

# The Word Segmentation Guidelines for the Vietnamese Treebank (0.1)

Quy T. Nguyen and Yusuke Miyao

16th October 2019

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Challenges . . . . .	3
1.2	Word definition . . . . .	3
1.3	Guidelines organization . . . . .	4
<b>2</b>	<b>Word-formation and segmentation principles</b>	<b>5</b>
2.1	Single word . . . . .	5
2.1.1	Lexical word . . . . .	5
2.1.2	Function word . . . . .	5
2.1.3	Exclamation word and modifier . . . . .	6
2.1.4	Pronoun . . . . .	6
2.1.5	Other . . . . .	6
2.2	Multiple syllable word . . . . .	7
2.2.1	Lexical word . . . . .	7
2.2.2	Function word . . . . .	17
2.2.3	Exclamation word and modifier . . . . .	17
2.2.4	Pronoun . . . . .	17
2.3	Other types of word . . . . .	18
2.3.1	Proper noun . . . . .	18
2.3.2	Chemical symbols . . . . .	22
2.3.3	Mathematical formula . . . . .	22
2.3.4	Idiom and locution . . . . .	22
2.3.5	Foreign word . . . . .	22
2.3.6	Cardinal number . . . . .	23
2.3.7	Cardinal number with unit noun . . . . .	23
2.3.8	Quantity nouns with tmesis factors . . . . .	23
2.3.9	Ordinal number . . . . .	23
2.3.10	Time . . . . .	24
2.3.11	Abbreviation and acronym . . . . .	26
2.3.12	Name of law, resolution, agreement, and document . . . . .	26
2.3.13	Punctuations . . . . .	26
2.3.14	Phone number, fax, e-mail address, website, nick name, or home number . . . . .	26
2.3.15	New words/terms . . . . .	26
<b>A</b>	<b>List of categorization nouns</b>	<b>27</b>

<b>B</b>	<b>Rules of the combination of Sino-Vietnamese elements and Vietnamese origin words</b>	<b>28</b>
B.1	Sino-Vietnamese elements are prefixes . . . . .	28
B.2	Sino-Vietnamese elements are suffixes . . . . .	30
<b>C</b>	<b>List of idioms</b>	<b>31</b>

# Chapter 1

## Introduction

### 1.1 Challenges

The task of WS has become complicated due to several characteristics of Vietnamese that include no explicit word delimiters or inflectional morphemes. Unlike Western languages, in which blank spaces denote word delimiters, blank paces in Vietnamese not only play the roles of word delimiters but also those of syllable delimiters. Furthermore, unlike English and Japanese, Vietnamese is not an inflectional language, in which morphological forms can provide useful clues for word segmentation, part-of-speech tagging, and bracketing. While similar problems also occur with Chinese [4], it may be more difficult to annotate Vietnamese words because the modern Vietnamese writing system is based on Latin characters, which represent its phonology but not its semantics. As a result, Vietnamese has many polysemous expressions, i.e., terminologies that have the same surface structure but different interpretations.

Difficulties of the Vietnamese word segmentation are also caused by conflicting word definitions among linguists. For example, the two expressions of *cá<sub>fish</sub> rô<sub>anabas</sub> {anabas}* and *bệnh<sub>illness</sub> sởi<sub>measles</sub> {measles}* have the same construction, which is a combination of a categorization noun<sup>1</sup> and a specific noun. However, *cá\_ rô* is treated as a compound word while *bệnh sởi* is treated as two single-syllabic words [5, 7].

### 1.2 Word definition

Our purpose for word segmentation is to build a treebank for Vietnamese. Therefore, we consider a word as the smallest syntactic unit having a complete meaning and preventing syntactic rules from analyzing word structure. On the basis of this word definition, we propose the following rules to solve the difficulties in Vietnamese word segmentation:

Words are the most basic units of a treebank [1], and the first step in the annotation process is to define words. We adhered to the word definitions provided by Di Sciullo and Williams [1] in this research. A word is specifically defined as the smallest syntactic unit that conveys complete meanings. Moreover, no syntactic rules can be applied to a word to analyze its structure. On the basis of this definition, we propose nine rules to address the difficulties faced in Vietnamese word segmentation:

---

<sup>1</sup>Categorization nouns indicate general entities, such as *cá<sub>fish</sub>* and *cây<sub>tree</sub>*.

1. If A and B<sup>2</sup> have different meanings and the meaning of the combined form (A\_B) is different from the split form (A B), we select the form that has a more appropriate meaning for the context.
2. If A and B have different meanings and A\_B has the same meaning as A or B, A\_B is a compound word.
3. If A and B have the same meaning, A\_B is treated as a compound word.
4. If another syllable can be inserted between A and B, A and B are words.
5. If A or B (or both A and B) is a bound syllable, A\_B is treated as a compound word.
6. For expressions composed by a categorization noun (A) and other words (B), if B indicates something different from what the expression indicates, A\_B is treated as a compound word. In contrast, if B has a similar meaning to A B, A and B are treated as two words.
7. An expression composed by one or more Sino-Vietnamese syllables and an original Vietnamese word is not treated as a word when the Sino-Vietnamese syllables are the elements used to create similar words, such as antonyms and hyponyms.
8. Special classifier nouns<sup>3</sup> are treated as single-syllabic words.
9. If reduplicative words or reiterative forms are comprised of two syllables, they are treated as single-syllabic words. In cases they have more than two syllables, if we can insert a comma or a conjunction between the syllables, we treat each syllable as a single-syllabic word. Otherwise, the whole expression is considered to be a word.

Details in using these rules to recognize the words in texts will be described in the following chapters.

### 1.3 Guidelines organization

Our guidelines are organized according to the word types so that searching a word type is fast as possible. For example, we go to the personal name section to find how to annotate a personal name. For each word type, we present the rules to recognize, the word types which it is ambiguous with and the way to distinguish, examples, list of exceptions and list of closed words (if any).

---

<sup>2</sup>Without loss of generalization, we assume the expression we want to segment is A B, where A and B can be syllables or words.

<sup>3</sup>Classifier nouns indicate two types of entities: animate and inanimate things. Special classifier nouns, e.g., *sự-ing/-ion/-ity/...*, *việc-ing/-ion/-ity/...*, and *nhà-er/-or*, have also been considered to be classifier nouns by Nguyen et al. [6]. However, the collocation of special classifier nouns is different from that of classifier nouns such as *cái* and *con*. Classifier nouns are placed before categorization nouns to indicate animate (*con cá<sub>fish</sub>*) and inanimate entities (*cái bàn<sub>table</sub>*). While special classifier nouns play the same role as affixes in English, they can combine with a verb, such as *việc-ion lựa\_chọn<sub>to select</sub> {selection}*, or an adjective, such as *sự-ity nhập\_những<sub>ambiguous</sub> {ambiguity}*.

# Chapter 2

## Word-formation and segmentation principles

### 2.1 Single word

Each single word includes one syllable having lexical meaning (lexical/content word) or grammatical meaning (function word).

#### 2.1.1 Lexical word

A single lexical word is composed of only one syllable, has a complete lexical meaning and nominative function, and can be sentence constituent on its own.

Examples: *cha* {father}<sup>1</sup>, *thầy* {teacher}, *ăn* {eat}, *hát* {sing}, *đẹp* {nice}, *dày* {thick}, *quần* {trousers}, *áo* {shirt}, etc.

**Caution:**

A single lexical word can be a component of the compound word. To distinguish them, we base on the following rule: If A and B have different meanings and the meaning of the combination form (A\_B) is different from the split form (A B), we select the form that has a meaning more appropriate for the context. Otherwise, when the meanings of combination form and split form are the same, A and B should be treated as single words.

Example:

*Tiền của cô ấy.* => *tiền của* are the single words which means *money of*

*Cô ấy có nhiều tiền\_của.* => *tiền\_của* is a compound word which means *fortune*

*Cô ấy mở cửa hàng quần\_áo.* => *quần\_áo* is a compound word which means *clothes*

*Bộ đồ này phối màu quần áo khác nhau.* => *quần áo* are the single words which means *trousers and shirt*

#### 2.1.2 Function word

A monosyllabic function word can be an adjunct, a conjunction, or a preposition that is composed of only one single syllable, does not have a complete lexical meaning and nominative function, can not be sentence constituent on it own, and expresses grammatical relationships with other words.

Examples: *sẽ* {will}, *vừa* {just}, *vẫn* {still}, *của* {of}, *bằng* {by}, *vì* {because; because of}, *nếu* {if}, *tuy* {although}, etc.

---

<sup>1</sup>Translation for a Vietnamese word or sentence is given in curly brackets following the raw text.

### **Caution:**

When 2 or more monosyllabic function words are adjoining, It is ambiguous to distinguish between a sequence of single words and a multiple syllable function words. See Section 2.2.2 to know how to distinguish.

## **2.1.3 Exclamation word and modifier**

Exclamation word or modifier specifies the attitude or mood of the speaker. It does not have a complete lexical and grammatical meaning and can not be sentence constituent on its own.

Examples: *à, ư, nhỉ, nhé, ời, hử, sao, à, ối, thế, nào, đâu, vậy, chính, riêng, etc.*

## **2.1.4 Pronoun**

A pronoun is used to point/indicate person/thing. It has a complete lexical meaning and nominative function. It can be sentence constituent on its own. There are monosyllable pronouns as follows: *tôi I, họ they, hắn he, ai who, gì what, này this/ these, đó that/ those, ấy that/ those, etc.*

## **2.1.5 Other**

Classifier nouns (including inanimate classifier nouns, such as *cái, áng, chiếc, quyển, bức, cuốn*, animate classifier noun *con*, and classifier nouns denoting person such as *con, ông, bà, chú, gã, tên, cô, chú*) combine with other nouns (categorization nouns or common nouns) can not form compound words because the later part can stand alone without classifier noun (high popularity and loose combination).

Examples:

*Cái bàn table, con vịt {duck}, cái kiến {ant}, ông bảo\_vệ {doorkeeper}, chị thủ\_thư {librarian}, etc.*

Expression of a special classifier noun and an adjective or verb to identify or to name a person or thing, in which the special classifier noun plays the role as an affix, can not be treated as a compound word. Examples show in Tables 2.1 and 2.2.

<b>Noun</b>	<b>Verb</b>	<b>Noun hrase</b>
cái	nhìn {look}	cái nhìn {insight/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 {favour/understand thoroughly}	điều tâm_đắc {favourite thing}
niềm	tin {believe/ trust}	niềm tin {belief}
nỗi	buồn {sad}	nỗi buồn {melancholy}
vụ	tham_ô {embezzle}	vụ tham_ô {embezzlement}

Table 2.1: Examples shows expressions of special classifier nouns and verbs that cannot be treated as compound words.

Noun	Adjective	Noun phrase
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}
nỗi	vất_vả {hard}	nỗi vất_vả {hardship}
vụ	rắc_rối {complicated}	vụ rắc_rối {complication}
điều	đơn_giản {simple}	điều đơn_giản {simple thing}

Table 2.2: Examples shows expressions of special classifier nouns and adjectives that cannot be treated as compound words.

## 2.2 Multiple syllable word

Multiple syllables word refers to word formed by combining more than one syllable. The multiple syllable words are classified into the following types.

### 2.2.1 Lexical word

#### 2.2.1.1 Equal compound word

Principle to recognize: A and B can be single words. However, meaning of combination form (A\_B) is different from splitting form (A B).

Examples:

*quần {trousers}, áo {shirt} => quần\_áo {clothes}*  
*ruột {gut}, thịt {flesh} => ruột\_thịt {kinship}*  
*chèo {row}, chống {prop} => chèo\_chống {manage}*

The words are constructed as follows.

- A and B have different meanings and meaning of A\_B is wider than or more indefinite than meanings of A and B. For examples, *đất {ground}* and *nước {water}* => *đất\_nước {country}*, *ăn {to eat}* and *uống {to drink}* => *ăn\_uống {to eat and drink/junket/diet/eating}*, etc.
- A and B have different meanings and A\_B has the same meaning as A or B. For examples, *ăn {to eat}* and *nói {to speak}* => *ăn\_nói {to speak}*, *đi {to go}* and *đứng {to stand}* => *đi\_đứng {to move}*, etc.
- A and B are antonyms and meaning of A\_B adds up meanings of A and B. For examples: *hên\_xui {lucky-unlucky} => luck*, *nóng\_lạnh {hot-cold} => to catch a chill*, *sống\_chết {alive-die} => no matter what*, etc.
- A, B, and A\_B have the same meaning. For example, *tìm {find}* and *kiếm {find}* => *tìm\_kiếm {to find}*. Origin of the syllables can be:
  - One Vietnamese element and one Sino-Vietnamese element. For examples, *núi\_non*, *xương\_cốt {bones}*, *chia\_ly {separate}*, *lính\_tráng {soldier}*, etc.



- Both are Sino-Vietnamese elements. For examples, *vĩ\_dại* {great}, *kiến\_thiết* {construct}, *tâm\_hồn* {soul}, etc.
- Both are Vietnamese elements. For examples, *đợi\_chờ* {wait}, *đổi\_thay* {change}, etc.
- One general and one local language element. For examples, *nóng\_nực* {hot}, *bông\_hoa* {flower}, etc.
- Meanings of A and B are close. We can use A or B to replace A\_B. For example, *kéo*{drag} and *đẩy*{push} => *kéo\_đẩy* {to find} is a word.

### **Cautions:**

If we can add the comma or the conjunctions “và” , “hoặc”, or “hay” between A and B that meaning of the sentence does not change, we should treat A and B as single syllable words.

Example:

*Quần\_áo* {clothes} *luộm thuộm* {The clothes are untidy}

*Bộ đồ này phối quần áo (quần và áo/quần, áo) khác màu* {Color of trousers and shirt is blended different}

### **2.2.1.2 Subordinate compound word (SCW)**

Each word includes a principal element and a dependent element. We base on the following criteria to recognize this word type:

- Meaning of combination form is different from the split form. For examples, in the word *chân\_vịt* {crew propeller}, *chân* {leg} is the main element and *vịt* {duck} is the dependent element. If we split *chân\_vịt* into 2 single words, this expression means *legs of a duck*.
- Meaning of dependent element is indefinite if it is not combined with the principle element. For example, in *đen\_đúa*, we cannot understand what does *đúa* means if it is treated as a single word.
- Dependent element (B) indicates something different from what A\_B indicates. For examples, in *cá\_heo* {dolphin}, *cá* {fish} that is a categorization noun is the principal element and *heo* that is a specific noun is dependent. If we treat *heo* as a single word, its meaning is *pig*. But in this case, *heo* is name of a fish species. Therefore, *heo* should be combined with *cá* to identify its meaning exactly. For *xe\_lam* {three-wheels taxi}, *lam* means *deep blue* if it is treated as a single word.
- We cannot insert another syllable (conjunction) between principal syllable and dependent syllable.

Subordinate compound words are classified into the following types.

### **Categorization noun (A) + noun/verb/adjective (B)**

Categorization nouns<sup>2</sup> are nouns indicating general entities. A categorization noun can be modified by a specific noun or a verb/adjective that express the use, purpose, quality,

Categorization noun	Dependent element	Compound word
cá {fish}	heo {pig}	cá_heo {dolphin}
hoa {flower}	đồng_tiền {coin/money/currency}	hoa_đồng_tiền {gerbera}
nước {water}	hoa {flower}	nước_hoa {perfume}
nước {water}	hàng {goods}	nước_hàng {a kind of spice making by suger}
nước {water}	ngọt {sweet}	nước_ngọt {freshwater}

Table 2.3: Examples shows expressions of categorization nouns and dependent element that are treated as compound words.

or manner. This combination will be treated as a compound word if B express a thing different from the thing A\_B express. Examples are showed in Table 2.3.

More examples about the combination of categorization noun and noun/verb/adjective to create SCWs are showed in Table 2.4.

For the cases that meaning of B is similar to A\_B, and B can be used to replace A\_B, A\_B should not be treated as SCW.

Examples:

*ung\_thư {cancer} => bệnh ung\_thư {cancer}*

*tâm\_thần {mental} => bệnh {illness} tâm\_thần {mental}*

*Hoa {flower} hồng {rose}, hoa cúc, quả/trái {fruit} xoài {mango}, quả/trái ổi, quả/trái mận, quả/trái măng\_cụt, cá {fish} bống {goby}, cá chép {carp}, cá chạch, cá rô {anabas}, cá lia\_thia, rau cải, rau dền, rau diếp\_cá, etc.*

If meaning of A\_B is similar to meaning of A B, this sequence should be splitted. In this case, we can insert other element between A and B without changing of the meaning. It is occur frequently for B indicating materials.

Examples:

*Thịt bò => thịt của con bò*

*Nhà tranh => nhà bằng tranh*

*Nồi đồng => nồi bằng đồng*

*Nồi đất => nồi bằng đất*

*Bò sữa => bò lấy sữa*

**Caution:**

The following examples show that insertion of the word causes the change of meaning of the SCW. These examples also indicate that B are usually verbs or adjectives expressing the sense, the use, or the quality of the priority components.

*Áo\_chòàng: áo rộng, dài đến đầu gối unlike Áo\_để\_chòàng: áo gì cũng được, để khoác tạm bên ngoài*

*Cà\_tím: cà quả to và dài, màu tím hoặc nửa tím nửa trắng unlike Cà\_màu\_tím: cà gì cũng được, màu tím*

## Noun + noun

Meaning of A\_B is different from the meaning of A B.

Examples:

---

<sup>2</sup>List of categorization nouns is given in appendix.

Categorization noun	SCW
Hoa	hoa_mép_dê, hoa_môm_chó, hoa_đồng_tiền
Cá	cá_đôi, cá_thu, cá_hồi {salmon}, cá_chim, cá_quả, cá_voi, cá_hường, cá_chai, cá_bột
Rau	rau_má, rau_ngót, rau_đắng, rau_thơm, rau_tập_tàng
Máy	máy_bay {airplane}, máy_bơm {pump}, máy_tiền {lathe}, máy_in {printer}, máy_gặt_đập, máy_quay, máy_vi_tính
Nồi	nồi_cơm_điện, nồi_hầm, nồi_hấp, nồi_hơi, nồi_áp_suất
Nhà	nhà_thương {hospital}, nhà_hàng {restaurant}, nhà_hát {theatre}, nhà_xuất_bản
Cà	cà_tím, cà_pháo, cà_chua
Áo	áo_choàng, áo_khoác, áo_cưới

Table 2.4: More examples show the combination of categorization noun and noun/verb/adjective.

Meaning of the syllable *con* in the compound word *con\_gái* {girl} is different from meaning of the word *con*{child} in the expression *con gái* {daughter}

#### Noun + adjective

Meaning of A\_B is different or wider than the meaning of A B.

Examples:

*đỉnh\_cao* {peak/pinnacle} unlike *đỉnh* {top} *cao* {high}

*điểm\_nóng* {hot spot/ flashpoint} unlike *điểm* {point/mark} *nóng* {hot}

*áo\_dài* {Vietnamese dress/tunics} unlike *áo* {shirt} *dài* {long}

#### Caution:

Distinguishing between noun phrase and SCW having the same structure, noun come before adjective: If meaning of A\_B is exactly the meaning of A B, we should select the splitting form.

Examples:

*người\_tốt* {kind man}, *kẻ\_xấu* {bad man}, *nét\_đẹp*, *vẻ\_đẹp* {beauty}, *mùi\_thơm*, *hương\_thơm*, etc.

#### Adjective + noun

Meaning of A\_B is different or wider than meaning of A B.

Examples:

*nhANH\_trí* {quick-mind => quick-witted}, *nhANH\_mắt* {quick-eyes => quick-eyed/quick-sighted}, *vui\_tính* {funny-personality => playful}, *vui\_tai* {funny-ear => funny to hear}, *vui\_mắt* {funny-eye => funny to see}, *vui\_miệng* {funny-mouth => funny for mouth}, *đẹp\_đôi* {nice-couple => well-matched}, *đẹp\_trai* {nice-man => handsome}, *mát\_tay* {cool-hand => skillfull}, *to\_gan* {big-liver => daring}, *tốt\_bụng*, *tốt\_tính*, etc.

#### Caution:

Distinguishing between phrase and SCW having the same structure, adjective comes before noun: If meaning of A\_B is exactly the meaning of A B, we should select the splitting form.

Examples:

*nhieu\_người*, *cùng\_ngành*, *ngắn\_hạn*, *giỏi\_văn*, etc.

### Adj + syllable denoting level

Meanings of B is indefinite if we treat it as a single word.

Examples:

*đen\_sì {very black}, đen\_thui {very black}, đen\_ngòm {deep black}, đen\_đúa, đen\_đuôi, đen\_nhẽm, đen\_gìòn, đen\_hắc, trắng\_phau {spotlessly white}, trắng\_toát {very white}, xanh\_ngắt {very green/deep blue}, đỏ\_au {bright and cheerful red}, nhạt\_phèo {extremely dull/insipid}, rỗng\_tuếch {completely empty, shallow}, etc.*

#### **Caution:**

*ngọt thanh* and *ngọt gắt* are not SCWs because *thanh* and *gắt* are adjectives in Vietnamese.

### Verb + noun

Meaning of A\_B is different from the meaning of A and B. In this combination, meaning of A is wide and indefinite.

Examples:

*làm\_duyên {do-charm => to charm}, chia\_tay {separate-hand => to take leave of somebody/to separate}, làm\_chứng {do-witness => to witness}, làm\_dáng {do-figure => posture}, làm\_mai/làm\_mối {do-matchmaker => to act as a matchmaker}, làm\_việc {work-something => work/question/contact/grill}, làm\_dâu {do-bride => be a bride}, ăn\_ảnh, dọn\_đường, làm\_ruộng, etc.*

#### **Caution:**

Distinguishing between phrase and SCW having the same structure, verb comes before noun: if meanings A\_B and A B are the same, we should select the splitting form. In this case, the meanings of the verb and the noun are definite and they can be treated as the single words.

Examples:

*làm\_móng {cutting nails}, trồng\_rau {grow vegetables}, đánh\_bài {play cards}, etc.*

### Verb + adjective

In this word type, meaning of the SCW is different from the meaning of its elements.

Examples:

*làm\_cao {do-high => put on superior airs}, làm\_đẹp {do-beautiful => aestheticize}, làm\_dữ {do-wicked => kick up a fuss}, làm\_giàu {do-rich => get rich}, làm\_khó {do-difficult => create difficulties}, đánh\_bóng {do-shiny => wax}, etc.*

### Sino-Vietnamese subordinate compound word

The expressions including Sino-Vietnamese elements will be treated as compound words when if they satisfy at least one of the following characteristics:

- There is an element which its meaning is indefinite if it separate from the other element. For example, for *hiệu\_trưởng*, *hiệu* cannot be treated as a word, so we treat *hiệu\_trưởng* as a compound word.
- Meaning of the combination form is different from splitting form. For example, *bác\_học {scholarly}* is unlike *bác\_học {paternal uncle studies}*

More examples show the expressions including Sino-Vietnamese elements which are treated as compound words as follows: *tổ\_quốc, độc\_lập, tự\_do, sản\_xuất, phát\_triển,*

*thể\_sự, giai\_nhân, phụ\_nữ, áp\_lực, thuyết\_mình, phát\_động, đồng\_bào, đồng\_chí, đồng\_hương, quân\_sự, đại\_diện, etc.*

**Caution:**

The Sino-Vietnamese elements<sup>3</sup> 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}, etc.* combine with the Vietnamese words to create the new words, such as *nguyên\_cứu {research} => nghiên\_cứu viên {researcher}*. These Sino-Vietnamese elements are treated as words.

**Examples:**

*phó\_tổng\_giám\_độc {deputy general director}, tân\_thủ\_tướng {prime minister}, phó\_thủ\_tướng {deputy prime minister}, nguyên\_chủ\_nhiệm {former chairman}, trưởng\_khoa {head of department}, huấn\_luyện\_viên {coach}, diễn\_viên {actor, actress}, giáo\_viên {teacher}, vô\_điều\_kiện {unconditional}, bất\_lịch\_sự {impolite}, bất\_bình\_đẳng {injustice}, siêu\_lợi\_nhuận {super profit}, tiểu\_sản\_xuất {small production}, [tiền Seagames]W pre-Sea Games, hạ\_Sahara {sub-Saharan or south of the Sahara}, chi\_cục\_trưởng, nguyên\_phó\_chủ\_tịch, công\_nghiệp\_hoá, viện\_trưởng, phó\_trưởng\_công\_an, ngọt\_hoá, siêu\_xe, bất\_chính {unrighteous}, giáo\_viên, giảng\_viên, tốp\_ca, đảng\_viên, đoàn\_viên, hội\_viên, toán\_học, vật\_lý\_học, etc.*

**Exceptions:**

- Idioms are treated as the words, such as *bất\_di\_bất\_dịch {immutable}*.
- Expressions having 3 or more syllables in which the combination of main factors do not form a word will be treated as words. For examples: *bán\_cầu\_não* (*cầu* and *não* do not form a word), *đa\_dạng\_sinh\_học* (*đa*, *dạng* and *sinh* do not form a word), *địa\_hoá\_học* (*địa* and *hoá* do not form a word), *đa\_thần\_giáo* (*thần* and *giáo* do not form a word), etc.

## Categorization noun + specific noun + other word

These expressions are used to name somethings. We do not treat them as compound words.

**Examples:**

*cá\_rô\_đồng, cá\_voi\_xanh, hoa\_cúc\_dại, máy\_bay\_trực\_thăng\_helicopter, máy\_in\_kim {dot-matrix printer}, máy\_in\_dập {impact printer}, máy\_xay\_sinh\_tổ, etc.*

**Caution:**

*máy\_bay\_bà\_già {cougar woman}*: 1 word

*máy\_bay\_bà\_già {morane fighter plane}*: 2 words

*phi\_công\_trẻ {young pilot}*: 2 words

*phi\_công\_trẻ {a man who love an older woman}*: 1 words

### 2.2.1.3 Reduplicative word, reduplicative form and reiteration

#### a. Reduplicative word

Some compound words are constructed from the phonetic repetition phenomenon of the syllables. Double reduplicative word constitutes the majority of phonetic repetition compound word. In addition, there are also triple and fourfold reduplicative word.

##### a1. Double reduplicative word

Principles to recognize:

---

<sup>3</sup>Refer to appendices for a fixed list of Sino-Vietnamese elements.

- All elements have no meaning or vague meaning. For examples, *long\_lanh* {glistening}, *chập\_chờn* {glimmering}, *bồn\_chồn* {disturbed}, *thướt\_tha* {graceful}, *lững\_thững* {amble}, etc.
- One element has meaning, and the other has not. The meaningful syllable can stand in front or behind the other syllable. For examples,
 

*sạch* {clean} + *sẽ* => *sạch\_sẽ* {clean},  
*lanh* {vivacious} + *lợi* => *lanh\_lợi* {vivacious},  
*vui* {cheerful} + *vẻ* => *vui\_vẻ* {cheerful},  
*vội* {hasty/hurried} + *vàng* => *vội\_vàng* {hasty/hurried},  
*lạnh* {cold} + *lẽo* => *lạnh\_lẽo* {cold and cheerless},  
*nhớ* {miss} + *nhung* => *nhớ\_nhung* {miss},  
*nhỏ* {small/little} + *nhắc* => *nhỏ\_nhắc* {petite},  
*lẻ* {lone/odd} + *loi* => *lẻ\_loi* {lone},  
*nhỏ* {small} + *nhen* => *nhỏ\_nhen* {illiberal},  
*vẽ* {paint} + *vờ* => *vẽ\_vờ* {conceive/ contrive},  
*loanh* + *quanh* {round} => *loanh\_quanh* {wander},  
*tưng* + *bùng* {flare up/burn} => *tưng\_bùng* {jubilant},  
*hống* + *hách* {imperious} => *hống\_hách* {imperious}, etc.
- The syllables have the same meaning. However, meaning of combination form is different from splitting form. For examples,
 

*xanh* {blue/green} => *xanh\_xanh* {bluish/greenish},  
*hao* {waste} => *hao\_hao* {look slightly like}, etc.

There are 2 sub-types of repetition compound word according to component of syllable to be repeated:

#### 1. Partial repetition:

- First consonant repetition. For examples, *lanh\_lợi* {vivacious}, *sạch\_sẽ* {clean}, *vui\_vẻ* {cheerful}, *nơm\_nớp* {apprehensive}, etc.
- Vowel repetition. For examples, *bồi\_hồi*, *càu\_nhàu*, *khéo\_léo* {clever}, *lúng\_túng*, etc.
- First consonant and vowel repetition (tones are difference). For examples, *cỏn\_con* {tiny}, *đỏ\_đỏ* {red}, *tím\_tím* {violet}, etc.
- First consonant and vowel repetition (tones and last consonants are difference). For examples, *đẹp\_đẹp* {beautiful}, *bạc\_bạc* {silvery}, *tốt\_tốt* {well}, etc.

#### 2. Full word repetition: Adjectives are repeated. The repetition lightens meaning of the word. For examples, *hao\_hao* {alike}, *đùng\_đùng* {a sound}, *róc\_róc* {murmur} *vàng\_vàng* {slightly yellow}, *thấp\_thấp* {slightly short}, etc.

#### a2. Triple reduplicative word and fourfold reduplicative word

In this kind of compound word, there is at least one syllable that cannot stand alone and has no meaning.

Examples:

*đứng\_đứng\_đứng, sạch\_sạch\_sạch {completely empty}, khít\_khít\_khít {close-fitting}, nhũn\_nhũn\_nhũn, lơ\_lơ\_lơ, hớn\_ha\_hớn\_hở {radiant/merry}, lờ\_lờ\_lờ\_lờ {slovenly}, ấm\_à\_ấm\_úc, đủng\_đả\_đủng\_đĩnh, long\_la\_long\_lanh, nhí\_nha\_nhí\_nhảnh, khít\_khít\_khít\_khín, hăm\_hăm\_hở\_hở, bồi\_hồi\_bồi\_hồi, etc.*

### Caution:

- There are expressions that syllables appear like repetition of each other. However, they are independent words and should be segmented separately. For examples, *đi thi {go for an examen}, ra xa {go further}, etc.*
- Some compound words look like phonetic repetition compound words. However, their syllables are meaningful. For examples, *tươi {fresh} + tốt {good} => tươi\_tốt {verdant}, mặt {face} + mũi {nose} => mặt\_mũi {look}, etc.*
- Name of somethings, which are compound words, look like reduplicative word. For examples, *đu\_đu, chuồn\_chuồn, châu\_châu, ba\_ba, su\_su, cào\_cào, thần\_lần, thường\_luồng, etc.*

## b. Reduplicative form

Reduplicative forms are treated as compound words. Their constructions are as follows.

- A\_B\_B is created on the basis of the subordinate compound word A\_B, where B that modifies A about degree is fully repeated. The examples are showed in Table 2.5.
- A\_B\_B' is created on the basis of the subordinate compound word A\_B, where B that modifies A about degree is partially repeated. The examples are showed in Table 2.6.

SCW	Reduplicative form
đỏ_lòm	đỏ_lòm_lòm
xanh_lè	xanh_lè_lè
đen_sì	đen_sì_sì
nông_choèn	nông_choèn_choèn
tối_om	tối_om_om

Table 2.5: Examples show reduplicative forms created by fully repeating B.

SCW	Reduplicative form
đỏ_hoét	đỏ_hoén_hoét
cao_ngheù	cao_lêu_ngheù
nông_choèn	nông_choèn_choét

Table 2.6: Examples show reduplicative forms created by partially repeating B.

## c. Reiteration

Reiteration looks like full reduplicative word. However, reiteration is used when we want to emphasize something. Double reiterations are treated as words. However, the reiterations having more than two words are not treated as words.

- Double reiteration (A A). For examples, *nhà\_nhà, người\_người, ngành\_ngành, sáng\_sáng, tối\_tối, etc.*
- Triple reiteration (onomatopoeia having 3 syllables) (A A A). For examples, *âm\_âm\_âm, ha\_ha\_ha, etc.*
- Fourfold reiteration:
  - A A B B: *quần\_quần áo\_áo, tầng\_tầng lớp\_lớp, cười\_cười nói\_nói, trùng\_trùng điệp\_điệp, buồn\_thiu buồn\_thiu, đi\_đi lại\_lại, lên\_lên xuống\_xuống, ra\_ra vào\_vào, sắc\_sắc không\_không, etc.*
  - A B A C: *vui nhà\_vui cửa, đẹp người\_đẹp nét, về lâu\_về dài, chạy tới\_chạy lui, chẳng\_nói\_chẳng\_rằng, người\_này\_người\_nọ, etc.*

**Caution:** Distinguish between reduplicative word and reiteration

- Onomatopoeia having 2 syllables is treated as word, for examples, *ào\_ào, ắc\_ắc, ha\_ha, bộp\_bộp, khặc\_khặc, etc.* However, onomatopoeia having 3 syllables is not treated as word, for examples, *ào\_ào\_ào, ha\_ha\_ha, cộp\_cộp\_cộp, cóc\_cóc\_cóc, etc.*
- Fourfold reiteration is unlike fourfold reduplication:
  - Fourfold reiteration: repetition of a compound word. For examples, *đi\_lại => đi\_đi lại\_lại, buồn\_thiu => buồn\_thiu buồn\_thiu, etc*
  - Fourfold reduplication: repetition of a double reduplicative word. For example, *vội\_vàng => vội\_vội\_vàng\_vàng*

#### 2.2.1.4 Other compound words

The rest of compound words are falling in this category where the elements do not have any phonetic or semantic relationship. They can be Vietnamese origin word, Chinese transcription, or English or French loanwords.

Examples:

Vietnamese origin word: *bồ\_hóng {soot}, bù\_nhìn {dummy/ puppet}, cổ\_hủ {conservative}, mè\_nheo {fussy}, ba\_láp {talkative}, ba\_hoa {talkative}, etc.*

Chinese transcription: *mật\_chước {mah-jong}, xì\_dầu {soya sauce}, etc.*

French loanword: *xà\_phòng {French: savon; English: soap}; xi\_măng {F: ciment; E: cement}, búp bê {F:Poupee; E:doll}, cà\_phê {F: café; E: coffee}, etc.*

English loanword: *oxy {oxygen}, internet, etc.*

#### 2.2.1.5 Blending word or phrase

##### a. Blending word

The words are created by combining elements of some compound words. There are types of blending word as follows.



- Combination of first syllables of each compound word.

– A\_B + C\_D => A\_C. For examples,  
*công\_nhân {workers} + nông\_dân {peasants} => công\_nông {workers and peasants},*  
*tươi\_đẹp + sáng\_sủa => tươi\_sáng,*  
*khoa\_học + giáo\_dục => khoa\_giáo,*  
*thiếu\_niên + nhi\_đồng => thiếu\_nhi,*  
*văn\_học + nghệ\_thuật => văn\_nghệ*

– A\_B + C\_D + E\_F = A\_C\_E. For examples,  
*công\_nhân {workers} + nông\_dân {peasants} + binh\_lính {soldiers} => công\_nông\_binh {workers, peasants and soldiers}*

- Combination of last syllables of each compound word. For example,  
*nhân\_dân + đại\_biểu => dân\_biểu.*
- Combination of the first syllable of the first compound word and the last syllable of the last compound word. For example,  
*công\_nhân {workers} + viên\_chức {official} => công\_chức {civil servants}.*
- Combination of the last syllable of the first word and the first syllable of the last word. For examples,  
*nhân\_dân + vận\_động => dân\_vận*  
*nhân\_dân + trí\_thức => dân\_trí*
- Combination of the first word and the first syllable of the last word. For examples,  
*công\_nhân {workers} + viên\_chức {official} => công\_nhân\_viên {employee}*
- Combination of first syllable of the first word and the last word (A\_B + C\_B => A\_C\_B). For examples,  
*y\_sĩ + bác\_sĩ => y\_bác\_sĩ*  
*ưu\_điểm + nhược\_điểm => ưu\_nhược\_điểm*  
*huân\_chương + huy\_chương => huân\_huy\_chương*  
*công\_nghiệp + nông\_nghiệp => công\_nông\_nghiệp*  
*thanh\_niên + thiếu\_niên => thanh\_thiếu\_niên*
- A\_B + C\_B + D\_B = A\_C\_D\_B. For example,  
*nông\_nghiệp + lâm\_nghiệp + ngư\_nghiệp => nông\_lâm\_ngư\_nghiệp*

### **Caution:**

There are some cases that A B and C B are phrases, however, the blending is a word. For example,

*khám\_bệnh + chữa\_bệnh => khám\_chữa\_bệnh*

## **b. Blending phrase**

There are 3 cases of blending phrase as follows.

- $A\_B + A\_C \Rightarrow A\ B\ C$

For example, *đài\_phát\_thanh + đài\_truyền\_hình*  $\Rightarrow$  *đài\_phát\_thanh\_truyền\_hình*

- $A\_B + C\_B + E\_F \Rightarrow A\_C\_B\ E\_F$

For example, *thanh\_niên + thiếu\_niên + nhi\_đồng* = *thanh\_thiếu\_niên nhi\_đồng*

### 2.2.2 Function word

An expression including more than one syllables which plays the role as an adjunct, a conjunction, or a preposition will be treated as multiple syllable function word if meaning of splitting form is different from combination form.

Examples: *cho\_nên, chẳng\_những, nói\_chung, nói\_riêng, hình\_như, dường\_như, bỗng\_dưng, thậm\_chí, bất\_cứ, cực\_kì, xung\_quanh, etc.*

#### **Caution:**

Some conjunctions going with pronouns such as: *vì\_thế, vì\_vậy, vì\_đó, bởi\_thế, bởi\_vậy, bởi\_đó, do\_vậy, do\_đó, dù\_vậy, dù\_thế, chẳng\_những\_thế, không\_những\_thế, không\_chỉ\_thế, mặc\_dù\_thế, etc.* like multiple syllable function words. However, they should be splitted because their constructions are not stable (can be analyzed). We can insert other words between syllable without changing of the meaning.

Examples:

*vì\_thế*  $\Rightarrow$  *vì\_như\_thế*

*do\_đó*  $\Rightarrow$  *do\_điều\_đó*

### 2.2.3 Exclamation word and modifier

There are the words used to specify the attitude or mood of the speaker that include more than one syllable. Where, the syllables can have meaning or not.

Examples:

*ui\_da, à\_ha, chao\_ôi, chẳng\_hạn, etc.*

There are some cases that meaning of combination form is different from splitting form.

Examples:

*Ông quá trời\_ơi. {He is very unbearable.}*

*Ông trời\_ơi. {Hey, God}*

### 2.2.4 Pronoun

There are some pronouns having more than one syllable, such as *chúng\_ta, chúng\_tôi.*

## 2.3 Other types of word

### 2.3.1 Proper noun

#### 2.3.1.1 Person name

Person name is treated as a word. For examples, *John\_Smith*, *Hồ\_Chí\_Minh*, *Mao\_Trạch\_Dông*, *Vích-to\_Huy-gô*, etc.

**Caution:**

1. The common nouns, numerals or titles, that combines with proper names to form a popular name or nickname will be segmented separately. For examples,  
Titles: *Trạng Quỳnh*, *Nghè Tân*, *Đồ Chiểu*, *Thủ\_khoa Huân* {First class Huan},  
*Bà Trưng* {Mrs Trung}, *Cụ Phan* {Mr Phan}, *Bác Hồ* {Uncle Ho}, *Hồ Chủ\_tịch*  
{President Ho}, *Minh\_Tri Thiên\_hoàng* {Mikado Minh Tri}, *Hoa\_Nhị phu\_nhân*  
{Madame Hoa Nhi}, *Nam\_Phương Hoàng\_hậu*, *Ngọc\_Hân Công\_chúa*, etc.  
Numbers combining with proper names to form a popular name or nickname will be  
treated as words: *Một\_Hiên* {second lieutenant Hien}, *Năm\_Tôn* {Colonel Ton},  
*Ba\_Giai* (Ba is a number), *Tám\_Danh* (Tám is a number), etc.
2. Descriptive factors (person's appearance or location) go with the names to address  
someone. These words will be segmented separately. For examples,  
*Năm Sài\_Gòn* {Nam Saigon},  
*Mình sẹo* {Mình scar},  
*Xuân tóc đỏ* {Xuan red hair},  
*Ba Cầu Bông* {Ba Bong Bridge},  
*Ba chữ\_kí* {Ba signature}, etc
3. Alias/nickname will be splitted from the real proper name. For examples, *Ốc Thanh\_Vân*  
{Thanh Van Snail}, *Công\_chúa bong\_bóng Bảo\_Thy*, etc.

#### 2.3.1.2 Person name with title or professional designator

Person name is separated from the title or professional designator.

Examples:

*PGS. TS. Vũ\_Văn\_An*,  
*Bác\_sĩ Vũ\_Văn\_An*,  
*Bs. Vũ\_Văn\_An*,  
*Phó\_hiệu\_trưởng Vũ\_Văn\_An*,  
*HT. Vũ\_Văn\_An*,  
*Bà chủ\_tịch Nguyễn\_Thị\_Bình*,  
*Lão Trương*, etc.

#### 2.3.1.3 Location name

Proper nouns indicating location name are treated as the words. For examples, *Thái\_Bình\_Dương*  
{Pacific Ocean}, *Việt\_Nam* {Vietnam}, *Hà\_Nội*, *Nội\_Bài*, *Trường\_Tiền*, *Mát\_xơ\_va*  
{Moscow}, *Mạc\_Tư\_Khoa* {Moscow}, etc.

The common nouns or administrative unit nouns come before location proper names  
are splitted from location proper names.

Examples:

*sông river Mississippi,*  
*sân\_bay Tân\_Sơn\_Nhất {Tan Son Nhat airport},*  
*nhà\_thờ {basilica} Đức\_Bà {Notre-Dame},*  
*công\_viên văn\_hoá Đàm\_Sen {Dam Sen Cultural Park},*  
*châu {continent} Á {Asia},*  
*thành\_phố {city} Hồ\_Chí\_Minh,*  
*Tp. {city} Hồ\_Chí\_Minh,*  
*quận {district} 10,*  
*Q. {district} 10,*  
*số nhà address 448/12A,*  
*số 10 - 12 Đinh\_Tiên\_Hoàng,*  
*khu\_du\_lịch {tourism area} Suối\_Tiên,*  
*ngoại\_ô {suburb} Paris, etc.*

**Caution:**

- An expression of directional words that is a location name will be treated as a word. For examples,  
*vùng Tây\_Bắc {Northwest Region of Vietnam},*  
*vùng Đông\_Bắc {Northeast Region of Vietnam},*  
*khu\_vực Bắc\_Mỹ*
- Directional words forming location names will be combined. For examples,  
*bến\_xe Miền\_Đông {Mien Dong bus station/ Eastern Coach Station},*  
*bến\_xe Miền\_Tây {Mien Tay bus station/ Western Coach Station},*  
*Đông\_Nam\_Bộ {Southeast Region}*
- An expression of directional words that is used to indicate a direction will be treated as a word. For examples,  
*(hướng) tây\_nam {south west},*  
*(phía) tây tây\_nam ,*  
*(gió\_mùa) đông\_bắc {north east monsoon wind},*
- Directional words are used to mention the direction (directional words are not intrinsic part of a location proper name), they are splitted from location names. For examples,  
*Nam California Southern of California,*  
*Nam Ailen Southern/South of Ireland,*  
*nam (sông) Trường\_Giang,*  
*bắc Biển Đông,*  
*miền nam {Southern} Luân\_Đôn {London},*  
*phía Nam {South} Việt\_Nam {Vietnam},*  
*hướng đông {east} Nhà\_Bè {Nha Be},*  
*trung\_tâm {central} Luân\_Đôn {London},*  
*nội\_thành Đà\_Nẵng,*

*ngoại\_thành Hà\_Nội {suburb of Ha Noi},  
 vùng\_ven Hà\_Nội {Hanoi environs},  
 ngoại\_ô Tp. Hồ\_Chí\_Minh {Ho Chi Minh C. suburb},  
 Đông\_Trường\_Sơn,  
 Tây\_Trường\_Sơn*

- Directional words are used to mention part of a location/area (directional words are intrinsic part of a location proper name, or to mention a continent, or a cluster of countries), they are segmented as one word. For examples,

*Đông\_Nam\_Á {Southeast Asia},  
 Nam\_Phĩ {South Africa},  
 Nam\_Mỹ {South America},  
 Bắc\_Triều\_Tiên {North Korea},  
 An\_Phú\_Đông {An Phu Dong, Vietnam},  
 Đông\_Timor {East Timor},  
 Tây\_Australia {Western Australia},  
 Bắc\_Âu {Scandinavia/ North Europe},  
 Bắc\_Ailen {Northern Ireland},  
 Thượng\_Lào {Upper Laos},  
 Hạ\_Sahara {Lower Sahara},  
 Trường\_Sơn\_Đông,  
 Trường\_Sơn\_Tây,  
 Nam\_Triều\_Tiên*

**Caution:** Distinguish between direction and part of a location.

Examples:

Direction: phía Nam Carolina {South of Carolina}

Part of location: Nam\_Carolina South\_Carolina

- The combination of names which have no connection in meaning will be splitted. For examples,

*người Ấn - Âu,  
 kênh Nhiêu\_Lộc - Thị\_Nghè,  
 mối quan\_hệ Trung - Hàn {China - Korea relationship},  
 các đảo Trường\_Sa - Hoàng\_Sa {Truong\_Sa - Hoang\_Sa islands}*

#### 2.3.1.4 Organization name, names of products, and trademarks

- The common nouns (words come before) are splitted from the proper name of the organization. For examples,

*Công\_ty KMDC,  
 Khách\_sạn Caravelle,  
 Nhà\_xuất\_bản Trẻ, etc.*

- Foreign name: The organization name in foreign languages is treated as a word. For examples,

*New\_York\_Times,  
 Công\_ty Petronas\_Carigali {Petronas Carigali company},  
 Angel\_Beauty,*

*Harper's\_Bazaar,*  
*White\_Palace,*  
*Diamond\_Plaza,*  
*Sai\_Gon\_Times, etc.*

- Vietnamese name:

- The sequence of organization name that all elements are proper name will be combined. For examples,

*Công\_ty Nguyễn\_Văn\_Trỗi Nguyen Van Troi company,*  
*Nhà\_hàng Ngọc\_Sương,*  
*Công\_ty Vinh\_Thủy, etc.*

- The combination of names is treated as one word. For examples,

*Hội\_Việt\_Mỹ,*  
*Bệnh\_viện Pháp-Việt,*  
*Hội hữu\_nghị Việt\_-\_Ý*

- The sequence of organization name that one or more elements are not proper names will be splitted. For examples,

*Trường\_Dại\_học Công\_nghệ Thông\_tin {University of Information Technology},*  
*Bộ Kế\_hoạch và Dầu\_tư,*  
*Báo Tiếp\_thị và Gia\_đình,*  
*Báo Tuổi\_Trẻ {Youth newspaper}*  
*Nhà\_hàng Hải\_sản Ngọc\_Sương {Ngoc Suong Seafood Restaurant},*  
*Quán\_ốc Quang\_Anh,*  
*Trường\_Dại\_học Hoa\_Sen,*  
*Nhóm nhạc Mặt\_Trời\_Đỏ*  
*Công\_ty TNHH Hoa\_Mai*  
*Quán\_cơm\_tám Thuận\_Kiều*  
*Nhà\_máy thủy\_điện Hoà\_Bình, etc.*

- Translated name is splitted. For examples,

*Thời\_báo New\_York {New York Times}*

- The name that is mixed foreign name and Vietnamese name will be splitted. For example,

*Thuận\_Kiều Plaza*

- Transcription is treated as a word. For examples, *Mạc\_Tư\_Khoa, Mát\_xơ\_va, etc.*

### 2.3.1.5 Proper names including number

- Ordinal number will be separated from proper name. For examples,

*Vua Henry II {Henry II King},*  
*Giáo\_hoàng John\_Paul II,*  
*Cầu Bình\_Triệu 2,*  
*Bệnh\_viện Nhi\_Đông 1 {Children's Hospital No.1},*  
*Công\_ty Cienco 6, etc.*

- Abbreviated phrases, such as: *U18*, *U50*, etc. are treated as words.
- Class name is a word. For examples, *lớp 6A*, *lớp 12C*, etc.

### 2.3.2 Chemical symbols

Chemical symbols are segmented as a word.

Examples:

*CH<sub>4</sub>*, *CO*, *HCl*, *Fe*, etc.

### 2.3.3 Mathematical formula

A mathematical formula is segmented as a word.

Examples:

*2x*, *2\*x*, *2\_x\_3*, *2x\_+\_3y\_=\_50*, etc.

A quantity phrase indicating fraction (the tmesis mark is “ / ”) that is written in numbers will be treated as a word, such as: *3/4*. However, it will be separated if it is written in letters, such as: *ba phần tư*.

### 2.3.4 Idiom and locution

Segmentation of idioms and locutions follows the following rules:

If idioms and locutions are entries in the idiom list (refer to appendices), they are segmented as one word. For examples, *rán\_sành\_ra\_mỡ*, *ăn\_cây\_nào\_rào\_cây\_đó*, etc.

Idioms and locutions which are sentences, such as proverbs and folk-songs, will not be segmented as one word. For example: *Đêm tháng năm chưa nằm đã sáng. Ngày tháng mười chưa cười đã tối*. will not be considered as one word.

The phrases in which the Sino factors dominate will be also grouped as one word. For examples, *tứ\_nhiếp\_pháp*, *thiên\_long\_bát\_bộ*, *kiến\_văn\_tiểu\_lục*, etc.

### 2.3.5 Foreign word

Foreign words are words taken from foreign languages (including languages of the minorities) and adopted into Vietnamese language. For examples, *internet*, *email*, *polymer*, *design*, *vak\_wei*, *drôk\_tue*, *Bih\_KJhông*, *video\_game*, etc.

The segmentation of foreign words follows the following rules:

When the name of a substances, illness, animal, and plant, program, etc. is written in foreign language, they will be treated as a word. For examples, *hydro*, *natrium*, *Asian\_Song\_Festival*, *Aie\_confiance\_mon\_amour*, *Woman's\_Day*, *Christmas\_Day*, etc.

When a foreign word is transliterated, they are treated as Vietnamese words. For examples, *cà\_phê*, *a\_xít*, *a\_pa\_tít*, etc.

When foreign country/city names are translated into Vietnamese, they will also be treated as one word. For examples, *Hoa\_kỳ* {US}, *Nhật\_Bản* {Japan}, etc.

When the foreign words indicating different things, they are segmented separately. For examples *Album vol 1*, *Album Missing\_You*, etc.

### 2.3.6 Cardinal number

A cardinal number is a word. For examples, *30*, *30,5*, etc.

Cardinal numbers indicating precise, imprecise, or approximate quantities written in letters are segmented separately.

Examples:

*ba mươi*, *ba mươi phẩy năm*, *3 triệu*, *hai chục*, *5 trăm*, *vài chục ngàn*, *trăm triệu*, *mấy tỉ*, *nửa tá*, *nửa chục*, *một nửa tá*, etc.

### 2.3.7 Cardinal number with unit noun

Unit noun is splitted from number.

Examples:

*2 \$*, *2 cm*, *2 đô-la*, *2 độ\_C*, *hai đô-la*,

*200 ngàn/người/tháng*,

*200 ngàn một người một tháng {VND200,000 per person per month}*,

*20 lít mỗi ngày {20 liters a day}*, etc.

**Caution:**

In the case that unit noun and number is not splitted by blank space, we add a blank space between unit noun and number.

Examples:

*2\$ => 2 \$*

*2cm => 2 cm*

For the expression that unit noun and number is not splitted by blank space, and have a slash between unit noun and number, we add blank spaces before and after that slash.

Example:

*200.000 / người/tháng* (3 words)

Symbol % does not treated as unit noun. An expression including number and % is treated as a word, for example *2%*, *2\_ %*. However, % written in letters will be splitted from number, for example *hai phần\_ trăm*, *2 phần\_ trăm*.

If the unit noun includes an expression, it is not treated as a word.

Example:

*2,2 tỷ đồng/căn nhà -> 2,2 tỷ đồng / căn nhà*

*1 lít sữa/ngày -> 1 lít sữa / ngày*

### 2.3.8 Quantity nouns with tmesis factors

Football score will be treated as a word: *2:3*, *2-3*, *2\_ :\_ 3*, *2\_ -\_ 3*, etc.

Size of a photo will be treated as a word: *(hình) 4x6*, *4\_x\_6*, etc.

Expressions denote from ... to ... will be annotated separately: *từ 2 - 3*, etc.

### 2.3.9 Ordinal number

An expression indicating ordinal number is not treated as a word. For examples: *thứ nhất*, *thứ hai*, *thứ ba*, *thứ tư*, *thứ một ngàn*, *thứ 1000*, etc.



### 2.3.10 Time

- Nouns of time are the words. For examples, *giây* {second}, *phút* {minute}, *giờ/tiếng* {hour}, *ngày* {day}, *tuần* {week}, *tháng* {month}, *quý* {quarter}, *năm* {year}, *thứ* {day of week}, *thế\_kỉ* {century}, *thiên\_niên\_kỉ* {millennium}, *buổi*, *mùa*, *Xuân*, *Hạ*, ect.
- The following types of expression which include two or more nouns of time are not treated as words.
  - Half a day. For example, *buổi tối* {evening}, etc.
  - Season. For example, *mùa hè* {summer}, etc.
  - Special holidays or festivals. For example, *ngày quốc\_khánh* {Independence day}, etc.
  - Age. For example, *thời Phục\_hưng* {Renaissance}, etc.
- For an expression indicating time point, time period, or measure of time in which has numbers and nouns of time, the numbers are separated from the nouns of time.

Examples:

- Time points:
  - 8 giờ sáng* {8 am},
  - năm 1987* {year 1987},
  - năm một nghìn chín trăm tám bảy* {year 1987},
  - tháng 1*,
  - tháng một*,
  - tháng 12*,
  - tháng mười hai*,
  - thứ 2*,
  - thứ hai* {Monday},
  - ngày 6*,
  - ngày sáu* {sixth},
  - thế\_kỉ 20*,
  - thế\_kỉ hai mươi*,
  - thế\_kỉ XX* {the 20th century}, etc.

**Caution:**

*chủ nhật* {Sunday}, *tháng\_giêng* {January}, and *tháng\_chạp* {December}, are words.

- Period of time:
  - Từ tháng 1 đến tháng 6*
- Measure of time:
  - 8 giờ* {in 8 hours},
  - hai giờ ba mươi phút* {2 hours and a half},
  - bốn mươi hai giây* {forty two seconds},
  - 6 ngày*,
  - hai năm* {two years}, etc.

- A phrase of time including giờ {hour}, phút {minute}, giây {second} written in short form is a word. For examples, 10g, 10\_g {10 h}, 30\_p {30 m}, 30' {30 m}, 45\_ " {45 s}, 10g30 {10 h 30}
- A time sequence including numbers and symbols will be treated as a word. For examples,  
 30-04  
 30/04/1975  
 30-04-1975  
 6/7/2011  
 30\_-\_04  
 30\_/\_04\_/\_1975  
 30\_-\_04\_-\_1975  
 1987\_-\_2013  
 6\_-\_7/11  
 6\_-\_7\_/\_11

If it is written by letters or letters and numbers, it should be splitted. For examples,  
*ngày ba mươi tháng tư năm một nghìn chín trăm bảy mươi lăm,*  
*từ ngày 10 đến ngày 15, etc.*

*sáng 19 tháng 4 {in the morning of 19th April}*  
*29 tháng bảy năm 2005 {29th, September, 2005}*  
*ngày 30-04*  
*ngày 30\_-\_04*  
*ngày 6 và 7/11*  
*ngày 6 và 7\_/\_11*  
*từ ngày 10-15*

**Caution:** Expressions denote from year to year, from day to day, ect. will be not treat as words.

*1987 - 2013*

- In some cases, the adjunct *qua* and *tới* follow nouns of time. They should be splitted from the nouns of time. For examples, *ngày qua {last day}*, *năm qua {last year}*, *tuần qua {last week}*, *tháng qua {last month}*, *năm tới {next month}*, *tuần tới {next week}*, *tháng tới {next month}*, etc.
- Words, such as *nay*, *này*, *rồi*, *mai*, *trước*, *sau*, *kia*, *kìa* often follow nouns of time. The should be splitted from the nouns of time. For examples, *chiều nay {this evening}*, *hè này {this summer}*, *tuần rồi {last week}*, *sáng mai {tomorrow morning}*, *bữa trước {previous day}*, *tháng sau {next month}*, *năm kia {the year before last}*, *hôm kia {three days ago}*, *mai sau*, *mai kia*, *mai này*, etc.

**Caution:**

There are some expressions that nouns of time come before *nay*, *mai*, etc. but they are considered as words because we can not insert any word into these expressions. For examples, *hôm\_qua*, *hôm\_nay*, *ngày\_mai*, *ngày\_nay*, *hiện\_nay*, *hôm\_mai*, *hôm\_sớm*, etc.

### 2.3.11 Abbreviation and acronym

Abbreviation of proper names will be treated as a word.

Examples:

*HCM {Ho Chi Minh}, KHTN, ĐHQG, KHXH&NV, etc.*

Title or professional designator will be separated from proper name.

Examples:

*TP. HCM {Ho Chi Minh city}, TP HCM, GS. TS. Nguyen\_An, ĐH KHTN, ĐHQG TP. HCM, ĐH KHXH&NV, etc.*

For an expression that is partially abbreviated, abbreviation will be separated from full word. For example, *Đại\_học CNTT*, etc.

Person names are always treated as words. For example, *Bà L.T.M\_Lê*, etc.

### 2.3.12 Name of law, resolution, agreement, and document

Name of law, resolution, agreement, document, etc. which includes number and acronym is treated as one word, such as: *Nghị\_định 43/2009ND-CP*.

### 2.3.13 Punctuations

Punctuation is treated as one word.

Examples:

*2, 3 người*

*Cô H. ( nhân viên mới ) nói : “ Tôi biết vụ việc này ” .*

### 2.3.14 Phone number, fax, e-mail address, website, nick name, or home number

Phone number, fax, e-mail address, website, nick name (chat), or address is segmented as one word.

Examples:

phone number: *090312345*, etc.

Website: *http://uit.edu.vn*, etc.

E-mail address: *xyz@yahoo.com, abc@gmail.com, etc.*

Nick name (chat): *lu\_khach\_co\_don*, etc.

Home number: *55/12K, K15/49, 10-12, etc.*

### 2.3.15 New words/terms

The expressions that appear in text but are not covered by guidelines should be discussed in the group for how to treat them.

# Appendix A

## List of categorization nouns

No.	Catgorization noun	Combine	Split
1	Bàn	bàn_cờ	bệnh ung_thư cá chép {carp} cây tre {bamboo} chim chào_mào {bulbul} hoa lan quả mít {jack-fruit}, trái cam {orange} rau mồng_tơi
2	Bánh	bánh_tráng {rice paper}	
3	Bệnh	bệnh_kín	
4	Bếp	bếp_dầu	
5	Cá	cá_chim	
6	Cà	cà_pháo	
7	Cây		
8	Chân	chân_trời	
9	Chim	chim_sâu	
10	Cối	cối_xay	
11	Dưa	dưa_chuột	
12	Hoa/bông	hoa_mười_giờ	
13	Máy	máy_bay {airplane}	
14	Mũ/Nón	mũ_bảo_hiểm	
15	Nhà	nhà_hát {theatre}	
16	Nồi	nồi_cơm_điện	
17	quả/trái		
18	Rau	rau_má	
19	Rượu	rượu_vang {wine}	
20	Xe	xe_ngựa	

Table A.1: Examples show treatments of the expressions including a categorization nouns and an other word

# Appendix B

## Rules of the combination of Sino-Vietnamese elements and Vietnamese origin words

### B.1 Sino-Vietnamese elements are prefixes

Showned in Table B.1 and B.2.

No.	Rules	Examples
1	bách + N = N	bách bệnh
2	bách + N = A	bách chiến, bách thắng
3	bán + N = N	bán bình_nguyên, bán cầu
4	bán + A = A	bán tự_động
5	bản + N = N	bản toạ, bản hiệu
6	bất + A = A	bất bình_đẳng, bất khả_thi
7	bất + V = V	bất bạo_động, bất hợp_tác
8	bất + N = N	bất động_sản, bất phương_trình
9	bất + V = R	bất luận
10	bổn + N = N	bổn toạ
11	chủ\_nghĩa + N/A = N	chủ_nghĩa tư_bản
12	chư + N = N	chư vị, chư huynh
13	cố+ N = N	cố thủ_tướng
14	cự+ N = N	cự giám_đốc
15	đa + N = N	đa phương_tiện, đa khoa
16	đại + N = N	đại bản_doanh, đại bộ_phận
17	đồng + N = N	đồng chủ_tịch
18	đồng + N/A = A	đồng nghĩa
19	đồng + V = V	đồng học
20	đồng + N = R/A	đồng thời
21	hậu + N = N	hậu tố, hậu hoạ

Table B.1: Some rules of combination of Sino-Vietnamese elements and other words

No.	Rules	Examples
23	hậu + V = V	hậu kiểm, hậu chi
24	hiếu + V = A	hiếu học
25	hiếu + N = A	hiếu khách
26	hữu + N = A	hữu hình, hữu thần
27	hữu + V = A	hữu dụng
28	lão + N = N	lão nghệ_nhân, lão ông
29	liên + N = N	liên bang, liên bộ
30	nam + N = N	nam thanh_niên, nam công_nhân
31	nguyên + N = N	nguyên thủ_tướng, nguyên trưởng phòng
32	nữ + N = N	nữ công_nhân, nữ nghệ_sĩ
33	phản + N = A	phản tác_dụng, phản khoa_học
34	phi + N = A	phi chính_phủ, phi lợi_nhuận
35	phó + N = N	phó chủ_nhiệm, phó viện trưởng
36	siêu + N = N	siêu giai_cấp, siêu lợi_nhuận
37	siêu + V = V	siêu dẫn
38	siêu + A = A	siêu thực
39	tái + V = V	tái cơ_cấu, tái đầu_tư
40	tiền + N = N	tiền lệ, tiền đồn
41	tiền + N/V = A	tiền khởi_nghĩa
42	tiền + A = A	tiền tư_bản_chủ_nghĩa
43	tiểu + N = N	tiểu bang, tiểu vương_quốc
44	tổng + N = N	tổng số
45	tổng + V = V	tổng chi, tổng kiểm_kê
46	trưởng + N = N	trưởng ban, trưởng phòng
47	vi + N = N	vi sinh_vật
48	vô + N = A	vô chủ, vô nhân_đạo
49	vô + V = A	vô can, vô học
50	vô + V = R	vô kể, vô luận
51	vô + A = R/A	vô cùng

Table B.2: Some rules of combination of Sino-Vietnamese elements and other words

## B.2 Sino-Vietnamese elements are suffixes

Showed in Table B.3.

No.	Rules	Examples
1	$N + \text{bộ} = N$	đảng bộ, huyện bộ
2	$A/N + \text{chủ\_nghĩa} = A$	tư_bản chủ_nghĩa
3	$N + \text{hoá} = V$	lão hoá, trí_thức hoá
4	$N + \text{học} = N$	tâm_lí học, toán học
5	$N + \text{kiểu} = N$	Ấn kiểu, Việt kiểu
6	$V + \text{phó} = N$	chỉ_huy phó
7	$N + \text{tính} = N/A$	nhạc tính, kịch tính
8	$N + \text{trưởng} = N$	đại_đoàn trưởng, phân_viện trưởng
9	$V + \text{trưởng} = N$	chỉ_huy trưởng
10	$V + \text{viên} = N$	cộng_tác viên, lập_trình viên
11	$N + \text{viên} = N$	công_an viên
12	$A + \text{viên} = N$	chuyên viên

Table B.3: Some rules of combination of Sino-Vietnamese elements and other words

# Appendix C

## List of idioms

Ác_cái_là	Đổ_mò_hôi_sôi_nước_mắt	Nhỏ_tận_gốc
Ác_giả_ác_báo	Đồ_nát_đụng_nhau	Như_ai
Ác_một_cái_là	Đỏ_như_đồng_hun	Như_bóng_với_hình
Ai_đời	Đo_ni_đóng_giày	Như_cơm_bữa
Am_thanh_cảnh_vắng	Dở_ông_dở_thằng	Như_điều_gặp_gió
Ăn_Bắc_mặc_Kinh	Đổ_sông_đổ_biển	Như_đỉnh_đóng_cột
An_bần_lạc_đạo	Dở_trắng_dở_đen	Như_hình_với_bóng
Ăn_bằng_nói_chắc	Đoái_công_chuộc_tội	Như_mắc_cửi
Ăn_bánh_trả_tiền	Đoán_già_đoán_non	Như_muối_bỏ_bể
Ăn_báo_cô	Đoàn_kết_một_chiều	Như_nước_vỡ_bờ
Ăn_bình_bất_động	Độc_mồm_độc_miệng	Như_nước_với_lửa
Ăn_bơ_làm_biếng	Độc_nhất_vô_nhị	Như_ong_vỡ_tổ
Ăn_bờ_ở_bụi	Đổi_cho_sạch_rách_cho_thơm	Như_rắn_mất_đầu
Ăn_bớt_ăn_xén	Đội_nón_ra_đi	Như_tay_với_chân
Ăn_bớt_đội_nói_bớt_lời	Đối_nhân_xử_thế	Như_ý_sở_cầu
Ăn_bữa_hôm_lo_bữa_mai	Đồi_phong_bại_tục	Nhũn_như_con_chi_chi
Ăn_cần_ở_kiệm	Đời_thuở_nhà_ai	Nhường_cơm_sẻ_áo
Ăn_cấp_ăn_này	Đổi_trắng_thay_đen	Nhứt_bổn_vạn_lợi
Ăn_cấp_như_ranh	Đội_trời_đạp_đất	No_bụng_đối_con_mất
Ăn_cây_nào_rào_cây_nấy	Đơm_đó_ngọn_tre	No_cơm_ấm_áo
Ăn_cay_nuốt_đắng	Đơn_thương_độc_mã	No_dồn_đổi_góp
Ăn_cây_táo_rào_cây_sung	Đòn_xóc_hai_đầu	Nộ_hoả_nhai_đầu
Ăn_chắc_mặc_bền	Đồng_bệnh_tương_lân	Nộ_khí_xung_thiên
Ăn_cháo_đá_bát	Đồng_cam_cộng_khổ	Nở_mặt_nở_mày
Ăn_chay_nằm_đất	Đồng_chua_nước_mặn	Nở_mày_nở_mặt
Ăn_chay_nằm_mộng	Đóng_cửa_bảo_nhau	Nội_bất_xuất_ngoại_bất_nhập
Ăn_chay_niệm_Phật	Động_địa_kinh_thiên	Nói_bóng_gió
Ăn_chưa_no_lo_chưa_tối	Đồng_hội_đồng_thuyền	Nói_bóng_nói_gió
Ăn_chực_nằm_chờ	Đồng_không_mông_quạnh	Nói_cạnh_nói_khoé
Ăn_chung_máng_ở_chung_chuồng	Đông_như_trảy_hội	Nói_cho_phải
Ăn_chung_ở_lộn	Động_phòng_hoa_chúc	Nói_có_sách_mách_có_chứng
Ăn_cổ_di_trước_lợi_nước_di_sau	Đồng_ra_đồng_vào	Nội_công_ngoại_kích
Ăn_cơm_chúa_múa_tối_ngày	Đồng_sàng_dị_mộng	Nói_của_đáng_tội
Ăn_cơm_có_canh_tu_hành_có_bạn	Đồng_sinh_đồng_tử	Nồi_da_nấu_thịt
Ăn_cơm_lừa_thóc_ăn_cóc_bỏ_gan	Đồng_tâm_hiệp_lực	Nồi_da_xáo_thịt
Ăn_cơm_mới_nói_chuyện_cũ	Đồng_tâm_nhất_trí	Nồi_đỉnh_đám

Table C.1: List of idioms.



Ăn cơm nhà vác tù và hàng tổng	Đông tây kim cổ	Nổi đình nổi đám
Ăn cơm thiên hạ	Đồng tịch đồng sàng	Nói đúng ra
Ăn cơm trước kẻng	Đồng trắng nước trong	Nói gần nói xa
An cư lạc nghiệp	Độp một cái	Nói gì thì nói
Ăn dầm nằm dề	Đốt cháy giai đoạn	Nói giáo cho giặc
Ăn dất nằm sương	Đốt đặc cán mai	Nói hươu nói vượn
Ăn dói mặc rách	Đột từ nóc	Nói khan nói vãi
Ăn dọi nằm chờ	Đột từ nóc đột xuống	Nói một đằng làm một nẻo
Ăn đời ở kiếp	Du thủ du thực	Nói ngon nói ngọt
Ăn đơm nói đặt	Đũa móc chòi mâm son	Nói nhăng nói cuội
Ăn dong ở dợ	Đùa với lửa	Nổi như cồn
Ăn dung ngồi rồi	Dục bất khả tuyệt	Nói phải giữ lấy lời
Ăn gan uống máu	Dục nước béo cò	Nói rã bọt mép
Ăn giả làm thật	Dục tốc bất đạt	Nói ra nói vào
An giấc nghìn thu	Dùi đục chấm mắm cáy	Nói sùi bọt mép
Ăn giập miếng trầu	Đùng có trách	Nổi tam bánh
Ăn gió nằm mưa	Dung dăng dung dẻ	Nói thánh nói tướng
Ăn gỏi nằm nhờ	Đùng một cái	Nói thánh tướng
Ăn hại đái nát	Đứng mũi chịu sào	Nói toạc móng heo
Ăn hiền ở lành	Đứng ngồi không yên	Nổi trận lôi đình
Ăn hơn nói kém	Đứng như trời trồng	Nói trộm bóng
Ăn hương ăn hoa	Đứng núi này trông núi nọ	Nói trộm vía
Ăn không ăn hỏng	Đứng sững như trời trồng	Nói trước quên sau
Ăn không ngồi rồi	Dững tóc gáy	Nói vãi bọt mép
Ăn không nói có	Dựng tóc gáy	Nói xa nói gần
Ăn lông ở lỗ	Dụng văn hoá nguyệt	Non sông gấm vóc
Ăn mắm mút dòi	Dựng vợ gã chồng	Nóng như lửa đốt
Ăn mận khát nước	Được (một) cái (là)	Nữ công học hội
Ăn mật trả gừng	Được ăn cả ngã về không	Nữ tướng dương môn
Ăn mày cầm tinh bị gậy	Được bữa nào xào bữa ấy	Nửa doi nửa chuốt
Ăn mày đòi xôi gấc	Được chảnh hay chó	Nửa đời nửa đoạn
Ăn mày quen ngô	Được đằng chân lân đằng đầu	Nửa đường đứt gánh
Ăn miếng trả miếng	Được một cái	Nửa nạc nửa mỡ
Ăn nên đòi nói nên lời	Được voi đòi tiên	Nửa úp nửa mở
Ăn nên làm ra	Đường ai nấy đi	Nước chảy bèo trôi
Ăn ngay nói thẳng	Đường cung yển	Nước chảy chỗ trũng
Ăn nhờ ở đậu	Đường đi lối lại	Nước chảy đá mòn
Ăn như mổ khoét	Đường đi nước bước	Nước đến chân mới nhảy
Ăn no ngủ kỹ	Dương đông kích tây	Nước đổ đầu vịt
Ăn no vác nặng	Dương dương tự đắc	Nước đổ lá khoai
Ăn ốc nói mò	Đường ngang lối tắt	Nước đổ lá môn
Ăn phải đũa	Đứt đuôi con nòng nọc	Nước đục thả câu
An phận thủ thường	Đứt gánh giữa đường	Nước lã ra sông
Ăn quả nhớ kẻ trồng cây	Đứt ruột đứt gan	Nước lợ cơm niêu
Ăn quả vả trả quả sung	Ế xưng ế xía	Nước mắt cá sấu
Ăn sóng nói gió	Ếch chết tại miệng	Nước mắt chảy xuôi
Ăn sóng nuốt tươi	Ếch ngồi đáy giếng	Nước mắt nhà tan
Ăn sung giả ngái	Ếch vô hoa mướp	Nước sôi lửa bỏng
Ăn sung mặc sướng	Êm như ru	Nước sông công lính
Ăn thật làm giả	Gà cồ mắc tóc	Nuôi báo cô
Ăn thua đủ	Gà đẻ gà cục tác	Nuôi cáo trong nhà
Ăn thùng nôi trôi rế	Gà đẻ trứng vàng	Nuôi ong tay áo
Ăn thùng uống vại	Gà què ăn quần cối xay	Nứt đố đố vách

Table C.2: List of idioms.

Ăn_to_nói_lớn	Gá_thổ_đổ_hồ	Ổ_bầu_thì_tròn_ổ_ống_thì_dài
Ăn_trắng_mặc_trơn	Gà_trống_nuôi_con	Ổ_hiền_gặp_lành
Ăn_trên_ngồi_tróc	Gác_chân_chữ_ngũ	Oan_thị_kính
Ăn_trên_ngồi_trước	Gác_tía_lầu_son	Oan_thị_màu
Ăn_trông_nồi_ngồi_trông_hướng	Gái_đi_già_mồm	Ôm_rơm_răm_bụng
Ăn_tục_nói_phét	Gãi_đúng_chỗ_ngừa	Ôn_cổ_tri_tân
Ăn_tươi_nuốt_sống	Gàn_bát_sách	Ôn_sâu_nghĩa_nặng
Ăn_tuyết_nằm_sương	Gan_chí_mề	Ôn_trả_nghĩa_đền
Ăn_vóc_học_hay	Gần_chùa_gọi_bụt_bằng_anh	Ông_ăn_chả_bà_ăn_nem
Ăn_vội_ăn_vàng	Gan_cóc_tía	Ông_bà_ông_vải
Ăn_vụng_ngon_miệng	Gần_đất_xa_trời	Ông_chẳng_bà_chuộc
Ăn_xôi_đòi_đĩa	Gạn_đục_khơi_trong	Ông_già_bà_cả
Ăn_xối_ở_thì	Gần_kề_miệng_lỗ	Ông_nói_gà_bà_nói_vịt
Anh_hùng_không_có_đất_dụng_võ	Gan_lì_tướng_quân	Ông_thối_xuôi_bà_thối_ngược
Anh_hùng_mặt_lộ	Gan_liền_tướng_quân	Ông_tơ_bà_nguyệt
Anh_hùng_nhất_khoảnh	Gần_lửa_rát_mặt	Phá_gia_chi_nữ
Anh_hùng_rơm	Gần_nhà_xa_ngõ	Phải_biết
Anh_hùng_xạ_điều	Gan_như_cóc_tía	Phải_cái
Áo_ấm_cơm_no	Gan_vàng_dạ_sắt	Phải_cái_tội
Ào_ào_như_thác_lũ	Gán_vợ_đợ_con	Phận_ẩm_duyên_ôi
Ao_có_bờ_sông_có_bến	Gạo_châu_củi_quế	Phận_hẩm_duyên_ôi
Áo_đơn_đội_hè	Gạo_chợ_nước_sông	Phận_mỏng_cánh_chuồn
Áo_đơn_lồng_áo_kép	Gạo_trắng_nước_trong	Phép_thông_công
Áo_gấm_đi_đêm	Gặp_chăng_hay_chớ	Phép_vua_thua_lệ_làng
Áo_gấm_khăn_hoa	Gấp_lửa_bỏ_bàn_tay	Phi_nội_tắc_ngoại
Áo_gấm_về_làng	Gấp_lửa_bỏ_tay_người	Phi_thành_vật_nhiều
Ao_liền_ruộng_cả	Gặp_nhau_trước_lạ_sau_quen	Phi_thương_bất_phú
Ao_sâu_nước_cả	Gắt_như_mắm_tôm	Phi_thường_hoàn_mỹ
Ao_tù_nước_đọng	Gấu_ăn_trăng	Phi_thương_vật_nhiều
Áo_vải_cờ_đào	Gãy_gánh_giữa_đường	Phòng_bệnh_hơn_chữa_bệnh
Bạ_ăn_bạ_nói	Gậy_ông_đập_lưng_ông	Phòng_đơn_gối_chiếc
Ba_bề_bốn_bên	Gậy_thù_kết_oán	Phong_gấm_lụa_là
Ba_chân_bốn_cẳng	Ghen_ăn_tức_ở	Phú_quý_sinh_lễ_nghĩa
Ba_chìm_bảy_nổi	Ghen_bóng_ghen_gió	Phú_xuân_sơn_cư_đồ
Ba_chìm_bảy_nổi	Ghét_cay_ghét_đắng	Phú_xuân_sơn_quy_đồ
Ba_chìm_bảy_nổi_chín_lênh_đênh	Ghét_như_đào_đất_đổ_đi	Phúc_bất_trùng_lai
Bà_cô_ông_mãnh	Ghi_lòng_tạc_dạ	Phun_châu_nhả_ngọc
Ba_cọc_ba_đồng	Ghi_xương_khắc_cốt	Qua_cầu_rút_ván
Ba_đầu_sáu_tay	Gì_thì_gì	Quá_đi_chứ
Ba_hoa_chích_choè	Giá_áo_túi_cơm	Quá_lắm_chỉ
Ba_hoa_thiên_địa	Giá_cầm_giả_điếc	Quá_lắm_cũng_chỉ
Ba_hoa_xích_đế	Già_chơi_trống_bởi	Quá_lừa_lỡ_thì
Ba_hồn_bảy_vía	Già_cốc_đế	Qua_ngày_đoạn_tháng
Ba_hồn_chín_vía	Già_còn_chơi_trống_bởi	Qua_sông_đim_thuyền
Ba_làng_bảy_chợ	Gia_định_thành_thông_chí	Qua_sông_lụy_đồ
Ba_mặt_một_lời	Già_đòn_non_lẽ	Qua_sông_phải_lụy_đồ
Ba_máu_sáu_cơn	Già_đòn_non_nhẽ	Quá_tội
Ba_mũi_giáp_công	Giả_đui_giả_điếc	Quắc_cần_câu
Ba_phé_ba_nơi	Già_kén_kén_hom	Quan_âm_sơn
Ba_que_xổ_lá	Già_néo_đứt_dây	Quân_hồi_vô_lệnh
Bác_cổ_thông_kim	Giả_ngõ_giả_ngọng	Quân_hồi_vô_phêng
Bạc_tình_bạc_nghĩa	Giả_nhân_giả_nghĩa	Quần_là_áo_lướt
Bách_chiến_bách_thắng	Giả_phụng_chân_hoàng	Quần_sư_quạt_mo
Bạch_diện_thư_sinh	Già_trái_non_hột	Quần_tử_nhất_ngôn
Bách_gia_chư_tử	Giải_cầu_tương_phùng	Quang_minh_chính_đại

Table C.3: List of idioms.

Bách_niên_giai_lão	Giẫm_chân_tại_chỗ	Quanh_đi_quẩn_lại
Bách_phát_bách_trúng	Giậm_chân_tại_chỗ	Quê_cha_đất_tổ
Bài_bình_bổ_trận	Giận_cá_chém_thót	Quen_hơi_bén_tiếng
Bầm_gan_tím_ruột	Giáng_long_chi_kiểm	Quý_hồ_tinh_bất_quí_hồ_đa
Bấm_không_lỗ_bổ_không_vào	Giang_sơn_gắm_vóc	Quốc_kế_dân_sinh
Bấm_ra_sữa	Giao_trứng_cho_ác	Quốc_triều_hình_luật
Bàn_cát	Giật_đầu_cá_vá_đầu_tôm	Quý_hồ_tinh_bất_quý_hồ_đa
Bàn_dân_thiên_hạ	Giật_gấu_vá_vai	Quý_như_vàng
Bán_đỏ_bán_thảo	Giấu_đầu_hở_đuôi	Quỷ_tha_ma_bắt
Bán_lai_chân_diện	Giấu_đầu_lòi_đuôi	Quyền_cao_chức_trọng
Ban_ngày_ban_mặt	Giàu_đổ_bì_m_leo	Quyền_rom_vạ_đá
Bán_nước_buôn_dần	Giàu_nứt_đổ_đổ_vách	Quyền_sinh_quyền_sát
Bàn_ra_tấn_vào	Giấy_rách_phải_giữ_lấy_lề	Quýt_làm_cam_chịu
Bán_sống_bán_chết	Giấy_trắng_mực_đen	Quýt_ngọt_cam_chua
Bàn_tay_vàng	Giẻ_cùi_tốt_mã	Ra_cái_điều
Bán_thân_bắt_toại	Gieo_gió_gặt_bão	Ra_cái_vẽ
Bán_thân_nuôi_miệng	Gieo_nhân_nào_gặt_quả_nấy	Ra_đầu_ra_đuă
Bán_thảo_cương_mục	Giơ_cao_đánh_khê	Ra_điều
Bán_tín_bán_nghi	Gió_chiều_nào_che_chiều_ấy	Ra_hồn
Bán_tổng_bán_táng	Giơ_đầu_chịu_băng	Ra_môn_ra_khoai
Bán_tổng_bán_thảo	Gió_táp_mưa_sa	Ra_ngô_ra_khoai
Bán_trời_không_vấn_tự	Giơ_tay_mặt_đặt_tay_trái	Ra_tuồng
Bán_trôn_nuôi_miệng	Gió_vào_nhà_trống	Ra_vẽ
Bán_vợ_đợ_con	Giội_gáo_nước_lạnh	Rán_sành_ra_mỡ
Bằng_chị_bằng_em	Giọt_nước_trần_ly	Rào_trước_đón_sau
Bằng_ngàn_vượt_bể	Giữ_cho_trong_âm_ngoài_êm_thuận_hòa	Rau_nào_sâu_ấy
Bằng_vai_phải_lửa	Giữ_mồm_giữ_miệng	Rày_đầy_mai_đó
Bằng_xương_bằng_thịt	Giữa_đường_đứt_gánh	Rẻ_như_bèo
Bánh_ít_đi_bánh_qui_lại	Giục_như_giục_tà	Rẽ_thuý_chia_uyên
Bạo_ăn_bạo_nói	Giương_mắt_ếch	Rỏ_rá_cạp_lại
Bảo_hoàng_hơn_vua	Góc_bể_chân_trời	Rồi_đây
Bạo_mồm_bạo_miệng	Góc_biển_chân_trời	Rối_như_canh_hẹ
Bạo_thiên_nghịch_địa	Gối_đất_nằm_sương	Rối_như_tơ_vò
Bắt_bóng_dè_chùng	Góp_gió_thành_bão	Rợn_tóc_gáy
Bắt_cá_hai_tay	Gửi_trứng_cho_ác	Rồng_đến_nhà_tôm
Bắt_chân_chữ_ngũ	Gương_võ_lại_lành	Rồng_mây_gặp_hội
Bắt_cóc_bỏ_đĩa	Hạ_hồi_phân_giải	Rũ_áo_ra_đi
Bắt_cộng_đái_thiên	Hả_lòng_hả_dạ	Rửa_tay_gác_kiểm
Bắt_đắc_kì_tử	Há_miệng_chờ_sung	Run_như_cây_sậy
Bật_đèn_xanh	Há_miệng_mắc_quai	Rừng_thiên_nước_độc
Bắt_di_bắt_dịch	Hai_bàn_tay_trắng	Rừng_vàng_biển_bạc
Bắt_khả_chiến_bại	Hai_năm_rõ_mười	Rừng_xanh_núi_đỏ
Bắt_khả_xâm_phạm	Hại_nhân_nhân_hại	Ruộng_cả_ao_liền
Bắt_khoan_bắt_nhặt	Hai_sương_một_nắng	Ruột_để_ngoài_da
Bắt_kiến_kỳ_hình	Hai_tay_buông_xuôi	Rút_dây_động_rừng
Bắt_ne_bắt_nét	Hải_trung_kim	Rút_phép_thông_công
Bắt_phân_thắng_bại	Ham_thanh_chuộng_lạ	Sắc_nước_hương_trời
Bắt_tận_tay_day_tận_trán	Hang_cùng_ngõ_hẻm	Sắc_sắc_không_không
Bất_tình_nhân_sự	Hàng_hà_sa_sổ	Sách_gối_đầu_giường
Bạt_vía_kinh_hồn	Hang_hùm_miệng_rắn	Sai_một_li_đi_một_dặm
Bật_vô_âm_tín	Hàng_tâm_hàng_sản	Sáng_bảnh_mắt
Bầu_đàn_thê_tử	Hàng_tôm_hàng_cá	Sang_sông_thì_phải_lụy_đò
Bầu_đoàn_thê_tử	Hành_tảo_thiên_quân	Sẩy_dần_tan_nghé
Bầu_rượu_túi_thơ	Hao_bình_tồn_tướng	Say_hoa_đắm_nguyệt
Bày_binh_bổ_trận	Hao_tiền_tồn_của	Say_như_điều_đổ
Bé_cái_lắm	Hậu_sinh_khả_uy	Sét_đánh_ngang_tai
Bể_hành_bể_tỏi	Hay_ăn_chóng_lớn	Siêu_sinh_tĩnh_độ
Bé_hạt_tieu	Hết_chỗ_nói	Sinh_cơ_lập_nghiệp
Bé_môn_toả_cảng	Hết_nước_hết_cái	Sinh_con_để_cái
Bé_quan_toả_cảng	Hết_ra_lửa	Sinh_li_tử_biệt
Bé_xé_ra_to	Hết_thuốc_chữa	Sinh_ly_tử_biệt
Bĩ_cực_thái_lai	Hỉ_mũi_chưa_sạch	Sinh_nghề_tử_nghiệp

Table C.4: List of idioms.

Bí_mật_nhà_nghe	Hỷ_yến	Sinh_sau_để_muộn
Bích_câu_kỳ_ngộ	Hổ_đầu_xà_vĩ	Sinh_sôi_nảy_nở
Biết_người_biết_của	Hộ_pháp_cẩn_trất	Sinh_sự_sự_sinh
Biệt_tăm_biệt_tích	Hộ_phong_hoán_vũ	Sờ_lên_gậy
Biết_thân_biết_phận	Hổ_phụ_sinh_hổ_tử	Sợ_chỉ_đỏ
Biệt_vô_âm_tín	Hổ_phụ_vô_khuyển_tử	Sợ_tốc_chê_làm_tư
Bình_an_vô_sự	Hồ_thĩ_tang_bồng	Sơn_cùng_thủy_tận
Bình_chân_như_vai	Hồ_voi_bắn_súng_sậy	Sơn_cùng_thủy_tận
Bình_cũ_rượu_mới	Hoa_cà_hoa_cải	Sơn_đầu_hoả
Bình_địa_ba_đào	Hoa_chân_múa_tay	Sơn_hào_hải_vị
Bình_hùng_tướng_mạnh	Hoa_hổ_hoa_bì_nan_hoa_cốt	Sơn_lam_chướng_khí
Bình_ngô_đại_cáo	Hoa_hoè_hoa_sói	Sống_cạn_đá_mòn
Bộ_chớ_múa_bắc	Hoa_hồ_hoa_sói	Sống_chết_mặc_bay
Bỏ_của_chạy_lầy_người	Hóa_mù_ra_mưa	Sống_có_khúc_người_có_lúc
Bỏ_hơi_tai	Hoa_tàn_nhị_rữa	Sống_dở_chết_dở
Mùi_mễ_gì	Hoa_trung_hữu_thi	Song_hỷ_lâm_môn
Mùi_tên_hòn_đạn	Hoa_vô_đơn_chí	Sống_khôn_thác_thiên
Bơ_thừa_sữa_cặn	Hoa_vô_đơn_chí	Song_kiểm_hợp_bích
Bỏ_vật_bỏ_vạ	Hoãn_bình_chi_kế	Sống_lâu_lên_lão_làng
Bờ_xôi_ruộng_mặt	Hoàng_kim_giáp	Song_phi_liên_hoàn
Bóc_áo_tháo_cây	Học_ăn_học_nói_học_gói_học_mở	Sống_tết_chết_giỗ
Bóc_ngấn_cẩn_dài	Học_một_hiểu_mười	Sóng_to_gió_lớn
Bối_bèo_ra_bộ	Học_thầy_không_tày_học_bạn	Sư_đồ_ngoạ_long
Bối_gio_trát_trầu	Hơi_đầu_mà	Sức_cùng_lực_kiệt
Bối_lông_tìm_vết	Hồi_mã_thương	Sức_dài_vai_rộng
Bối_tro_trát_trầu	Hơn_bù_kém	Sướng_như_tiên
Bối_tro_trét_trầu	Hồn_canh_hồn_cư	Sứt_đầu_mẻ_trán
Bốn_dài_hai_ngắn	Hồn_quân_hồn_quan	Suy_bụng_ta_ra_bụng_người
Bốn_phương_tám_hướng	Hồn_xiêu_phách_lạc	Tả_đột_hữu_xung
Bòn_tro_đãi_trầu	Hồng_lâu_mộng	Tả_hoả_tam_tinh
Bóng_câu_qua_cửa_sổ	Hồng_nhan_bạc_mệnh	Tả_xung_hữu_đột
Bóp_hầu_bóp_cổ	Hồng_nhan_bạc_phận	Tắc_đất_cắm_dùi
Bóp_mồm_bóp_miệng	Hợp_tình_hợp_lí	Tắc_đất_tắc_vàng
Bớt_miệng	Hư_trương_thanh_thế	Tắc_oai_tắc_quái
Bớt_mồm_bớt_miệng	Hứa_hươu_hứa_vượn	Tai_bay_vạ_gió
Bữa_cơm_bữa_cháo	Hùng_bá_thiên_hạ	Tài_nào_mà_chẳng
Bữa_đói_bữa_no	Hùng_mิêu_song_hiệp	Tại_ngoại_hậu_cứu
Bữa_đực_bữa_cái	Hương_đồng_cổ_nội	Tại_qua_nạn_khỏi
Bùn_lầy_nước_đọng	Hương_đồng_gió_nội	Tai_to_mặt_lớn
Bụng_bảo_dạ	Hút_máu_hút_mủ	Tai_trời_ách_đất
Bụng_làm_dạ_chịu	Hữu_danh_vô_thực	Tai_vách_mạch_rừng
Bụng_mang_dạ_chữa	Hữu_dũng_vô_mư	Tầm_chương_trích_cú
Buộc_chỉ_chân_voi	Hữu_sắc_vô_hương	Tam_đại_đồng_đường
Buổi_đực_buổi_cái	Hữu_sinh_vô_dưỡng	Tâm_đầu_ý_hợp
Buôn_có_bạn_bán_có_phường	Hữu_tài_vô_hạnh	Tám_đời
Buôn_dân_bán_nước	Hữu_xạ_tự_nhiên_hương	Tam_khoanh_tứ_đốm
Buôn_phấn_bán_hương	Huỳnh_đệ_tương_tàn	Tam_long_khởi_nguyên
Buôn_phấn_bán_son	Hỷ_yến	Tam_luân_cửu_chuyển
Buôn_tảo_bán_tân	Ích_kĩ_hại_nhân	Tam_nhân_thiền_thử
Buôn_thần_bán_thánh	Ích_kỷ_hại_nhân	Tam_phân_tổn_ích
Buôn_thúng_bán_bưng	Ích_nước_lợi_dân	Tâm_phục_khẩu_phục
Buôn_thùng_bán_chậu	Ích_nước_lợi_nhà	Tam_quốc_chí
Buôn_thúng_bán_mẹt	Ích_phu_vượng_tử	Tam_quốc_truyền_kì
Bụt_chùa_nhà_không_thiên	Im_hơi_lặng_tiếng	Tam_quốc_truyền_kỳ
Bút_sa_gà_chết	Im_lặng_như_tờ	Tam_sao_thất_bản
Cá_chậu_chim_lồng	Im_như_thóc	Tam_sao_thất_bổn
Cá_đối_bằng_dầu	Ít_xuyết_ra_nhiều	Tam_sơn_tứ_hải_nhất_phần_điền
Cà_kê_dê_ngỗng	Kẻ_ăn_người_làm	Tầm_sư_học_đạo
Cá_lớn_nuốt_cá_bé	Kẻ_ăn_người_ở	Tân_dòng_sông_ly_biệt
Cá_mè_một_lứa	Kẻ_cấp_gặp_bà_già	Tân_hoàn_châu_cách_cách
Cá_nằm_trên_thốt	Kẻ_ít_người_nhiều	Tân_hồng_lâu_mộng
Cà_rịch_cà_tang	Kẻ_miệng_lỗ	Tân_thiếu_lâm_tự

Table C.5: List of idioms.

Cà_riêng_cà_tỏi	Kẻ_tám_lạng_người_nửa_cân	Tàn_thủy_hử
Cả_thèm_chống_chán	Kẻ_tung_người_hứng	Tàn_cửa_nát_nhà
Cả_vú_lấp_miệng_em	Kẻ_vai_sát_cánh	Tan_đàn_xẻ_nghé
Cải_chạy_cải_cối	Kén_cá_chọn_canh	Tán_gia_bại_sản
Cải_chữa_ra_cây	Kéo_bè_kéo_cánh	Tán_hươu_tán_vượn
Cái_khó_ló_cái_khôn	Kéo_cày_trả_nợ	Tan_thành_mây_khói
Cải_lão_hoàn_đồng	Kết_cổ_ngâm_vành	Tan_xương_nát_thịt
Cải_lương_chi_bảo	Kết_tóc_xe_tơ	Tang_bồng_hồ_thĩ
Cài_răng_lược	Khắc_cốt_ghi_tâm	Táng_đờm_kinh_hồn
Cái_sảy_nảy_cái_ung	Khắc_cốt_ghi_xương	Táng_tận_lương_tâm
Cải_tà_qui_chính	Khắc_máu_tanh_lòng	Tang_thương_ngẫu_lục
Cải_tử_hoàn_sinh	Khách_ba_chúa_nhà_bảy	Tao_nhân_mặc_khách
Cầm_cân_nảy_mực	Khai_cơ_lập_nghiệp	Tắt_lửa_tối_đèn
Cầm_cân_nảy_mực	Khai_sơn_phá_thạch	Tắt_lửa_tối_đèn
Cắm_đầu_cắm_cổ	Khai_thiên_lập_địa	Tắt_nước_theo_mưa
Cắm_miệng_hến	Khấp_chợ_cùng_quê	Tẩu_hỏa_nhập_ma
Cắm_như_hến	Khất_lần_khất_lữa	Tay_bắt_mặt_mừng
Cần_cù_bù_thông_minh	Khẩu_đầu_khẩu_đuôi	Tây_du_kí
Cấn_hột_cơm_không_vỡ	Khẩu_phật_tâm_xà	Tây_du_ký
Cấn_rơm_cấn_cỏ	Khẩu_thiệt_vô_bằng	Tay_hòm_chìa_khoá
Cấn_tắc_vô_áy_náy	Khen_phò_mã_tốt_áo	Tay_làm_hàm_nhai
Cấn_tắc_vô_ưu	Khỉ_ho_cò_gáy	Tay_làm_hàm_nhai_tay_quai_miệng_trẻ
Cạn_tàu_ráo_máng	Khó_ăn_khó_nói	Tẩy_rửa_tiền
Cánh_tay_phải	Khổ_một_nỗi	Tầy_sơn_tam_kiệt
Cành_vàng_lá_ngọc	Khó_người_dễ_ta	Tay_trắng_tay_đen
Cao_bay_xa_chạy	Khô_như_ngồi	Tay_xách_nách_mang
Cao_chạy_xa_bay	Khổ_nỗi	Tề_cấp_bất_điều
Cao_lương_mĩ_vị	Khổ_rách_áo_ôm	Tề_gia_nội_trợ
Cáo_mượn_oai_hùm	Khổ_tận_cam_lai	Té_nước_theo_mưa
Cầu_bơ_cầu_bắt	Khoa_chân_múa_tay	Tề_thiên_dại_thánh
Cầu_được_uớc_thấy	Khoá_hư_lực	Tên_rơi_đạn_lạc
Cầu_toàn_trách_bị	Khóc_dở_mếu_dở	Thả_con_săn_sắt_bắt_con_cá_rô
Cây_đa_cây_đề	Khóc_đứng_khóc_ngồi	Thả_con_săn_sắt_bắt_con_cá_sộp
Cây_muốn_lặng_mà_gió_chẳng_ngừng	Khốc_như_ri	Thả_hổ_về_rừng
Cây_ngay_không_sợ_chết_đứng	Khỏe_như_voi	Thả_mồi_bắt_bóng
Cây_nhà_lá_vườn	Khỏi_phải_nói	Tha_phương_cầu_thực
Cày_sâu_cuốc_bầm	Khôn_ba_năm_dại_một_giờ	Thạch_Sùng_còn_thiếu_mẻ_kho
Cha_căng_chú_kiệt	Khốn_một_nỗi	Tham_bát_bỏ_mâm
Cha_chung_không_ai_khóc	Khôn_nhà_dại_chợ	Thâm_cẩn_cổ_đế
Cha_già_con_cọc	Khốn_nỗi	Thậm_cấp_chí_nguy
Chả_là	Không_biết_chừng	Tham_công_tiệc_việc
Cha_làm_thầy_con_đốt_sách	Không_bù	Thâm_cung_điệp_ảnh
Cha_mẹ_sinh_con_trời_sinh_tính	Không_cánh_mà_bay	Tham_đó_bỏ_đăng
Cha_nào_con_nảy	Không_đầu_không_đuă	Tham_phú_phụ_bần
Cha_sinh_mẹ_đẻ	Không_đầu_vào_đầu	Tham_quan_ô_lại
Cha_truyền_con_nói	Không_đội_trời_chung	Tham_quyền_cổ_vị
Chắc_như_đình_đóng_cột	Không_hẹn_mà_nên	Tham_sinh_ụy_tử
Chạm_như_rùa	Không_ít_thì_nhiều	Thăm_sơn_cùng_cốc
Chân_chỉ_hạt_bột	Không_kèn_không_trống	Thăm_thăng_y_truyền_kỳ_hệ_liệt
Chân_cứng_đá_mềm	Không_khảo_mà_khai	Tham_thì_thăm
Chân_đăm_đá_chân_chiêu	Không_khảo_mà_xưng	Thăm_thương_trộm_nhớ
Chán_đến_mang_tai	Không_mơ_thì_chợ_vẫn_đông	Thăm_yêu_trộm_nhớ
Chân_diện_mực	Không_một_tắc_đất_cắm_dùi	Thân_bại_danh_liệt
Chăn_đơn_gối_chiếc	Không_nhiều_thì_ít	Thân_bằng_cổ_hữu
Chân_đồng_vai_sắt	Không_nói_không_rằng	Thân_bằng_quyền_thuộc
Chân_giày_chân_đép	Không_thầy_đồ_mày_làm_nên	Thân_cô_thể_cô
Chân_lắm_tay_bùn	Không_thầy_đồ_mày_làm_nên	Thần_điều_dại_hiệp
Chân_mệnh_thiên_tử	Không_tiền_khoảng_hậu	Thần_giao_cách_cắm
Chán_mớ_đời	Khua_chiêng_gỗ_mỡ	Thần_hồn_nát_thần_tính
Chân_nam_đá_chân_chiêu	Khua_môi_múa_mép	Thần_làm_tội_đời
Chân_nam_đá_chân_xiêu	Khuôn_vàng_thước_ngọc	Thần_tàn_ma_dại
Chân_son_mình_rỏi	Khuynh_gia_bại_sản	Than_thân_trách_phận
Chân_tơ_kẻ_tóc	Khuynh_thể_hoàng_phi	Than_vấn_thờ_dài
	Kí_cả_hai_tay	Thắng_ba_ngày_tám

Table C.6: List of idioms.

Chân_trời_góc_bể	Kì_hoa_dị_thảo	Thăng_cánh_cò_bay
Chân_trong_chân_ngoài	Kì_phùng_dịch_thủ	Thăng_như_ruột_ngựa
Chân_ướt_chân_ráo	Kiểm_phong_kim	Thăng_quan_tiền_chức
Chân_yếu_tay_mềm	Kiến_bò_miệng_chén	Thăng_ruột_ngựa
Chẳng_chóng_thì_chầy	Kiến_giả_nhất_phận	Thanh_cung_tuyệt_luyện
Chẳng_dầu_vào_dầu	Kiến_văn_tiểu_lục	Thành_dầu_thổ
Chẳng_là	Kim_chi_ngọc_diệp	Thanh_mai_trúc_mã
Chẳng_mấy_chốc	Kim_chung_tráo	Thành_tám_thành_món
Chẳng_mấy_khi	Kim_lăng_thập_tam_thoa	Thanh_thiên_bạch_nhật
Chẳng_mấy_nổi	Kín_cổng_cao_tường	Thanh_xà_bạch_xà
Chật_như_nêm	Kín_như_bưng	Thao_binh_luyện_tướng
Châu_chầu_đá_voi	Kinh_bang_tế_thể	Tháo_cũi_sổ_lồng
Châu_chầu_đá_xe	Kinh_cung_chi_diệu	Tháo_cũi_xổ_lồng
Chạy_đôn_chạy_đáo	Kinh_hồn_bạt_vía	Thao_thao_bất_tuyệt
Chạy_long_tóc_gậy	Kính_lão_đắc_thọ	Thấp_cổ_bé_hong
Chạy_ngược_chạy_xuôi	Kính_nhi_viên_chi	Thấp_cổ_bé_miếng
Cháy_nhà_ra_mặt_chuột	Kinh_tâm_động_quỷ	Thập_dại_ác_nhân
Chạy_như_cờ_lông_công	Kinh_thiên	Thập_dại_danh_trà
Chạy_sấp_đạp_ngửa	Kinh_thiên_động_địa	Thập_diện_mai_phục
Cháy_thành_vạ_lầy	Ký_cả_hai_tay	Thập_tam_thoa
Chạy_vội_chạy_vàng	Kỳ_phùng_dịch_thủ	Thập_toàn_thập_mỹ
Chẻ_sợi_tóc_làm_tư	Lá_cờ_đầu	Thập_tử_nhất_sinh
Chém_đinh_chặt_sắt	Lá_lành_dùm_lá_rách	Thất_cơ_lỡ_vận
Chém_to_kho_mặn	Lá_mặt_lá_trái	Thất_đám_kinh_hồn
Chém_tre_không_dè_dầu_mặt	Lá_ngọc_cánh_vàng	Thất_đáy_lưng_ong
Chén_chú_chén_anh	Lạ_nước_lạ_cái	Thất_điên_bát_đảo
Chén_tạc_chén_thù	Lá_rụng_về_cội	Thất_lưng_buộc_bụng
Chen_vai_thích_cánh	La_trung_phản_hồn	Thất_thập_cổ_lai_hi
Chết_cha_còn_chú_xây_mẹ_bú_đi	Lá_xanh_trời_biết	Thất_thập_cổ_lai_hy
Chết_dần_chết_mòn	Lắc_đầu_lè_lười	Thấu_tình_đạt_lí
Chết_dở_sống_dở	Lãi_mẹ_dể_lãi_con	Thấu_tình_đạt_lý
Chết_đứng_còn_hơn_sống_quỷ	Làm_bia_dỡ_dạn	Thầy_bói_xem_voi
Chết_đuối_vớ_được_cọc	Làm_bộ_làm_tịch	Thay_da_đổi_thịt
Chết_đuối_vớ_phải_bọt	Làm_bỏ_xác	Thay_hình_đổi_dạng
Chết_không_kịp_ngáp	Làm_dầu_trăm_họ	Thay_lòng_đổi_dạ
Chết_mê_chết_mệt	Làm_đường_lạc_lối	Thấy_mồ_tổ
Chết_vinh_còn_hơn_sống_nhục	Làm_mình_làm_mấy	Thay_ngựa_giữa_dòng
Chỉ_dầu_dánh_dó	Lầm_mồm_lầm_miệng	Thấy_người_sang_bắt_quàng_làm_họ
Chỉ_mảnh_treo_chuông	Làm_mưa_làm_gió	Thay_thầy_đổi_chủ
Chỉ_mảnh_treo_chuông	Làm_ơn_nên_oán	Thệ_hải_minh_sơn
Chỉ_mỗi_tội	Lầm_sãi_không_ai_đóng_cửa_chùa	Thề_non_hẹn_biển
Chỉ_nam_ngọc_âm	Lam_sơn_chưưng_khí	Thề_sống_thề_chết
Chỉ_nam_ngọc_âm_giải_nghĩa	Lầm_thầy_thối_ma	Thế_thái_nhân_tình
Chỉ_tay_năm_ngón	Lầm_tiền_nhiều_tật	Thêm_mắm_thêm_muối
Chìa_khoá_trao_tay	Làm_tình_làm_tội	Thềm_vào
Chia_năm_xế_bảy	Làm_trao_cho_đáng_thân_trao	Theo_dóm_ăn_tàn
Chia_ngọt_sẻ_bùi	Lặn_ngồi_ngồi_nước	Theo_voi_ăn_bã_mía
Chia_uyên_rê_thuý	Lẩn_như_chạch	Thét_ra_lửa
Chiêm_khê_mùa_thối	Lang_bạt_kì_hồ	Thì_chớ
Chiêu_bình_mãi_mã	Lang_bạt_kỳ_hồ	Thì_chớ_kể
Chiều_như_chiều_vong	Lang_tai_ký	Thì_có
Chiều_tài	Lành_ít_dữ_nhiều	Thi_gan_dấu_trí
Chiều_tài_hoá_sắt	Lạnh_như_băng	Thi_gan_dọ_sức
Chiều_tài_tác_lộc	Lạnh_như_tiền	Thì_phải
Chim_kêu_vượn_hú	Lão_giả_an_chi	Thi_trung_hữu_hoạ
Chìm_sa_cá_lặn	Lão_nông_trí_diễn	Thiên_biến_vạn_hoá
Chìm_trời_cá_nước	Lao_tâm_khổ_tứ	Thiên_binh_thiên_tướng
Chín_bỏ_làm_mười	Lạt_mềm_buộc_chặt	Thiên_cao_địa_hạu
Chín_chữ_cù_lao	Lầu_son_gác_tía	Thiên_kinh_địa_nghĩa
Chín_người_mười_ý	Lầu_tôm_lầu_cá	Thiên_la_địa_vông
Chính_dại_quang_minh	Lạy_cả_nón	Thiên_lang_ấn
Chinh_phụ_ngâm	Lấy_độc_trị_độc	Thiên_long_bất_bộ

Table C.7: List of idioms.

Chịu thương chịu khó	Lạy như tể sao	Thiên lý tẩu đơn kỵ
Chó cắn áo rách	Lạy ông tôi ở bụi này	Thiện nam tín nữ
Chó cắn ma	Lấy thịt đè người	Thiện nữ u hồn
Chó cày gần nhà	Lấy thúing úp voi	Thiên quan tứ phúc
Chó chạy cùng rào	Lạy trời chuyển đất	Thiếu lâm bổng pháp
Chó chê mèo lắm lông	Lên bờ xuống ruộng	Thiếu lâm thất thập nhị huyền công
Chó có trách	Lên thác xuống ghềnh	Thóc đầu mà dài gà rừng
Chở củi về rừng	Lên voi xuống chó	Thọc gậy bánh xe
Chó cùng dứt đau	Lên xe xuống ngựa	Thối hư tật xấu
Chó cùng rút giầu	Liên chi hồ điệp	Thông đồng bên giọt
Chó ngáp phải ruồi	Liên hoàn chiến thuyền	Thông kim bác cổ
Cho qua chuyện	Liên hoàn mã	Thông tắc bất thông
Chó săn chim mồi	Liên hồi kì trận	Thua chi kém em
Chó thấy sóng cả mà ngã tay chèo	Liệu cái thần hồn	Thừa giấy vẽ voi
Chó treo mèo dầy	Liệu cơm gắp mắm	Thừa gió bẻ măng
Chọc gậy bánh xe	Liệu gió phất cờ	Thua me gỗ bài cào
Chối bay chối biến	Liệu thần hồn	Thừa sống thiếu chết
Chối cùn rế rách	Liều trai chí dị	Thuận buồm xuôi gió
Chơi với lửa	Liều yếu đào thơ	Thuận chèo mát mái
Chọn mặt gửi vàng	Liều yếu đào tơ	Thuận mua vừa bán
Chôn nhau cắt rốn	Lĩnh nam chích quái	Thuần phong mỹ tục
Chôn rau cắt rốn	Lo bỏ trống rỗng	Thuần phong mỹ tục
Chồng chung vợ chạ	Lơ trung hoả	Thực bất tri kỳ vị
Chồng loan vợ phượng	Lời ăn tiếng nói	Thực khuya dậy sớm
Chưa biết mèo nào cắn mỉu nào	Lợi bất cập hại	Thực mục sở thị
Chưa cứng lòng mao đã rào vẫy dít	Lời bè kéo cánh	Thực phổ bách thiên
Chùa đất phật vàng	Lời đường mật	Thực tình mà nói
Chưa đỡ ông nghề đã đe hàng tổng	Lời mặt chuột	Thực tức bình cường
Chùa nát bụi vàng	Lợi ngược dòng	Thực vi chi thuật
Chưa ráo máu đầu	Lời nói gió bay	Thuốc đắng dã tật
Chưa thấy quan tài chưa đổ lệ	Lời ong tiếng ve	Thuộc như lòng bàn tay
Chuẩn đường bình sở	Lời qua tiếng lại	Thượng cẳng chân hạ cẳng tay
Chức trọng quyền cao	Lời ra tiếng vào	Thương hoa tiếc ngọc
Chửi cha không bằng pha tiếng	Lợn lành chữa thành lợn què	Thương tâm đồng thoại
Chửi chó mắng mèo	Lớn nhanh như thổi	Thượng vàng hạ cám
Chúi đầu chúi mũi	Lòng chim dạ cá	Thương vay khóc mướn
Chung chăn gối	Lộng giả thành chân	Thủy chung như nhất
Chung lưng đầu cật	Lòng lang dạ sói	Thủy hử vô gian đạo
Chứng nào tật ấy	Lòng lang dạ thú	Tích cốc phòng cơ
Chuồn chuồn đập nước	Long môn phi giáp	Tích tiểu thành đại
Chuồn tai gai mắt	Long nam sư	Tiền hậu bất nhất
Chuột chạy cùng sào	Lòng son dạ sắt	Tiền hô hậu ủng
Chuột sa chĩnh gạo	Long thành cầm giả ca	Tiền lữ kì duyên
Có bát ăn bát để	Long tranh hổ đấu	Tiền lữ kỳ duyên
Cò bay thẳng cánh	Long trời chuyển đất	Tiền mất tật mang
Có bè gì	Long trời lở đất	Tiền nào cửa ấy
Cờ bí dí tốt	Lòng vả cũng giống lòng sung	Tiền nghìn bạc vạn
Cổ cày vai bừa	Lòng vả cũng như lòng sung	Tiền oan nghiệp chướng
Có còn hơn không	Lọt sàng xuống nia	Tiên phong đạo cốt
Có công mài sắt có ngày nên kim	Lú ruột lú gan	Tiên thiên bất tức
Có da có thịt	Lửa cháy đổ thêm dầu	Tiến thoái lưỡng nan
Cổ dầm ăn xôi	Lửa thầy phản bạn	Tiên trách kỉ
Có đi có lại	Lừa trâu vô rú	Tiền trăm hậu tẩu
Có ít nói sít ra nhiều	Lực bất tòng tâm	Tiền trao cháo múc
Cổ kim hoà ca tập	Lực mạch thần kiếm	Tiền trình dĩ cầm
Có máu mặt	Lười chầy thây	Tiền tươi thóc thật
Có mới nói cũ	Lười không xương	Tiếng bắc tiếng chì
Có một không hai	Luôn chân luôn tay	Tiếng cả nhà không
Có nơi có chốn	Lươn ngắn chê chạch dài	Tiếng chẳng tiếng chuộc
Có qua có lại	Lươn ngắn lại chê chạch dài	Tiểu ngao giang hồ
Cổ sống cổ chết	Lường thầy phản bạn	Tím gan tím ruột
Có tật giặt mình	Lương y như từ mẫu	Tin sái cổ

Table C.8: List of idioms.

Có thực mới vực được đạo	Lưu hương ký	Tinh hao khí tổn
Có tình có lí	Ma chê quỷ hờn	Tình ngay lí gian
Có tình có lý	Ma thiêng nước độc	Tình sâu nghĩa nặng
Có trăng quên đèn	Ma xui quỷ khiến	Tình tâm cóc
Có trên có dưới	Ma xui quỷ khiến	Tình xưa nghĩa cũ
Cốc bỏ đĩa	Mặc khách tao nhân	To gan lớn mặt
Cốc mò cò ăn	Mặc lòng	Tò son điểm phấn
Cốc mò cò xơi	Mai danh ản tích	Tò son trát phấn
Coi người bằng nửa con mắt	Mâm cao cỗ đầy	Toạ hưởng kì thành
Coi trời bằng vung	Mãn chiều xế bóng	Toạ sơn quan hổ đấu
Cơm ăn áo mặc	Mãn nguyệt khai hoa	Toàn tâm toàn ý
Cơm áo gạo tiền	Màn trời chiếu đất	Toàn thiện toàn mỹ
Cơm bưng nước rót	Mang con bỏ chợ	Toát mồ hôi hột
Cơm ghe bè bạn	Mang nặng đẻ đau	Tốc chiến tốc quyết
Cơm hàng cháo chợ	Mang tai mang tiếng	Tốc chiến tốc thắng
Cơm không lành canh không ngọt	Manh áo miếng cơm	Tối cường hỷ sự
Cơm niêu nước lọ	Mất ăn mất ngủ	Tội gì
Cơm thừa canh cặn	Mất bò rồi mới làm chuông	Tối lửa tắt đèn
Con bé con bông	Mặt búng da chì	Tối mắt tối mũi
Con bông con mang	Mặt búng ra sữa	Tối mày tối mặt
Con đại cái mang	Mất cả chì lẫn chài	Tối như mực
Con dao hai lưỡi	Mặt cắt không còn hột máu	Tối nơi tối chốn
Con gái con đũa	Mặt cưa mướp đắng	Tối tăm mặt mũi
Con giun xéo lắm cũng quăn	Mát da mát thịt	Tối tăm mày mặt
Còn khuya	Mát dạn mày dày	Tồn sư trọng đạo
Còn mệt	Mát dày mày dạn	Tòng bá mộc
Còn mồ ma	Mặt đỏ tía tai	Tổng cựu nghênh tân
Con nhà lính tính nhà quan	Mặt đối mặt	Trà dư tửu hậu
Còn nước còn tát	Mặt hoa da phấn	Trách nào chẳng
Con ông cháu cha	Mặt ít ruồi nhiều	Trách nào mà chẳng
Còn phải nói	Mất la mày lét	Trách nào mà chẳng
Con rồng cháu tiên	Mát lòng mát dạ	Trái gió gió giò
Con sâu làm rầu nồi canh	Mát mặt mắt nhạt	Trái gió trở trời
Con trai con đũa	Mát mày mát mặt	Trái nắng trở trời
Còn xơi	Mặt ngọt chết ruồi	Trái tai gai mắt
Công ăn việc làm	Mặt người dạ thú	Trai tài gái sắc
Công dã tràng	Mất nhắm mất mở	Trai thanh gái lịch
Công làm thủ phá	Mất phượng mày ngài	Trái tính trái nết
Công phu vịnh xuân	Mặt sửa gan lim	Trăm cay ngàn đắng
Công rắn cắn gà nhà	Mất thấy tai nghe	Trăm công ngàn việc
Công thành danh toại	Mất trời ông địa	Trăm công nghìn việc
Công tôn vương tử	Mất tròn mất det	Trăm đầu đổ đầu tằm
Công tư lưỡng lợi	Mất trước mất sau	Trăm hay không bằng tay quen
Cốt nhục tương tàn	Máu chảy ruột mềm	Trăm phương ngàn kế
Cù bơ cù bất	Màu cờ sắc áo	Trăm thứ bà giằn
Cửa ăn cửa để	Máu ghen Hoạn Thư	Trầm tư mặc tưởng
Cửa chìm cửa nổi	Màu mỡ riêu cua	Trần bảo kì thú
Cửa đáng tội	Mau mồm mau miệng	Trần như nhộng
Cửa đau con xót	Mẫu nghi thiên hạ	Trần triều quốc bảo
Cửa đi thay người	Màu sắc ản hiện	Trang bị tận răng
Cửa đóng then cài	Mấy ai	Tràng giang đại hải
Cửa đứt đục suốt	Mấy chóc	Trắng như cước
Cua gặp ếch	Mấy đời	Trắng như trứng gà bóc
Cửa ít lòng nhiều	Mấy khi	Tranh bá đồ vương
Cửa ngon vật lạ	Mấy nă	Tranh tối tranh sáng
Cửa người phúc ta	Mấy nổi	Tránh vỏ dưa gặp vỏ dừa
Cửa sừng làm ghé	Mày tao chi tổ	Tránh voi chẳng xấu mặt nào
Cửa thiên trả địa	Mẹ gà con vịt	Trâu buộc ghét trâu ăn
Cực chẳng đã	Mẹ goá con cô	Trâu chậm uống nước đục
Cúc cung tận tụy	Mẹ hất con khen hay	Trầy da tróc vẩy
Củi mực khó cháy	Mê hồn trận	Trầy vi tróc vẩy
Cùm chân xích cánh	Mê như điều đồ	Tre già măng mọc
Cùn đời	Mẹ tròn con vuông	Trẻ người non dạ
Cùn đời mặt kiếp	Mềm nắn rắn buông	Trên bến dưới thuyền

Table C.9: List of idioms.



Cùng bắt đắc dĩ	Mèo đàng chó điểm	Trên đe dưới búa
Cứng đầu cứng cổ	Mèo già hoá cáo	Trên kính dưới nhường
Cùng hội cùng thuyền	Mèo khen mèo dài đuôi	Trên trời dưới biển
Cứng như đá	Mèo mả gà đồng	Trên trời dưới đất
Cung oán ngâm khúc	Mèo mù vớ cá rán	Treo đầu dê bán thịt chó
Cũng quá tội	Mèo nào cắn mẩu nào	Trèo đèo lội suối
Cung tâm kế	Mèo nhỏ bắt chuột con	Tri thiên mệnh
Cung toả liêm châu	Một lữ cò bợ	Trở trời
Cung toả tâm ngọc	Miệng ăn núi lở	Trời có mắt
Cùng trời cuối đất	Miếng cơm manh áo	Trời đánh không chết
Cười đầu cười cổ	Miệng còn hoi sữa	Trời đánh thánh vật
Cười hỏ	Miệng hùm gan sữa	Trối gà không chặt
Cười ngựa xem hoa	Miệng nam mô bụng bồ dao găm	Trời sinh voi sinh cỏ
Cười như nắc nẻ	Mình đồng da sắt	Trời sinh voi trời sinh cỏ
Cười ra nước mắt	Mình hạc xương mai	Trôi sông lạc chợ
Cười vỡ bụng	Mình sơn thệ hải	Trời tru đất diệt
Cuồng phong bạo vũ	Mỡ bầy mỡ ba	Trối voi bỏ rọ
Cửu âm chân kinh	Mỡ để miệng mèo	Trời xui đất khiến
Cứu nhân độ thế	Mỏ hôi nước mắt	Trời yên biển lặng
Cửu thiên tiên mộng	Mỏ kim đáy biển	Trộm vía
Da bọc xương	Mỏ mặt mỡ mày	Trốn chúa lộn chồng
Đá chéo sân	Mỏ mày mỡ mặt	Trống đánh xuôi kèn thổi ngược
Đa cổ hoài lang	Mỏ yên mả đẹp	Trống dong cờ mớ
Đa đa ích thiện	Mọc mũi sủi tăm	Trông gà hoá cuốc
Đá gà đá vịt	Mỗi tội	Trông gà hóa cuốc
Da mỗi tóc sương	Mồm loa mép giải	Trông giở bỏ thóc
Da mưi túc trí	Mồm miệng đỡ chân tay	Trống huơ trống hoác
Đa nghi như Tào Tháo	Mồm năm miệng mười	Trông mặt mà bắt hình dong
Đa ngôn đa quá	Môn đăng hộ đối	Trọng nam khinh nữ
Đa ngựa bọc thây	Mong được ước thấy	Trọng nghĩa khinh tài
Đá quả bóng	Một chín một mười	Tru di tam tộc
Đá thúng đựng nia	Một chón đôi què	Trúc chẻ ngói tan
Đã trót phải trét	Một chữ bẻ đôi	Trúc lâm thất hiền
Đại cầm nã thủ	Một cổ hai tròng	Trưng cầu dân ý
Dãi cát lấy vàng	Một con ngựa đau cả tàu bỏ cỏ	Trưng chọi với đá
Dãi cát tìm vàng	Một công đôi ba việc	Trưng cuốc
Dãi dòng văn tự	Một công đôi việc	Trưng để đầu đẳng
Dại gì	Một đồng một cốt	Trưng gà trứng vịt
Dại gì mà	Một lòng một dạ	Trưng khôn hơn rận
Dãi gió dầm mưa	Một mắt một còn	Trưng khôn hơn vịt
Dại khê thủy	Một mắt mười ngờ	Trước sau như một
Dại mình nhất thống chí	Một mình một bóng	Trường giả học làm sang
Dại nam nhất thống chí	Một mình một chợ	Trường hận ca
Dại nam nhất thống toàn đồ	Một nắng hai sương	Trường sinh bất tử
Dại nam quốc sử diễn ca	Một sớm một chiều	Truy cùng đuổi tận
Dãi nắng dầm mưa	Một tác đến trời	Tự biên tự diễn
Dại nguyên nhất thống chí	Một thân một mình	Tử biệt sinh li
Dại như đĩa	Một trời một vực	Tự cao tự đại
Dại sự ngoạ long	Một vừa hai phải	Tự cấp tự túc
Dại thanh nhất thống chí	Mũ ni che tai	Từ cổ chí kim
Dại thánh vô song	Múa ba tác lười	Tự cổ chí kim
Đái tội lập công	Mua đứt bán đoạn	Tứ cố vô thân
Đâm ba chày củ	Mua gian bán lận	Tự cung tự cấp
Đâm bị thóc chọc bị gạo	Mưa như trút nước	Tứ đại đồng đường
Đâm chân tại chỗ	Múa riu qua mắt thợ	Tự lực cánh sinh
Đâm chồi nảy lộc	Mưa thuận gió hoà	Tu nhân tích đức
Đâm lao phải theo lao	Mưa to gió lớn	Tứ nhiếp pháp
Dâm mưa dãi nắng	Mục hạ vô nhân	Tự sản tự tiêu
Dám nghĩ dám làm	Mùng như bắt được vàng	Tử sát giao long
Đắm nguyệt say hoa	Muối bỏ bể	Tu tâm dưỡng tánh
Đàn gảy tai trâu	Muối bỏ biển	Tứ thân phụ mẫu
Dân ngu khu đen	Mười bữa nửa tháng	Tứ thời bát tiết
Dẫn thủy nhập điền	Mượn gió bẻ măng	Tự trang tự chế
Dân vi quý	Muôn hình vạn trạng	Từ trong trứng

Table C.10: List of idioms.

Dáng_dồng_tiền_bát_gao	Muôn_màu_muôn_vẻ	Tự_tự_tự_lợi
Đặng_trăng_quên_trang	Muốn_sang_thì_bắc_cầu_Kiều	Tự_tung_tự_tác
Đằng_vân_giả_vũ	Mưư_ma_chước_quỉ	Tuần_tự_nhi_tiến
Đánh_bài_ngửa	Mưư_ma_chước_quỷ	Tức_cảnh_hồng
Danh_bất_hư_truyền	Mỹ_nhân_tâm_kế	Tức_cảnh_sinh_tình
Đánh_bùn_sang_ao	Năm_cha_ba_mẹ	Tức_lộn_ruột
Danh_chính_ngôn_thuận	Năm_châu_bốn_biển	Tức_nước_vỡ_bờ
Đanh_đá_cá_cày	Năm_cùng_tháng_tận	Tùm_đen_tùm_đỏ
Đánh_đồng_đẹp_bắc	Nắm_đăng_chuôi	Túm_năm_túm_ba
Đánh_đu_với_tinh	Nằm_gai_nằm_mặt	Tương_kế_tự_kế
Danh_lam_thắng_cảnh	Nam_giao_học_tổ	Tuỳ_cơ_ứng_biến
Đánh_rắm_siêu_khắc	Năm_lần_bảy_lượt	Tuyệt_tình_ca
Đánh_trống_bỏ_dùi	Nam_phụ_lão_ấu	Ừ_ừ_cạc_cạc
Đánh_trống_ghi_tên	Nam_thanh_nữ_tú	Uốn_ba_tắc_lưỡi
Đánh_trống_lẳng	Năm_thê_bảy_thiếp	Uổng_công_phí_sức
Đánh_trống_lấp	Năm_thì_mười_họa	Uổng_nước_nhớ_nguồn
Đánh_trống_qua_cửa_nhà_sấm	Năm_thì_mười_họa	Uớt_như_chuột_lột
Dao_canh_hoả_chủng	Năm_xung_tháng_hạn	Úp_úp_mở_mở
Dao_hai_lưỡi	Năn_nỉ_ỉ_ôi	Vá_chằng_vá_đụp
Dao_to_búa_lớn	Năng_khăn_sửa_túi	Vạch_áo_cho_người_xem_lưng
Đào_tơ_liều_yếu	Năng_nhật_chặt_bị	Vạch_lá_tìm_sâu
Đáo_xứ_tuỳ_thân	Nắng_như_đỏ_lửa	Vải_thưa_che_mắt_thánh
Đắp_tai_cài_tróc	Nát_một_đời_hoa	Vai_u_thịt_bắp
Đất_bằng_nổi_sóng	Nát_như_tương	Ván_đá_đóng_thuyền
Đất_khách_quê_người	Nấu_sử_sôi_kinh	Vạn_diệp_tập
Đất_lành_chim_đậu	Nay_đây_mai_đó	Văn_dốt_vũ_dát
Đất_lề_quê_thối	Nể_tình_vị_nghĩa	Văn_kỳ_thình
Đạt_lí_thấu_tinh	Nem_công_chả_phượng	Vạn_ly_trường_thành
Đất_như_tôm_tươi	Ném_đá_giấu_tay	Văn_mình_vợ_người
Đất_rộng_trời_cao	Ném_tiền_qua_cửa_sổ	Văn_ôn_võ_luyện
Đầu_ấp_tay_gối	Nền_cơm_cháo_gì	Vạn_phụng_chi_vương
Đầu_bạc_răng_long	Ngã_theo_chiều_gió	Vạn_sự_khởi_đầu_nan
Đầu_bò	Ngậm_bồ_hòn	Vạn_sự_như_ý
Đầu_bò_đầu_bướu	Ngậm_bồ_hòn_làm_ngọt	Vắng_chủ_nhà_gà_vọc_niêu_tôm
Đầu_bù_tóc_rối	Ngậm_cười_chín_suối	Vang_mình_sốt_mấy
Đầu_chày_đít_thốt	Ngậm_cười_nơi_chín_suối	Váng_mình_sốt_mấy
Đấu_chiến_thắng_phật	Ngậm_đắng_nuốt_cay	Vàng_thau_lăn_lộn
Đầu_cua_tai_nheo	Ngậm_hột_thị	Vào_lỗ_hà_ra_lỗ_hổng
Đầu_dây_mối_dợ	Ngậm_máu_phun_người	Vào_luôn_ra_cúi
Đầu_dây_mối_nhợ	Ngân_cân_treo_sợi_tóc	Vào_sinh_ra_tử
Đầu_đuôi_xuôi_ngược	Ngăn_sông_cắm_chợ	Vào_tù_ra_tội
Đầu_đường_xó_chợ	Ngày_ba_tháng_tám	Vật_bất_li_thân
Đầu_gối_tay_ấp	Ngày_lành_tháng_tốt	Vật_bất_ly_thân
Đau_như_cắt	Ngày_một_ngày_hai	Vắt_chân_chữ_ngũ
Đầu_sôi_lửa_bỏng	Ngày_qua_tháng_lại	Vắt_chân_lên_cổ
Đầu_sóng_ngọn_gió	Ngày_rộng_tháng_dài	Vắt_chanh_bỏ_vỏ
Đầu_sóng_ngọn_nguồn	Nghê_đời_nó_thể	Vắt_cổ_chày_ra_nước
Đầu_tắt_mặt_tối	Nghe_hơi_nổi_chỗ	Vật_đổi_sao_dời
Đầu_thừa_đuôi_thẹo	Nghệ_thuật_vị_nghệ_thuật	Vắt_mũi_chưa_sạch
Đầu_trâu_mặt_ngựa	Nghệ_thuật_vị_nhân_sinh	Vẽ_đường_cho_hươu_chạy
Đầu_trộm_đuôi_cướp	Nghèo_rớt_mòng_tơi	Vẽ_rắn_thêm_chân
Đầu_voi_đuôi_chuột	Nghèo_rớt_mùng_tơi	Vì_thần_phải_nể_cây_đa
Đầu_voi_đuôi_chuột_minh_heo	Nghĩa_tử_là_nghĩa_tận	Việc_gì
Đầu_xuôi_đuôi_lột	Nghiêng_nước_nghiêng_thành	Việc_nhà_thì_nhắc_việc_cô_bác_thì_siêng
Đáy_bể_mò_kim	Nghìn_cân_treo_sợi_tóc	Viễn_giao_cận_công
Đáy_biển_mò_kim	Ngộ_biến_tùng_quyền	Việt_điện_u_linh
Dây_cà_ra_dây_muống	Ngó_trước_nhìn_sau	Vinh_quy_bái_tổ
Dày_gió_dạn_sương	Ngoại_hố_tàng_long	Vinh_thân_phỉ_gia
Dây_máu_ăn_phần	Ngoại_bất_nhập	Vỡ_chợ
Dây_mơ_rễ_má	Ngoài_lề	Vô_công_rồi_nghề
Đề_chùng_bất_bóng	Ngoài_mặt	Vô_đang_kiểm
Đề_đầu_cưỡi_cổ	Ngoài_miệng	Vô_danh_tiểu_tốt
Đẽ_làm_khó_bỏ	Ngoảnh_đi_ngoảnh_lại	Vò_đầu_bóp_trán
Đem_chuông_đi_đánh_xứ_người	Ngọc_bất_trắc_bất_thành_khí	Vơ_đũa_cả_nắm

Table C.11: List of idioms.

Dem_con_bỏ_chợ	Ngồi_chưa_nóng_đít	Vô_kế_khả_thi
Đêm_dài_lắm_mộng	Ngồi_lê_đôi_mách	Vỏ_quýt_dày_có_móng_tay_nhọn
Đếm_trên_đầu_ngón_tay	Ngồi_lê_mách_lẻo	Vô_thưởng_vô_phạt
Đến_đầu_đến_đuă	Ngồi_mát_ăn_bát_vàng	Vô_thuỷ_vô_chung
Đến_hẹn_lại_lên	Ngọt_như_mía_lùi	Vô_tiền_khoảng_hậu
Đền_nhà_ai_nấy_tổ	Ngủ_gà_ngủ_gật	Vô_tri_bất_mộ
Đền_nhà_ai_nhà_nấy_rạng	Ngủ_hổ_tướng	Vô_tri_vô_giác
Đen_như_cột_nhà_cháy	Ngũ_long_công_chúa	Vu_oan_giá_hoạ
Đen_như_mực	Ngu_như_bò	Vũ_trang_đến_tận_răng
Đến_nơi_đến_chốn	Ngựa_non_hầu_đá	Vừa_ăn_cướp_vừa_la_làng
Đèo_heo_hút_gió	Ngựa_phản_chủ	Vừa_đắm_vừa_xoa
Đi_cổng_sau	Ngựa_quen_đường_cũ	Vừa_đánh_trống_vừa_ăn_cướp
Đi_đăng_đầu	Người_dùng_mắt_lã	Vừa_đôi_phải_lửa
Đi_đêm_có_ngày_gặp_ma	Người_trần_mất_thịt	Vui_như_trầy_hội
Đi_đêm_lắm_có_ngày_gặp_ma	Ngư_tầm_ngư_mã_tầm_mã	Vụng_chèo_khéo_chống
Đi_đêm_về_hôm	Nhà_cao_cửa_rộng	Vụng_tay_quá_trán
Đi_đến_nơi_,_về_đến_chốn	Nhả_ngọc_phun_châu	Vụng_tiền_qua_cửa_sổ
Đi_đến_nơi_về_đến_chốn	Nhà_tranh_vách_đất	Vườn_không_nhà_trống
Đi_đời_nhà_ma	Nhắm_mắt_đưa_chân	Vườn_rộng_rào_thừa
Đi_đường_vòng	Nhắm_mắt_xuôi_tay	Vuông_thành_sắc_cạnh
Đi_guốc_trong_bụng	Nhân_bắt_học_bắt_tri_lý	Vuốt_mặt_không_kịp
Dĩ_hoà_vĩ_quí	Nhàn_cư_vĩ_bất_thiện	Vuốt_mặt_không_nể_mũi
Dĩ_hoà_vĩ_ quý	Nhân_tình_thế_thái	Vuốt_râu_hùm
Dĩ_hòa_vĩ_ quý	Nhân_vô_thập_toàn	Vứt_một_cái
Đi_mây_về_gió	Nhanh_như_cắt	Xa_chạy_cao_bay
Đi_tắt_đón_đầu	Nhanh_như_chớp	Xa_mặt_cách_lòng
Đĩa_đôi_deo_chân_hạc	Nhập_gia_tuỳ_tục	Xa_sông_cách_núi
Điếc_không_sợ_sấm	Nhất_bản_van_lợi	Xanh_vỏ_đỏ_lòng
Điếc_không_sợ_súng	Nhất_cử_lưỡng_tiền	Xé_ruột_xé_gan
Điều_ăn_tiếng_nói	Nhất_cử_nhất_động	Xem_mặt_mà_bắt_hình_dong
Điều_bình_khiển_tướng	Nhất_dương_chỉ	Xin_đi_đăng_đầu
Điều_hay_lẽ_phải	Nhất_hồ_bá_ứng	Xinh_như_mộng
Điều_hổ_li_sơn	Nhất_nghệ_tinh	Xôi_hồng_bồng_không
Điều_hổ_ly_sơn	Nhất_ngôn_đường	Xuân_thu_chiến_quốc
Điều_hơn_lẽ_thiệt	Nhất_như_cây	Xuất_đầu_lộ_diện
Điều_nặng_tiếng_nhẹ	Nhất_như_thỏ_đế	Xuất_giá_tòng_phu
Điều_này	Nhất_phạm_phong_thuận	Xuất_khẩu_thành_thi
Điều_này_thì	Nhật_phần_phai_hương	Xuất_khẩu_thành_thơ
Điều_nọ_tiếng_kia	Nhất_thân_vinh	Xuất_quĩ_nhập_thần
Điều_ong_tiếng_ve	Nhất_thành_bất_biến	Xuất_quỹ_nhập_thần
Điều_qua_tiếng_lại	Nhất_thiểm_yêu_hoàng	Xúc_cảnh_sinh_tình
Điều_ra_tiếng_vào	Nhất_tự_vĩ_sư_,_bán_tự_vĩ_sư	Xui_nguyên_giục_bị
Điều_vô_dương_oai	Nhất_tự_vĩ_sư_bán_tự_vĩ_sư	Xui_trẻ_ăn_cứt_gà
Dính_như_keo	Nhẹ_mềm_nhẹ_miệng	Xúm_đen_xúm_đỏ
Đo_bò_làm_chuồng	Nhẹ_như_bắc	Xưng_hùng_xưng_bá
Đỏ_con_mắt	Nhi_nữ_thường_tình	Xuôi_chèo_mát_mái
Dò_đá_qua_sông	Nhiệt_ái_đảo	Xương_đồng_da_sắt
Đỏ_da_thấm_thịt	Nhiều_sãi_không_ai_đóng_cửa_chùa	Ý_hợp_tâm_đầu
Dơ_dáng_dại_hình	Nhịn_miệng_đãi_khách	Y_như_răng
Đổ_dầu_vào_lửa	Nhìn_xa_trông_rộng	Ý_tại_ngôn_ngọai
Dở_dở_ương_ương	Nhỏ_cổ_nhỏ_cả_rể	Yếm_thấm_trơn_kim
Dở_dơi_dở_chuột	Nhọ_mặt	Yên_giấc_ngàn_thu
Dở_khóc_dở_cười	Nhọ_mặt_người	Yên_giấc_nghìn_thu
Đỏ_mặt_tía_tai	Nhớ_như_in	Yên_lạc_tự_chung

Table C.12: List of idioms.

# Bibliography

- [1] Anne-Marie Di Sciullo and Edwin Williams. 1987. *On the definition of word*, volume 14. Springer.
- [2] DCS LacViet Corp. 2011. *Vietnamese Dictionary*. LacViet Corp.
- [3] Duc-Trong Le. 1993. *Vietnamese glossary*. Ho Chi Minh City Publishers.
- [4] Fei Xia. 2000. *The segmentation guidelines for the Penn Chinese Treebank*.
- [5] Phe Hoang. 1998. *Vietnamese Dictionary*. Scientific & Technical Publishing.
- [6] Phuong-Thai Nguyen, Xuan-Luong Vu, and Thi-Minh- Huyen Nguyen. 2010. *Vietnamese word segmentation guidelines*.
- [7] Quang-Ban Diep. 2005. *Vietnamese grammar*. Vietnam Education Publisher.
- [8] Van-Hanh Hoang. 2008. *Vietnamese Idioms*. Social Sciences Publishers
- [9] Vietnam Social Sciences Commission. 1983. *Vietnamese grammar*. Social Sciences Publishers.
- [10] Xuan-Hao Cao. 1999. *Vietnamese glossary*. Social Sciences Publishers.