorization noun	Nn_swsp	Noun sub-phrase A sequence of a categorization noun (Nn) and its modifier (noun)
	Repetition form	Vv_swsp_Rt	Verb Repetition form having meaning as verb Ex: <i>Đi đi lại lại</i>
		Aa_swsp_Rt	Adjective Repetition form having meaning as adjective Ex: <i>Lóng la lóng lánh, trùng trùng điệp điệp</i>

		Nn_swsp_Rt	Common noun	Repetition form having meaning as noun Ex: <i>Chiều chiều</i>
		ON_swsp_Rt	Sound	Repetition form is a sound Ex: <i>Đùng ùng ùng</i>

4.2. Principles of W and SWSP annotation

4.2.1. Principles of W annotation

4.2.1.1. Common nouns (Nn_w)

The sequences of the Sino-Vietnamese elements and the words they modify are used to name the things or phenomena. These sequences will be tagged as Nn_w. There are 2 types of Nn_w as follows:

- **Sino-Vietnamese + noun/verb/adjective/FW = Nn_w**

In this sequence type, the Sino-Vietnamese elements, such as: *bách, bá, bán, bản, bắt, bốn, chủ_nghĩa, chư, cố, cữu, đa, đại, đệ, đồng, hậu, lão, liên, nam, nguyên, nữ, nhà, phó, siêu, tiên, tiểu, tổng, trưởng, vi, vô* are pre-modifiers of nouns, verbs, adjectives, or FWs.

Examples:

(Nn_w *bách*/SV *bệnh*/Nn)

(Nn_w *bá*/SV *quan*/Nn), (Nn_w *bá*/SV *bệnh*/Nn)

(Nn_w *chư*/SV *vị*/Nn) , (Nn_w *chư*/SV *huynh*/Nn)

(Nn_w *cố*/SV (Nn_w *bộ*/Nn *trưởng*/SV)) , (Nn_w *cố*/SV *giáo_sư*/Nn)

(Nn_w *trưởng*/SV *ban*/Nn) , (Nn_w *trưởng*/SV *phòng*/Nn)

(Nn_w *vi*/SV *sinh_vật*/Nn) , (Nn_w *vi*/S/Nn_w *điện_tử*/Nn)

(Nn_w *chủ_nghĩa*/SV *xã_hội*/Nn)

(Nn_w *chủ_nghĩa*/SV *tư_bản*/Aa)

(Nn_w *đệ*/SV *tam*/Aa), (Nn_w *đệ*/SV *nhất*/Aa)

(Nn_w *bốn*/SV *toạ*/Av)

(Nn_w *tiền*/SV *seagames*/FW)

(Nn_w *hậu*/SV *paragames*/FW)

- **Noun/verb/adjective + Sino-Vietnamese = Nn_w**

The Sino-Vietnamese elements, such as: *bộ, học, kiểu, tính, trưởng, viên* are post-modifiers of the nouns/verbs/adjectives.

Examples:

(Nn_w đảng/Nn bộ/SV), (Nn_w huyện/Nn bộ/SV)
 (Nn_w tâm_lí/Nn học/SV), (Nn_w toán/Nn học/SV)
 (Nn_w Việt/Nn kiểu/SV), (Nn_w Hoa/Nn kiểu/SV)
 (Nn_w nhạc/Nn tính/SV), (Nn_w giới/Nn tính/SV)
 (Nn_w đại_đoàn/Nn trưởng/SV), (Nn_w phân_đoàn/Nn trưởng/SV)
 (Nn_w công_an/Nn viên/SV), (Nn_w xã/Nn viên/SV), (Nn_w hội/Nn viên/SV)
 (Nn_w cộng_tác/Vv viên/SV), (Nn_w lập_trình/Vv viên/SV), (Nn_w điều_tra/Vv viên/SV)
 (Nn_w chuyên/Aa viên/SV)

4.2.1.2. Verb (Vv_w)

The sequences of the Sino-Vietnamese elements and the words they modify which denote the actions or states of things will be tagged as Vv_w. The constructions of these sequences are following:

- **Sino-Vietnamese + verb = Vv_w**

In this sequence type, the Sino-Vietnamese elements, such as: *bắt, đồng, siêu, tái, tổng* are pre-modifiers of the verbs.

Examples:

(Vv_w bắt/SV bạo_động/Vv), (Vv_w bắt/SV hợp_tác/Vv)
 (Vv_w đồng/SV học/Vv)
 (Vv_w siêu/SV dẫn/Vv), (Vv_w siêu/SV thoát/Vv), (Vv_w siêu/SV thắng/Vv)
 (Vv_w tái/SV cơ_cấu/Vv), (Vv_w tái/SV đầu_tư/Vv), (Vv_w tái/SV định_cư/Vv), (Vv_w tái/SV sản_xuất/Vv)
 (Vv_w tổng/SV chi/Vv), (Vv_w tổng/SV kiểm_kê/Vv)

- **Noun + Sino-Vietnamese = Vv_w**

In this sequence type, the Sino-Vietnamese element *hóa* is post-modifier of the verbs.

Examples:

(Vv_w lão/Nn hóa/SV), (Vv_w tri_thức/Nn hóa/SV), (Vv_w công_nghiệp/Nn hóa/SV)

4.2.1.3. Adjective (Aa_w)

The sequences of the Sino-Vietnamese elements and the words they modify having meaning as adjective will be tagged as Aa_w. The constructions of these sequences as follows:

- **Sino-Vietnamese + noun/verb/adjective = Aa_w**

In this case, the Sino-Vietnamese elements, such as: *bách, đồng, hiểu, hữu, phản, phi, tiền, vô, bán, bắt, siêu*, are pre-modifiers.

Examples:

(Aa_w bách/SV chiến/Nn), (Aa_w bách/SV thắng/Nn)

(Aa_w đồng/SV nghĩa/Nn), (Aa_w đồng/SV thời/Nn)

(Aa_w hiểu/SV khách/Nn)

(Aa_w hữu/SV hình/Nn), (Aa_w hữu/SV sự/Nn), (Aa_w hữu/SV thân/Nn)

(Aa_w phản/SV tác_dụng/Nn), (Aa_w phản/SV khoa_học/Nn)

(Aa_w phi/SV chính_phủ/Nn), (Aa_w phi/SV lợi_nhuận/Nn), (Aa_w phi/SV nhân_đạo/Nn)

(Aa_w tiền/SV khởi_nghĩa/Nn)

(Aa_w vô/SV chủ/Nn), (Aa_w vô/SV đạo_đức/Nn), (Aa_w vô/SV văn_hoá/Nn)

(Aa_w hiểu/SV học/Vv)

(Aa_w hữu/SV dụng/Vv), (Aa_w hữu/SV khuynh/Vv), (Aa_w hữu/SV sinh/Vv), (Aa_w hữu/SV trách/Vv)

(Aa_w tiền/SV khởi_nghĩa/Vv)

(Aa_w vô/SV can/Vv), (Aa_w vô/SV địch/Vv), (Aa_w vô/SV học/Vv)

(Aa_w bán/SV tự_động/Aa), (Aa_w bán/SV công_khai/Aa)

(Aa_w bắt/SV bình_đẳng/Aa), (Aa_w bắt/SV đặc_chí/Aa), (Aa_w bắt/SV hợp_lí/Aa), (Aa_w bắt/SV khả_thi/Aa)

(Aa_w siêu/SV thực/Aa), (Aa_w siêu/SV trường/Aa), (Aa_w siêu/SV trọng/Aa)

(Aa_w tiền/SV tư_bản/Aa chủ_nghĩa/SV)

(Aa_w vô/SV cùng/Aa)

- **Adjective/noun + Sino-Vietnamese = Aa_w**

In this sequence type, the Sino-Vietnamese elements, such as: *chủ_nghĩa, tính*, are post-modifiers.

Examples:

(Aa_w tư_bản/Aa chủ_nghĩa/SV)

(Aa_w kịch/Nn tính/SV)

4.2.1.4. *Adjunct (R_w)*

The sequences of the Sino-Vietnamese elements and the words they modify having meaning as adjuncts can be tagged as R_w. In these cases, the Sino-Vietnamese elements, such as: *đồng, bắt, vô* are pre-modifiers. The main-words can be noun, verb, or adjective.

Examples:

(R_w đồng/SV thời/Nn)

(R_w bắt/SV luận/Vv)

(R_w vô/SV kẻ/Vv), (R_w vô/SV luận/Vv)

(R_w vô/Vv cùng/Aa)

4.2.2. Principles of SWSP annotation

SWSP (supra-word sub-phrase) tag set is used to annotate the sequences which there is little consensus as to the methodology for annotating them. SWSP tag set is described as follows:

4.2.2.1. *Noun sub-phrase (Nn_swsp)*

Noun sub-phrase is a sequence of a classifier noun (Nc) and a noun, a sequence of a special classifier noun (Ncs) and a verb or adjective, or a sequence of a categorization noun and a noun. This sequence is used to name a thing or a phenomenon. It will be tagged as Nn_swsp. The following is constructions of these sequences.

- **Nc + modifier = Nn_swsp**

In this sequence type, the modifiers are common nouns (Nn) or foreign words (FW).

Examples:

(Nn_swsp cái/Nc nhà/Nn)

(Nn_swsp bài/Nc thơ/Nn)

(Nn_swsp bức/Nc tranh/Nn)

(Nn_swsp con/Nc cá/Nn)

(Nn_swsp cuốn/Nc album/FW)

(Nn_swsp tấm/Nc poster/FW)

(Nn_swsp chiếc/Nc camera/FW)

- **Ncs + modifier = Nn_swsp**

In this sequence type, the modifiers can be nouns (Nn), verb (Vv) or adjective (Aa).

Examples:

(Nn_swsp nhà/Ncs khoa_học/Nn)

(Nn_swsp sự/Ncs nhiệt_huyết/Nn)

(Nn_swsp niềm/Ncs cảm_xúc/Nn)

(Nn_swsp cuộc/Ncs thi/Vv)

(Nn_swsp nổi/Ncs đam_mê/Vv)

(Nn_swsp vụ/Ncs tham_ô/Vv)

(Nn_swsp sự/Ncs yên_bác/Aa)

(Nn_swsp điều/Ncs kì_diệu/Aa)

(Nn_swsp nổi/Ncs vất_vả/Aa)

(Nn_swsp sự/Ncs nhiệt_tình/Aa và sôi_nổi/Aa)

- **Subordinate compound words**

A sequence of the components of a subordinate compound word will be tagged as Nn_swsp.

Examples:

(Nn_swsp Cá/Nn chép/Nn hồng/Aa) {pink carp}

(Nn_swsp Cá/Nn rô/Nn đồng/Nn) {pink carp}

(Nn_swsp máy_bay/Nn bà_già/Nn) {morane fighter plane}

(Nn_swsp máy_bay/Nn trực_thăng/Nn) {helicopter}

(Nn_swsp máy_in/Nn kim/Nn) {dot-matrix printer}

(Nn_swsp máy_in/Nn đập/Vv) {impact printer}

(Nn_swsp máy/Nn-H (VP trộn/Vv-H (NP hồ/Nn-H)))

4.2.3. Note

- In the structure including A B C, if C modifies A, the internal structure includes A B. Otherwise, if C modifies B, we bracket B C with an appropriate phrase tag and bracket A B C with an internal structure tag.

Examples:

(NP

(Nn_swsp (Ncs-H sự) (Vv-H hỗ_trợ))

(PP (Cs-H của)

(NP (Nn-H Tỉnh_đoàn)

(NP (Nr-H Đắc_Nông))))))

(NP-SBJ

(Nn_swsp (Ncs-H Việc)

(VP (Vv-H chuyển_nhượng)

(NP (Nn-H VĐV))))))

- For the cases that an internal structure is a part of another internal structure, we determine head of each internal structure.

(Nn-swsp nhà/Ncs-H

(Nn_w tâm_lý/Nn-H học/Sv))

4.2.3.1. Repetition form

The repetition forms will be tagged as follows:

- Having meaning as nouns: Nn_swsp_Rt
- Having meaning as verbs: Vv_swsp_Rt
- Have meaning as adjectives: Aa_swsp_Rt
- Being the sounds: ON_swsp_Rt .

Examples:

[âm/ON âm/ON âm/ON]/ON_swsp_Rt

[đi/Vv đi/Vv lại/D lại/D]/ Vv_swsp_Rt

5. Appendices

5.1. Combination rules of Sino-Vietnamese elements and other words

Prefixes

bách + N = N: bách bệnh, bách thắng

bách + N = A: bách chiến, bách thắng

bá + N = N: bá quan, bá bệnh, bá chứng

bán + N = N: bán bình nguyên, bán nguyên âm, bán sơn địa, bán thành phẩm, bán cầu

bán + A = A: bán tự động, bán vũ trang, bán công khai, bán phong kiến

bản + N = N: bản toạ, bản hiệu (cửa hiệu), bản báo

bất + A = A: bất bình đẳng, bất đắc chí, bất hợp lí, bất khả thi.

bất + V = V: bất bạo động, bất hợp tác.

bất + N = N: bất đẳng thức, bất động sản, bất phương trình.

bất + V = R: bất luận

bổn + N = N: bổn toạ, bổn thân

chủ_nghĩa + N/A = N: chủ nghĩa tư bản, chủ nghĩa xã hội

chư + N = N: chư vị, chư huynh

cố + N = N: cố bộ trưởng, cố giáo sư, cố thủ tướng, v.v.

cựu + N = N: cựu bộ trưởng, cựu giám đốc, cựu thủ tướng, v.v.

đa + N = N: đa phương tiện, đa tác vụ, đa khoa

đại + N = N: đại bản doanh, đại bộ phận, đại công nghiệp, đại gia đình.

đệ + A/N = N: đệ tam, đệ nhất

đồng + N = N: đồng chí, đồng thời, đồng chủ tịch

đồng + N/A = A: đồng nghĩa

đồng + V = V: đồng học

đồng + N = R/A: đồng thời

hậu + N = N: hậu tố, hậu hoạ

hậu + V = V: hậu kiểm, hậu chi

hiếu + V = A: hiếu học

hiếu + N = A: hiếu khách

hữu + N = A: hữu hạn, hữu hình, hữu sự, hữu thần.

hữu + V = A: hữu dụng, hữu khuynh, hữu sinh, hữu trách

lão + N = N: lão nghệ nhân, lão ông, lão bà

liên + N = N: liên bang, liên bộ, liên ngành, liên cầu khuẩn, liên chi uỷ, v.v.

nam + N = N: nam thanh niên, nam công nhân

nguyên + N = N: nguyên bộ trưởng, nguyên thủ tướng, nguyên trưởng phòng, v.v.

nữ + N = N: nữ công nhân, nữ nghệ sĩ

phản + N = A: phản tác dụng, phản khoa học

phi + N = A: phi chính phủ, phi lợi nhuận, phi nhân đạo, phi nông nghiệp.

phó + N = N: phó chủ nhiệm, phó chủ tịch, phó viện trưởng, phó giám đốc.

siêu + N = N: siêu giai cấp, siêu hạng, siêu sao, siêu cầu thủ, siêu lợi nhuận, siêu trầm

siêu + V = V: siêu dẫn, siêu thoát, siêu thắng

siêu + A = A: siêu thực, siêu trường, siêu trọng

tái + V = V: tái cơ cấu, tái đầu tư, tái định cư, tái sản xuất

tiền + N = N: tiền lệ, tiền đồn

tiền + N/V = A: tiền khởi nghĩa

tiền + A = A: tiền tư bản chủ nghĩa

tiểu + N = N: tiểu bang, tiểu công nghệ, tiểu gia súc, tiểu khí hậu, tiểu loại, tiểu vương quốc

tổng + N = N: tổng số

tổng + V = V: tổng chi, tổng kiểm kê

trưởng + N = N: trưởng ban, trưởng phòng, trưởng thôn, trưởng tộc

vi + N = N: vi khuẩn, vi sinh vật, vi điện tử

vô + N = A: vô chủ, vô đạo, vô đạo đức, vô gia cư, vô nhân đạo, vô thần, vô văn hoá

vô + V = A: vô can, vô địch, vô học

vô + V = R: vô kể, vô luận

vô + A = R/A: vô cùng

Suffixes

N + bộ = N: đảng bộ, huyện bộ

A/N + chủ_nghĩa = A: hiện thực chủ nghĩa, tư bản chủ nghĩa

N + hoá = V: lão hoá, lao động hoá, công nông hoá, trí thức hoá

N + học = N: tâm lí học, toán học

N + kiều = N: Ấn kiều, Hoa kiều, Việt kiều

N + tính = N/A: nhạc tính, kịch tính, độc tính, giới tính

N + trưởng = N: đại đoàn trưởng, phân viện trưởng, tiểu đoàn trưởng

V + viên = N: cộng tác viên, lập trình viên, điều tra viên, diễn viên

N/V + viên = N: cộng sự viên, mưu dịch viên, phẫu thuật viên

N + viên = N: công an viên, xã viên, hội viên

A + viên = N: chuyên viên

5.2. List of classifier nouns

No.	Classifier noun (Nc)	Examples
1	à	nô tì, con gái, mèo mèo, ngựa
2	áng	văn {literature}, thơ {poem}
3	anh	sinh viên {student}, cán bộ {official}
4	ánh	đèn, trắng, kim loại
5	bà	mẹ, chủ nhiệm {chairman}, vợ {wife}, nội trợ, giám thị
6	bác	lãnh đạo, bảo vệ, công nhân {worker}, phu xe {rickshaw driver}
7	bài	thơ {poem}, diễn văn {speech},
8	bản	tuyên ngôn {speech}, tài liệu {materials},
9	bạn	sinh viên, học sinh, diễn, nhảy, đồng hành
10	băng	ghế {chair}, khẩu hiệu {slogan}
11	bậc	vĩ nhân {great man}, anh hùng {hero}
12	bé	trai {son/ boy}, gái {daughter/ girl/}
13	bộ	từ điển {dictionary}, máy tính {computer}
14	bức	tranh {picture}, thư {letter}, tượng {statue}, vách {wall}, ảnh {image}
15	cái	bàn {table}, ghế {chair}, đầu {head}, thuyền {boat} khuyết điểm {mistake}, tâm trạng {mood}
		ong, kiến
16	căn	phòng {room}, nhà {house}
17	cậu	bé, học trò {student}, con trai {boy}, kiếm lâm
18	cây	nến {candle}, đèn {light}, roi {whip}, bút {pen}, sung {fig}, đàn {instrument}, tăm {toothpick},...
19	chàng	thi sĩ {poet}, văn nhân {man of letters}, tình báo

20	chị	giáo viên{teacher}, nhà báo{journalist}, kế toán
21	chiếc	bàn{table}, ghế{chair}, thuyền{boat}
22	chú	bé, tiểu, liên lạc{communicate}, tài xế{driver}, kiểm lâm để mền, chuột
23		
24	con	dao{knife}, thuyền{boat}, sông{river} gián điệp{spy}, mẹ mìn{kidnapper} trâu{buffalo}, bò{cow}, gà{chicken} trai {son}, gái {daughter}, chị {older sister}, em {younger sister}
25	cọng	rom, dây, bún, tóc, lông
26	cô	bé, thợ, dược sĩ{chemist}, y tá{nurse}, biên tập
27	cỗ	máy, pháo
28	cộng	rom, dây, bún, tóc, lông
29	con	gió{wind}, dông{storm}
30	cụ	ông, bà, phụ lão
31	cuốn	sách{book}, tập{notebook}, vở tiểu thuyết{novel}
32	dải	thắt lưng{belt}, áo{shirt}
33	dệ	dệ đường, đệ đệ, đệ cò, đệ sông
34		
35	đóa	hồng{rose}, mây{cloud}
36	đòn	bánh tét
37	đứa	bạn, con, bé, cháu, trẻ{kid}, con nít{child}
38	em	học sinh{student}, nhi đồng{child}
39	gã	sĩ quan địch{enemy officer}, nguy binh{puppet soldier}, bảo kê
40		
41	hạt	muối, sạn, mưa, ngọc, cát
42	hòn	đạn{bullet}, bi{ball}, núi{mountain} đá {stone, rock}
43	hột	gạo{rice}, bắp{corn}
44	kẻ	tiểu nhân, đàn anh
45	khẩu	súng{gun}, đại bác{canon}
46	Lá	thư{letter}, bùa{spell},
47	làn	gió{wing}, mây {cloud}, tóc, da
48	lão	thầy bói, chủ
49	màn	sương{dew}, đêm{night}, mưa/Nn
50	mụ	chủ, vợ, đàn bà{woman}, vợ{wife}
51	nàng	công chúa{princess}, tiên{fairy}, cộng sự
52	ngài	đại sứ, chủ tịch, toàn quyền, thủ tướng
53	ngọn	cờ{flag}, núi{mountain},
54	ngôi	nhà{house}, đền{temple}, mộ{tomb},
55	nóc	xe{car}, tủ{closet}
56	ông	quản gia{steward}, tri huyện{district chief}, bảo vệ, đạo diễn, cố vấn
57	phiến	đá, gỗ

58	pho	tượng{statue}, truyện{story}, sách{book}
59	quả	bom{bomb}, núi{mountain}, tim{heart},
60	quyển	sách{book}, vở{notebook}
61	tàu	chuối, dừa, lá
62	tay	thầu khoán{contractor}, nhà buôn{merchant}, súng, vợt, búa, anh chị, phó nháy
63	tấn	tuồng{drama}, kịch{drama}
64	tên	sĩ quan địch{enemy officer}, nguy binh{puppet soldier}, môi giới, bảo kê
65	thanh	gươm{sword}, kiếm{sword}
66	thằng	quỷ sứ{devil}, mật thám{secret agent}
67	toà	nhà{house}, lâu đài{castle}
68	trái	bom{bomb}, núi{mountain}, tim{heart}
69	vệ	vệ đường, vệ đê, vệ cỏ, vệ sông
70	vì	sao{star}, tinh tú{star}
71	vị	phụ lão{elder}, chủ tịch{president}, chủ toạ, đại diện, lãnh đạo, đạo diễn
72	viên	sĩ quan{officer}, đại uý{captain}, thanh tra, chỉ huy
73	vở	kịch{drama}, tuồng{drama}

5.3. List of special classifier nouns

No.	Classifier noun (Ncs)	Examples
1	cuộc	thi đấu{competition}, điều tra{investigation}
2	điều	phi pháp{illegal}, ước mơ{wish}
3	mối	sâu, bất bình, đe dọa, tình, quan hệ
4	nền	dân chủ, cộng hoà, văn hoá{culture}, độc lập{independent}, khoa học{science}
5	nhà	văn, thơ, khoa học{science}, nghiên cứu{research}, tư vấn, thiết kế, tài trợ
6	niềm	vui{happy}, kiêu hãnh{proud}
7	nỗi	buồn{sad}, nhớ, bất công
8	sự	thành công{success}, đời{life}
9	vụ	kh khiếu nại{complaint}, rắc rối{trouble}, tham ô, cháy, án
10	việc	học{study}, sản xuất{produce}, chăm sóc
11	cái	nhìn{look}, đẹp, ăn, mặc, hay, dở
12	chúng	ho/Vv, buồn nôn/Aa, ba hoa/VvAa
13	đáng	anh hùng{hero}, thánh thần{saint}, cứu tinh, sinh thành/Vv
14	giấc	mơ{dream}, ngủ{sleep}
15	phiên	chứng khoán{stock}, chợ{market}, điều trần, xử, họp
16	tấm	ảnh{picture}, tranh{picture}, bảng{table}, bìa{cover}, quảng cáo, phủ, lợp

5.4. List of unit nouns

No.	Word	Examples
1	\$ {dollar}	
2	% {percent}	
3	£ {pound}	
4	€ {Euro}	
5	ampe {ampere}	
6	ao-xơ	
7	át-môt-phe	
8	au-xơ	
9	bảng	
10	bit	
11	bít	
12	c {coulomb, celsius, carat }	
13	cắc	
14	cal {calori}	
15	calo {calori}	
16	ca-lo {calori}	
17	cân {kilo}	3 cân thịt {3 kilos of meat}, 2 cân đường {2 kilos of sugar}
18	candela	
19	cao {=1/10 hectare}	
20	cara {carat}	
21	carat	
22	cây {kilometer}	
23	cây số {kilometer}	
24	cc {centimetre-cube}	
25	cedi	
26	cent	
27	centimet	
28	chạ	
29	châu	
30	chỉ	
31	chiềng	
32	chợ	
33	chòm	
34	chùa	
35	cm	
36	cm ²	
37	cm ³	
38	colon	
39	công {=1/7 - 1/10 hectare}	
40	công mẫu	

41	cordoba	
42	coulomb	
43	cron	
44	culon	
45	culong	
46	cu-lông	
47	cụm	
48	cút	
49	đ { Vietnam Dong }	
50	đạc { = 60 meters }	
51	đài tề	
52	dalasi	
53	dặm	
54	đạo	
55	đầu { = 1 liter or 1/2 m3 }	
56	deca	
57	đê-ca	
58	đeca-	
59	deci	
60	decibel	
61	denar	
62	đexiben	
63	điểm	
64	đin	
65	dinar	
66	dinh	
67	điop	
68	đi-ốp	
69	dioptr	
70	dirham	
71	dl	
72	đô	
73	độ { degree }	nước sôi 100 độ { boiling water }
74	đô la	
75	dobra	
76	đội	
77	đôla	
78	đô-la	
79	dollar	
80	đôn	
81	đồng	
82	động	
83	đồng cân	
84	dpi	
85	drachma	
86	dram	

87	dyn	
88	entropi	
89	en-tro-pi	
90	escudo	
91	euro	
92	euros	
93	exi	
94	F	
95	fara	
96	fa-ra	
97	feet	
98	florin	
99	foot	
100	foot vuông	
101	forint	
102	fps	
103	franc	
104	g	
105	gallon	
106	ga-lon	
107	gam	
108	gamma	
109	già	
110	giác	
111	giáp	
112	gigabyte	
113	gigawatt	
114	gourde	
115	gr	
116	gram	
117	guarani	
118	guilder	
119	ha	
120	hải lí	
121	hải lý	
122	hào	
123	hạt	
124	hec	
125	hec ta	
126	hecta	
127	hec-to	
128	hec-tô	
129	hecto-	
130	hertz	
131	hp	
132	hryvnia	

133	i	
134	in	
135	inch	
136	J	
137	joule	
138	Jun	
139	K	
140	k	
141	k-	
142	kb	
143	kcal	
144	kelvin	
145	kenvin	
146	ken-vin	
147	kg	
148	khâu	khâu vàng
149	khối	cm khối
150	kí	
151	ki lô	
152	kí lô	
153	kilô	
154	kilo-	
155	kilogam	
156	kilôgam	
157	kilomet	
158	kilômet	
159	kilooat	
160	kilooat-giờ	
161	kilôtôn	
162	kilowatt	
163	kilowatt-giờ	
164	kina	
165	kip	
166	kíp	
167	km	
168	km/h	
169	km ²	
170	km ³	
171	korun	
172	koruna	
173	kp	
174	kp.	
175	krona	
176	krone	
177	kronor	
178	kroon	

179	kruna	
180	kuna	
181	kv	
182	kva	
183	kwacha	
184	kwanza	
185	kwh	
186	ký	
187	ký lô	
188	kyat	
189	l	
190	lai	lai vàng = phân vàng
191	lạng	
192	lạng ta	
193	lari	
194	lat	
195	lei	
196	lek	
197	lempira	
198	leone	
199	lev	
200	li	
201	lia	
202	lilangeni	
203	lira	ruợu {wine}, nước {water}
204	lít {litre}	
205	litas	
206	loti	
207	lumes	
208	lượng	
209	ly	
210	m	
211	m-	
212	m ²	
213	m ³	
214	ma	
215	mã	
216	mã lực	
217	mác	
218	manat	
219	mark	
220	markka	ruợng, đất
221	mẫu	
222	mẫu ta	
223	mẫu tây	
224	mbps	

225	mcg	
226	me-ga	
227	mê-ga	
228	megabit	
229	megabyte	
230	megapixel	
231	megawatt	
232	mét	
233	mét khối	
234	mét vuông	
235	metical	
236	mg	
237	mg%	
238	mhz	
239	mỹ kim	
240	mi li mét	
241	micro-	
242	microgam	
243	microgram	
244	micromet	
245	mi-crô-mét	
246	micron	
247	mi-cron	
248	mi-li	
249	mili-	
250	mililit	
251	milimet	
252	milligram	
253	millirem	
254	ml	
255	mm	
256	mm ²	
257	mm ³	
258	mmhg	
259	mmol	
260	mp	
261	mpx	
262	msv	
263	mva	
264	mỹ kim	
265	naira	
266	năm ánh sáng	
267	nano-	
268	nanogram	
269	nén	
270	newton	

271	ng	
272	ngựa	
273	ngultrum	
274	nhân dân tệ	
275	niuton	
276	niu-ton	
277	nm	
278	nt\$	
279	ohm	
280	om	
281	ouguiya	
282	ounce	
283	oz	
284	p	
285	pa'anga	
286	peseta	
287	peso	
288	phân	
289	phân khối	
290	phút	
291	pint	
292	pixel	
293	plây	
294	pound	
295	ppi	
296	ppm	
297	pula	
298	quan	
299	quart	
300	quetzal	
301	r	
302	rad	
303	radian	
304	ra-đi-an	
305	rand	
306	real	
307	rĩa	
308	rial	
309	richter	
310	riel	
311	riên	
312	ringgit	
313	riyal	
314	rouble	
315	ruble	
316	rufiyaa	

317	rup	
318	rúp	
319	rupee	
320	rupi	
321	rupiah	
322	sào	
323	sấp	
324	schilling	
325	sd	
326	sek	
327	shekel	
328	shilling	
329	sol	
330	soles	
331	son	
332	soum	
333	sucré	
334	t	
335	tạ	
336	tắc	
337	taka	
338	tala	
339	tấn{ tone }	sắt{iron}, vàng{gold}
340	tê	
341	tenge	
342	terabyte	
343	thạch	
344	thước	vải
345	tolar	
346	tughrik	
347	uoát	
348	V	
349	vạn	
350	vatu	
351	verst	
352	vnd	
353	vnd	
354	volt	
355	vôn	
356	w	
357	watt	
358	watt-giờ	
359	won	
360	xenti-	
361	xu	
362	yard	

363	yd	
364	yen	
365	yến	
366	yuan	
367	zloty	

5.5. Unit nouns of administrative location

No.	Word	Examples
1	huyện	
2	huyện đảo	
3	huyện lị	
4	huyện lỵ	
5	hương	
6	hộ	
7	khối phố	
8	khóm	
9	khu	
10	khu phố	
11	lân	
12	làng	
13	lô	
14	lộ	
15	lũng	
16	miếu	
17	mường	
18	nậu	
19	nguồn	
20	phố	
21	phủ	
22	phường	
23	quán	
24	quận	
25	sách	
26	sở	
27	sóc	
28	thành	
29	thành phố	
30	thị	
31	thị trấn	
32	thị tứ	
33	thị xã	
34	thôn	
35	thừa tuyên	
36	thuộc	
37	tích	

38	tiết	
39	tiểu bang	
40	tiểu khu	
41	tỉnh	
42	tỉnh đảo	
43	tổ	
44	tổ dân phố	
45	tộc	
46	tổng	
47	trại	
48	trần	
49	trang	

5.6. List of nouns of time

No.	Word	Examples
1	ban	Trưa, chiều, đêm
2	bình minh	
3	buổi	
4	bữa	
5	cá tháng tư	
6	canh	
7	canh gà	
8	chạp	
9	chập	
10	chập	
11	chiều	
12	chốc	
13	chốc lát	
14	chủ nhật	
15	chúa nhật	
16	chừ	
17	con trăng	
18	cổ lai	
19	cổ kim	
20	công nguyên	
21	dạo	
22	đại	
23	đận	
24	đầu hôm	
25	đêm	
26	đêm hôm	
27	đêm ngày	
28	đêm tối	
29	đêm trừ tịch	
30	đêm trường	

31	đông	
32	đương thời	
33	g {hour}	
34	giác	
35	giáng sinh	
36	giao thời	
37	giao thừa	
38	giáp	
39	giáp vụ	
40	giắc	
41	giây	
42	giây lát	
43	giây phút	
44	giêng	
45	giêng hai	
46	giờ	
47	giờ khắc	
48	giờ lâu	
49	giờ phút	
50	h	
51	hạ	
52	hạ tuần	
53	hai	
54	hàn thực	
55	hè	
56	hiện	
57	hiện giờ	
58	hiện nay	
59	hiện tại	
60	hiện thời	
61	hiệp	
62	hoàng hôn	
63	hồi	
64	hỏi	
65	hôm	
66	hôm kia	
67	hôm kia	
68	hôm mai	
69	hôm nay	
70	hôm qua	
71	hôm sớm	
72	hôm	
73	ít nữa	
74	khắc	
75	khi	
76	khi hồi	

77	khuya	
78	kì	
79	kỉ	
80	kỉ nguyên	
81	kia	
82	kỳ	
83	kỷ	
84	lát	
85	lâu nay	
86	lóng	
87	lúc	
88	mai	
89	mai đây	
90	mai hậu	
91	mai_kia	
92	mai_một	
93	mai_sau	
94	mai_này	
95	mọi khi	
96	mông	
97	một	
98	mới đây	
99	mới rồi	
100	mùng	
101	mười	
102	nay	
103	nay_mai	
104	năm	
105	năm_kia	
106	năm ngoái	
107	năm tháng	
108	ngày	
109	ngày mai	
110	ngày nay	
111	ngày rày	
112	ngày xưa	
113	ngày_xưa ngày_xưa	
114	nguyên đán	
115	nguyên tiêu	
116	nhát	
117	ni	
118	niên	
119	nửa đêm	
120	phục sinh	
121	phút	
122	phút_giây	

123	quá khứ	
124	quí	
125	quý	
126	rày	
127	rằm	
128	s	
129	sáng	
130	sáng mai	
131	sáng qua	
132	sớm	
133	sớm chiều	
134	sớm hôm	
135	sớm khuya	
136	sớm tối	
137	sớm trưa	
138	tết	
139	tháng	
140	tháng củ mật	
141	tháng đủ	
142	tháng năm	
143	tháng ngày	
144	thập kỉ	
145	thập kỷ	
146	thập niên	
147	thâu	
148	thế kỉ	
149	thế kỷ	
150	thiên kỉ	
151	thiên kỷ	
152	thiên niên kỉ	
153	thiên niên kỷ	
154	thời	
155	thời đại	
156	thời điểm	
157	thu	
158	tiếng	
159	tiết	
160	tối	
161	trung thu	
162	trung tuần	
163	trừ tịch	
164	trưa	
165	trước kia	
166	tuần	
167	tuần lễ	
168	tương lai	

169	vu_lan	
170	vừa_nhảy	
171	vừa_qua	
172	vừa_rồi	
173	xế	
174	xế_xế	
175	xuân	
176	xuân_phân	
177	xưa	
178	xưa_nay	
179	kỷ_nguyên	
180	thượng_jura	
181	chính_trung	
182	mi-li-giây	

5.7. Distinguishing between Cs and other word types

Table 1: Examples to distinguish between D and Cs

Word	D	Cs
	Distinguishable point	Distinguishable point
đến	Indicating the direction of an activity Ex: <i>Đi đến trường</i>	Ex: <i>Nghĩ đến việc ấy lại thấy lo</i> <i>Đi đến quyết định cuối cùng</i>
qua	Ex: <i>Bơi qua sông</i>	Ex: <i>Học tập qua sách vở</i> <i>Kể qua chuyện khác</i>
sang	Ex: <i>Đi sang Nga</i>	Ex: <i>Chuyển sang vấn đề khác</i>
tới	Ex: <i>Đi tới trước</i>	Its meaning is similar to the word <i>đến</i> Ex: <i>Nói chuyện tới khuya</i>
về	<i>Đi về nhà</i> <i>Tàu chạy về Bình Định</i>	Ex: <i>Nhìn về phía trước</i> <i>Giỏi về toán học</i>
vô	Ex: <i>Đi vô nhà</i>	Ex: <i>Nhìn vô nhà</i> <i>Nô lệ vô sách vở</i>

Table 2: Examples to distinguish between R and Cs

Word	R	Cs
	Distinguishable point	Distinguishable point
đặng	Indicating the ability to perform an affair. Ex: <i>Ngủ không đặng</i>	Indicating the goal that we want to obtain (similar to the word <i>để</i>) Ex: <i>Nói ra đặng (để) nhẹ lòng</i>
qua	It means <i>carelessly</i> or <i>have been</i> Ex:	Ex: <i>Kiểm nghiệm qua thực tế</i>

	<i>Quét qua cái nhà</i> <i>Tôi đã ghé qua những trang web đó rồi</i>	<i>Chuyển qua vấn đề khác</i>
tới	Ex: <i>Đi tới đi lui</i> <i>Làm tới làm lui</i>	Ex : <i>Nhớ tới chuyện cũ</i> <i>Nói chuyện tới khuya</i>

Table 3: Examples to distinguish between Vv and Cs

Word	Vv	Cs
	Distinguishable point	Distinguishable point
bằng	Ex: <i>Bằng nhau</i> <i>Anh bằng vào đâu mà nói thế?</i>	Ex: <i>Bàn bằng gỗ</i> <i>Làm bằng tay</i> <i>Đi bằng tàu hoả</i> <i>Ăn bằng hết mới thôi</i>
để	Ex: <i>Để cây bút trên bàn</i> <i>Vết thương để lại sẹo</i> <i>Cửa để ngỏ</i>	Ex: <i>Nhà để ở</i> <i>Hứa trước làm gì để cho nó mong</i>
đến	Ex: <i>Đông qua, xuân đến</i> <i>Việc đến bất ngờ</i>	Ex: <i>Nghĩ đến bạn</i> <i>Nói chuyện đến khuya</i>
nhằm	Ex: <i>Nhằm mục tiêu mà bắn</i> <i>Nhằm lúc sơ hở</i>	Ex: <i>Nhằm ngày 25/6 âm lịch</i> <i>Nói thêm nhằm thanh minh</i>
nhờ	Ex: <i>Nhờ trông xe</i> <i>đi nhờ xe</i> <i>ngủ nhờ nhà bạn</i> <i>Đi đi cho tôi nhờ</i>	Ex: <i>Nhờ anh giúp đỡ mọi việc mới êm xuôi</i>
ở	Ex: <i>Tôi ở Sài Gòn</i> <i>Ở bản</i> <i>Ở bạc</i> <i>Đưa ở</i>	Ex: <i>Xem phim ở rạp</i> <i>Tin tưởng ở tương lai</i> <i>Lỗi ở con</i>
qua	Ex: <i>Qua năm mới</i> <i>Qua lớp tập huấn</i> <i>Cố qua mắt địch</i>	Ex: <i>Kiểm nghiệm qua thực tế</i> <i>Chuyển qua vấn đề khác</i>
sang	Ex: <i>Sang năm mới</i> <i>Sang tên</i>	Ex: <i>Chuyển sang vấn đề khác</i>
tới	Ex: <i>Xuân đang tới</i>	Ex: <i>Nhớ tới chuyện cũ</i>

	<i>Trẻ chưa qua, già chưa tới</i>	<i>Nói chuyện tới khuya</i>
vào/vô	Ex: <i>Vào biên chế</i> <i>Học vào loại giỏi</i>	Ex: <i>Nhìn vô nhà</i> <i>Vào dịp tết</i>
về	Ex: <i>Hàng về bến</i> <i>Trời ngả về chiều</i> <i>Tiền về tay tôi</i>	Ex: <i>Nhìn về phía trước</i> <i>Giỏi về toán học</i>

REFERENCES

- [1] Quang-Ban Diep. 2005. *Vietnamese grammar*. Education Publishing House.
- [2] DCS Center – LacViet Corp. 2011. *Vietnamese Dictionary*.
- [3] Fei Xia. 2000. *The part-of-speech tagging guidelines for the Penn Chinese Treebank (3.0)*.
- [4] Phe Hoang. 1998. *Vietnamese Dictionary*. Scientific & Technical Publishing.
- [5] Thai P. Nguyen, Luong X. Vu, Huyen T.M. Nguyen. 2010. *VTB part-of-speech tagging guidelines*.
- [6] Vietnam Social Sciences Commission. 1983. *Vietnamese grammar*. Social Sciences Publishers.