# Some Lexical Issues in Building Electronic Vietnamese Dictionary

Dinh Dien[1]
Pham Phu Hoi[2], Ngo Quoc Hung[2]

[1] Faculty of Information Technology, Vietnam National University of HCM City
20/C2 Hoang Hoa Tham, Ward 12, Tan Binh Dist., HCM City, Vietnam
ddien@saigonnet.vn
[2] Center of Information Technology Development, Vietnam National University of HCM City,
78 Truong Dinh, Dist. 3, HCM City, Vietnam
{hungnq, hoipp} @citd.edu.vn

**Abstract.** So far, the computer-based processing for Vietnamese has been limited due to different reasons and one of those is the lack of fundamental databases for automatic processing of natural languages by computers. One such necessary database is the electronic Vietnamese Dictionary used to process Vietnamese by computers. Computer Processing of Natural Languages consists of various tasks and all these tasks need to access the electronic dictionary database. So, the pre-requisite condition for every NLP (Natural Language Processing) is to build the electronic dictionary which computers are able to read from (in technical terms, this dictionary is called MRD: Machine Readable Dictionary). In this paper, we present some issues needed to be solved in building a large-scale MRD for Vietnamese, e.g.: the macrostructure and microstructure of more than 35,000 entries in the dictionary, the internationality of criteria for selecting word-entries in order that this dictionary can inherit, interface to other popular English NLP systems in the world.

## 1 Overview of Vietnamese Dictionaries

Vietnamese dictionaries here consist of 2 kinds: the traditional dictionaries (paper-dictionaries, hardcopy) used for human beings and MRD (electronic dictionaries, softcopy) used for machines. After investigating the macrostructure and microstructure of traditional Vietnamese dictionaries, we proceed to build the Vietnamese MRD for computers.

### 1.1 Traditional Vietnamese Dictionaries for Human Beings

So far, there have been a lot of studies in Vietnamese dictionaries but due to the limitation of this paper; we only mention the most related work – that is the Vietnamese dictionaries edited by Prof. Hoang Phe [8]. When we build a dictionary, there are two

basic issues: the macro-structure and the microstructure of the dictionary.For Vietnamese language, identifying the macro-structure faces with an extremely difficult problem, that is the word recognition criteria, distinguishing the boundary among words, Vietnamese syllables, compound words, phrases, … and so far it has not achieved the complete agreement of Vietnamese linguists [5]. Similarly, identifying the micro-structure (information contained in each entry) is not a simple problem. Even the basic information which exists in almost every dictionary, such as information about Part-Of-Speech is not easy to identify, and so far there are many cases which have not been unified by Vietnamese linguists. So, building dictionaries in which Vietnamese is the source language (e.g. Vietnamese-English, Vietnamese-French, Vietnamese only, etc.) is much more difficult than others (e.g. English-Vietnamese, French-Vietnamese, English-French-Vietnamese, etc. ).

However, because of the essential requirement of building the basic dictionary to solve the practical problems, we still have to base on some certain criteria that were accepted. This criterion also has to ensure the connectivity to international databases via English language. Therefore, in the building of our Vietnamese MRD dictionary, for macro-structure, we based on the criteria which choose words of Hoang Phe dictionary along with the combination of other criteria basing on Vietnamese-English dictionaries [9] and English-Vietnamese dictionary [10]. For microstructure, beside the information of the part-of-speech similar to that of Hoang Phe dictionary, we also have to add a lot of other information that was statistically extracted from Vietnamese corpora or other international knowledge bases in English form via the word alignment in the English-Vietnamese [3] and Vietnamese-English dictionaries.

### 1.2 Vietnamese Machine Readable Dictionary (MRD)

In order to prepare the database to serve the computer automatic Vietnamese language processing system, building a machine-readable dictionary is an obligation. The machine-readable dictionary has the different structure and storing information organization from dictionaries for human beings. For example, it is not necessary to include information about phonetics, etymology, explanation, etc. in an MRD. However, it contains obvious information, or the information which is not necessary to be included in the dictionary for human beings (because we can deduce this information with the knowledge about the real world or the living experiences).

An MRD should have a coherent, logical, precise and sufficient organization in structure and amount of information so that the computer can completely base on it to accomplish its tasks in the machine manner. Therefore, to build this MRD dictionary, we have to completely solve the problem of the criteria of choosing word-entry (macro-structure). Information contained in each word entry (micro-structure) must be exact, unified in spelling, code table and presentation.

We have solved above-mentioned problems by using our tools such as: the Vietnamese normalizer (to normalize Vietnamese syllables in spell), the Vietnamese Code Converter (to convert from other character codes into TCVN3-code), the Vietnamese Word Segmenter (to segment Vietnamese word under a list of pre-defined words), the Vietnamese spelling-checkers (to check spell in syllable and word-level), etc.

## 2 The Macrostructure of Vietnamese MRD

In order to build the macrostructure for the Vietnamese MRD automatically, we have to solve these following problems: choosing word-entry criteria, word-order criteria, and characteristics of connectivity to international dictionary databases. Besides, organizing Vietnamese MRD also need to be open so that we will be able to update, change easily and fast. At last, building, managing, and updating such a rather large database (tens thousand of words) with such a high accuracy can't be hand-made, but must be automatically learned from related databases and electronic dictionaries.

### 2.1 Examining the Macrostructure of Hoang Phe's Vietnamese Dictionary

This is the basic and most important problem in building a Vietnamese MRD as well as other Vietnamese dictionaries. The issue of choosing word-entries of the dictionary depends on the criteria of choosing word-entry, recognizing words, distinguishing words from other units (lower or higher than word). This issue is very difficult and we do not have the ambition to solve by ourselves. Therefore, we have researched the Vietnamese dictionary written by Hoang Phe to extract this criteria of choosing word-entries, then combine with our proposed criteria in order to build a Vietnamese MRD which is suitable to the proposed purpose before. From examining the macrostructure of Hoang Phe's dictionary, we will chose consistent criteria so that computer can base on them to build our Vietnamese MRD automatically.

### 2.2 The Macrostructure of Vietnamese MRD

Basically, we follow the criteria of word selection in Hoang Phe dictionary. In addition to those criteria, we also apply our new criteria as follows:

1. One of several most distinguished features in Vietnamese is *classifier* which is absent in European languages (in some cases, this classifier is equivalent to the determiners/articles "the" in English or "le, la" in French). Classifiers (or vice-noun) are often used to specify the class of nouns [11]. Each noun will go along with only certain classifiers arbitrarily. Ex: "sách" (book) often goes with classifiers "quyển" or "cuốn" and somewhere it goes alone. For examples: it is often said that "Đây là một quyển sách" (This is a book) not "Đây là một sách" whilst "Tôi thích đọc sách" (I like reading book). So, in the macrostructure of Vietnamese MRD, it is not advisable to include all these possible combinations "*quyển sách/cuốn sách*". This is the reason why classifiers will not be integrated in entries of our Vietnamese MRD. That means, in the macrostructure of our Vietnamese MRD, it has entries "thư", "sách", "bò",… but not *bức thư/lá thư/cánh thư* (letter), *quyển sách/cuốn sách* (book), *con bò* (cow/ox),… Regarding these entries, the information of possible classifiers will be added into its microstructure. List of these classifiers is in Appendix 1.

2. Unlike classifiers in Vietnamese, words denoting categories or subcategories will be integrated in the entry of dictionary. Ex: "máy" (machine)→ *máy tính* (computer), *máy in* (printer), *máy quét* (scanner), *máy vẽ* (plotter), *máy phát* (generator), *máy đọc mã vạch* (bar code reader), ; "bộ"(device)→ *bộ đếm* (counter), *bộ xử lý* (processor),etc. Regarding the words denoting categories which have high generality and popularity and in practical maybe absent in use, we will note this feature in their microstructure. Ex: "bệnh" (disease) in *bệnh lao* (tuberculosis), *bệnh ho gà* (whooping cough), *bệnh uốn ván* (tetanus),… will have this feature. In order to determine completely and exactly words denoting species and categories, we have to base on the general classification tree of lexical semantic network – WordNet [4]. List of words denoting categories are in Appendix 2.
3. Similar to the inflectional typology (e.g. English, Russian, French), some Vietnamese word are coined by affix-method. For examples: *-hoá* (-ize), *-viên* (-er/-or/-ist/-ian), *-học* (-ology), *bất-* (in-/non-), *liên-* (inter-), *siêu-* (meta-/super-/hyper-),... in *điện toán hoá* (computerize), *lập trình viên* (programmer), *tâm lý học* (psychology), *siêu sao* (superstar), etc. These derivations are formed by contrasting English derivational affixes and Vietnamese morphems (which has Chinese-Vietnamese origin). List of derivation affixes are in Appendix 3 [6].
4. Beside well-accepted Vietnamese words, there are some entries which are not unanimous in word level, e.g.: *đường thẳng* (line), *nhà tranh* (cottage), *nhà gạch* (brick house), *dưa hấu* (watermelon), *xe đạp* (bicyle), etc. In this case, we made use additional lexicalization information of the corresponding English words to decide whether this entry is word or not. Ex: *đường thẳng*, *nhà tranh, dưa hấu,...* are words (because their corresponding English meaning are lexicalized), but *nhà gạch* (brick house) is not a word.
5. *Tone marks* is another distinguished features for Vietnamese syllables (compositions of Vietnamese words). Due to this feature, we are not able to use the standard ASCII code for storing Vietnamese letters. In our Vietnamese MRD, we have automatically sorted them under following order: "a ă â b c d đ e ê f g h i j k l m n o ô ơ p q r s t u ư v w x y z". The order of Vietnamese tone marks is: "none – brève – question – tilde – acute – dot below". The order of entries in the macro-structure of Vietnamese MRD is: *Vietnamese letter, tone marks, and the next letters* (while the order in Hoang Phe dictionary is: *Vietnamese letter, the next letters, and tone marks*).
6. In order to calculate the usage – frequency of words, we have built automatically the Vietnamese frequency dictionary by statically counting on a 10-million Vietnamese corpus belonging to various styles. We based on these frequencies to select most popular words, avoiding special terms or rare words (Dinh Dien,2002).
7. To solve the problem of missing popular words: in English, these words will be retrieved easily by the spelling checker. In Vietnamese, however, this is a difficult problem due to ambiguous word boundaries and we have solved most part of this problem via the model of Vietnamese word segmentation and English-Vietnamese word alignments [1], [3].

## 3 The Microstructure of Vietnamese MRD

The microstructure of Vietnamese dictionary includes structured information in every entry in order to control the automatic processing for natural languages. Regarding the standard of Vietnamese spell, we made use the TVCN3 standard. Nevertheless, we also meet other spell variations by coding tone marks of Vietnamese words in storing and using fuzzy comparator in searching.

### 3.1 Morphological information

- Word form, Ex: "sách" (book), "thắng lợi" (win),…
- Word variations: *lemma, variations, reduplicative*, …
- Morphological features: possible combinations or particles: *classifiers, post-position*,…Ex: "sách" (book) usually go accompany with classifier "quyển", "cuốn".
- Position code (word order): its frequent position in phrases/ sentences,...

### 3.2 Grammatical Information

- Parts-of-speech of word, e.g.: *noun, verb, adjective,…*
- Subcategory: e.g. subcategories of nouns: *countable nouns, uncountable nouns*; subcategories of verbs: *transitive verbs, intransitive verbs,…*
- Syntactic features: tense: *past, present, future*; voice: *passive, active;* gender: *male, female, neutral*; number: *singular, plural*,…
- Structure/pattern features: which structure/pattern is this word used ?
- Collocation/phrase/idiom: which collocation/phrase/idiom does this word usually go accompanied with ? Ex: verb "nhắm" (*close*) usually goes accompanied with "mắt" (*eye*).

### 3.3 Semantic Information

- The meaning of word in English. Ex: "book", "win",…
- The semantic code of word: e.g. :*HUMAN, ANIMATE, MOVE*,… These sense tags are built from the lexical semantic network WordNetV [4], [7].

### 3.4 Pragmatic Information

- The domain/field of usage: this word is usually used in domains: *Computer, Mathematics, Medicine,...*
- The frequency code: has this word been used frequently or not ? The occurring frequency of word is measured by following formula $f = -\log_{10}\frac{m}{N}$ where m is the number of occurrences and N is the length of corpus used for measuring (pls refer sect. 4.4). Ex: f=3 means that this word has occurred at the frequency 1/1000 [2].
- The modality code: which style is this word used: *formal, informal, oral*,….

## 4 Experimental Results

The statistical results of our new Vietnamese MRD are as follows:

### 4.1 Calculating on the length of words

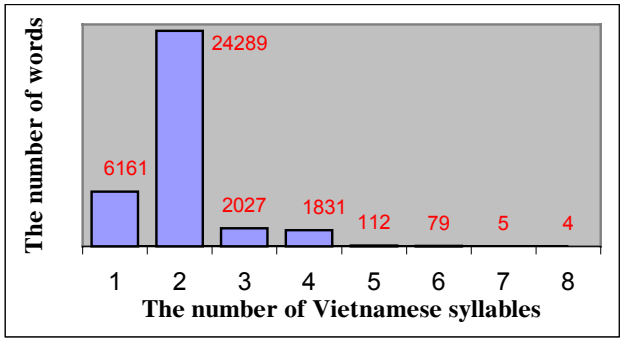


**Fig. 1.** Calculating on the length of words (Vietnamese syllables/word)

### 4.2 Calculating on the ambiguous POS

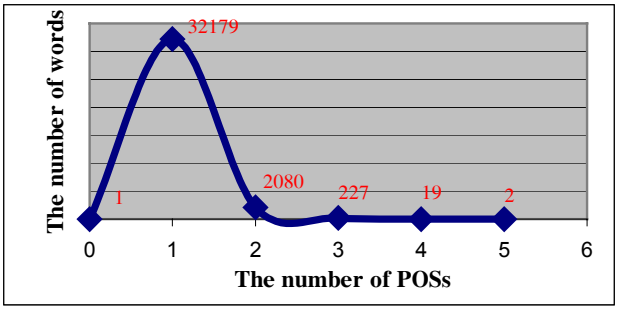


**Fig. 2.** Calculating on words and the number of their POSs

### 4.3 Calculating on the number of words for each POS

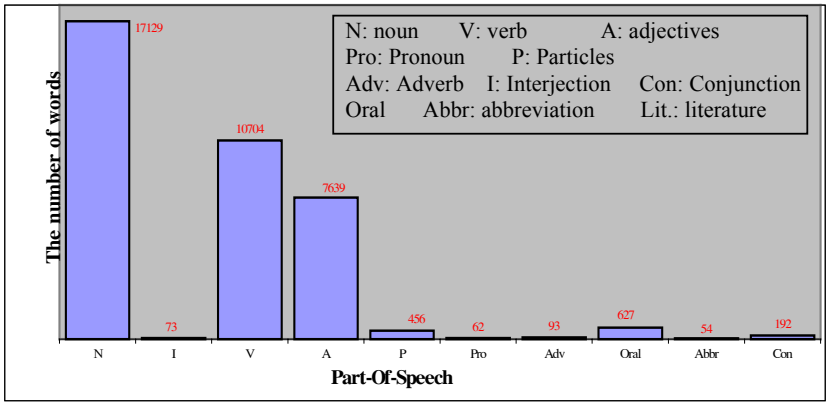


**Fig. 3.** Calculating on the number of words for each POS

### 4.4 The Usage-frequency of Vietnamese MRD

**Table 1.** Most frequent words in Vietnamese MRD (calculating on 10-M Vietnamese corpus of CINET – a webserver of Ministry of Culture of Vietnam)

| Word | English meaning | Occurences | Frequency |
|---|---|---|---|
| và | and | 215765 | 1.6985 |
| các | (classifier) | 170997 | 1.7995 |
| của | of | 152217 | 1.8501 |
| có | there/have | 118805 | 1.9577 |
| trong | in | 109308 | 1.9939 |
| được | (particle) | 89519 | 2.0806 |
| là | tobe | 88173 | 2.0872 |
| đã | (function word → past tense) | 86915 | 2.0934 |
| cho | for | 81175 | 2.1231 |
| ... | ... | … | … |
| tổ chức | organize/organizations | 35127 | 2.4869 |
| ... | ... | … | … |
| xây dựng | build/construct | 29528 | 2.5623 |
| ... | ... | ... | ... |
| văn hóa | culture | 29029 | 2.5697 |
| ... | ... | ... | ... |

## 5 Conclusion – Application – Limit – Future Development

We have just presented some issues concerning the automatic construction of macro- and micro-structure for the large-scale computerized Vietnamese dictionary – a kind of MRD to be used for Vietnamese-related NLP systems.

This Vietnamese MRD will serve as the basic database to such automatic Vietnamese processing systems as the word segmentation, spelling checker, POS-tagger, parser, semantic analyzer, etc. With the accompanied English information, this dictionary is easily inter-linkable with international dictionary database.

However, due to the dictionary scale as well as time limit up till now, this microstructure of this Vietnamese MRD can only contain such principal information (as bold-marked bullets) as: word forms, morphological codes; POS, subcategories; English meanings, sense tags; domains and frequency. In the future, the missing information will be statistically analyzed in an automatic way from more exhaustive and comprehensive Vietnamese database.

## References


1. Dien D., Kiem H., Toan N.V.: Vietnamese Word Segmentation. Proceedings of NLPRS'01 , Tokyo, Japan, 11/2001, (2001) 749-756.
2. Dinh Dien: The organization of electronic dictionary for English-Vietnamese Machine Translation. Journal of Science and Technology Development, National University of HCMC, Vol 5, No 1&2, (2002) 5-11.
3. Dinh Dien, et al. : Word alignment in English – Vietnamese bilingual corpus. Proceedings of EALPIIT'02 (The 2nd East Asian Language Processing and Internet Information Technology), HaNoi, Vietnam, 1/2002, (2002) 3-11.
4. Christian Fellbaum: WORDNET: an electronic lexical database. MIT Press, London (1999)
5. Hoang Van Hanh – Ha Quang Nang – Nguyen Van Khang: Từ tiếng Việt: hình thái – cấu trúc – từ láy – từ ghép – chuyển loại (Vietnamese words: morphology – structures – reduplicatives – compounds – type conversion). Social Sciences Publisher, Ha Noi (1998)
6. Do Dinh Lan: Lexicology. Vol. 1 & 2. Training College of HCM City (1993)
7. Van Chi Nam et al.: Building A WordNet for Vietnamese Nouns. Proceedings of Conference for Vietnamese Development, HCMC, Vietnam,12/2002, (2002), 41-45.
8. Hoang Phe: Từ điển tiếng Việt (Vietnamese Dictionary). Center of Lexicography. Da Nang Publisher (1998)
9. Bui Phung: Từ điển Việt – Anh (Vietnamese-English Dictionary). The World Publisher. HCM City (2001)
10. Foreign Languages University: Từ điển Anh-Việt (English-Vietnamese Dictionary). Vietnam National University of Hanoi. Education Publisher. Hanoi (1998)
11. The Institute of Linguistics of Vietnam: Loại từ trong các ngôn ngữ ở Việt Nam (Classifiers in Languages in Vietnam). Social Sciences Publisher. HaNoi (2000)
12. The Institute of Linguistics of Vietnam: Một số vấn đề Từ điển học (Some Issues in Lexicography). Social Sciences Publisher. HaNoi (1997)

## Appendix 1: List of Classifiers

**A. Using for inanimate nouns**

| No | Word | Examples |
|---|---|---|
| 1. | áng | văn (literary works) |
| 2. | bài | thơ (poem), diễn văn (speech), |
| 3. | bản | tuyên ngôn (declaration), tài liệu (materials), tiểu thuyết (novel) |
| 4. | bộ | từ điển (dictionary), máy (machine) |
| 5. | bông | hoa (flower), |
| 6. | bức | tranh (painting), thư (letter), tượng (statue), vách (wall), ảnh (photo) |
| 7. | **cái** | bàn (table), ghế (chair), đầu (head), thuyền (boat); khuyết điểm (fault), tâm trạng (emotional state) |
| 8. | cây | nến (candle), đèn (lamp), roi (rod), bút (pen), súng (gun), đàn (musical instrumen), tăm (tooth-stick) |
| 9. | căn | phòng (room), nhà (house) |
| 10. | chiếc | bàn (table), ghế (chair), thuyền (boat), |
| 11. | con | dao (knife), thuyền (boat), sông (river), |
| 12. | cuốn | sách (book), tập (notebook), vở (notebook), tiểu thuyết (novel) |
| 13. | đoá | hoa (flower) |
| 14. | hòn | đạn (bullet), bi (marble), núi (mountain) |
| 15. | khẩu | súng (gun), đại bác (cannon) |
| 16. | lá | bùa (charm), thư (letter), phiếu (ballot), đơn (application form) |
| 17. | làn | gió (wind) |
| 18. | màn | kịch (play), |
| 19. | món | quà (gift), nợ (debt), |
| 20. | nền | văn hoá (culture), độc lập (independence), khoa học (science), |
| 21. | nóc | nhà (house) |
| 22. | ngọn | cờ (flag), núi (mountain), |
| 23. | ngôi | nhà (house), đền (temple), chùa (pagoda), mộ (grave), sao (star) |
| 24. | pho | tượng (statue), truyện (story), sách (book) |
| 25. | quả | bom (bomb), núi (mountain), |
| 26. | quyển | sách (book), vở (notebook) |
| 27. | tấm | ảnh (photo), tranh (painting), bảng (board), bìa (card) |
| 28. | tấn | tuồng (play), kịch (comedy/tragedy) |
| 29. | toà | nhà (house), lâu đài (palace) |
| 30. | thanh | gươm (sword) |
| 31. | thửa | ruộng (field) |
| 32. | vì | sao (star), vua (king) |
| 33. | vở | kịch (comedy/tragedy), tuồng (play) |

**B. Using for animate nouns**

| No | Words | Examples |
|---|---|---|
| 34. | anh | sinh viên (student), cán bộ (official) |
| 35. | bà | chủ nhiệm (chief), vợ (wife) |
| 36. | bác | thợ (worker), phu xe (rickshaw puller) |
| 37. | chị | giáo viên (teacher), nhà báo (journalist) |
| 38. | bậc | vĩ nhân (great man), anh hùng (hero) |
| 39. | cái | Tí, Tỉu (proper names) |
| 40. | cậu | học trò (boy pupil), con trai (boy) |
| 41. | con | gián điệp (spy), mẹ mìn (child kidnapper) |
| 42. | **con** | trâu (buffalo), bò (cow/ox), gà (cook/hen/chicken) |
| 43. | cô | dược sĩ (dentist), y tá (nurse) |
| 44. | chàng | thi sĩ (poet), văn nhân (man of letters) |
| 45. | chú | liên lạc (contact man), tài xế (driver) |
| 46. | đấng | anh hùng (hero), thánh thần (the Holy Spirit) |
| 47. | em | học sinh (pupil), nhi đồng (child) |
| 48. | lão | quản gia (steward), tri huyện (district chief) |
| 49. | mụ | đàn bà (woman), vợ (wife) |
| 50. | nàng | công chúa (princess), tiên (fairy) |
| 51. | người | giáo viên (teacher), thợ nề (mason) |
| 52. | tay | thầu khoán (contractor), nhà buôn (dealer) |
| 53. | tên | sĩ quan địch (enemy officer), nguỵ binh (puppet army) |
| 54. | thằng | quỷ sứ (messenger of Satan), mật thám (investigator spy) |
| 55. | vị | phụ lão (aged man), chủ tịch (chairman) |
| 56. | viên | sĩ quan (officer), đại uý (captain) |

## Appendix 2: List of Words Denoting Categories

| No | Words | Examples |
|---|---|---|
| 1. | bông (flower) | hồng (rose), lài (jasmine) |
| 2. | cá (fish) | hồng (red snapper), thu (cod) |
| 3. | cây (tree) | tre (bamboo), chuối (banana) |
| 4. | chim (bird) | sẻ (sparrow), đa đa (partridge) |
| 5. | củ (root) | chuối (banana), ấu (caltrops), cải (beet) |
| 6. | hoa (flower) | hồng (rose), lài (jasmine) |
| 7. | máy (machine) | tính (computer), in (printer), bơm (pumper) |
| 8. | nhà (house) | máy (factory), đèn (power house), in (publisher) |
| 9. | rau (vegetable) | diếp (lettuce), cải (mustard greens) |
| 10. | quả (fruit) | chuối (banana), mít (breadfruit) |
| 11. | xe (vehicle) | đạp (bicyle), hơi (car), ủi (bulldozer) |

## Appendix 3: List of Derivational Affixes

### A. Derivational Prefixes

| No | Prefix | Vietnamese meanings | Examples |
|---|---|---|---|
| 1. | anti | kháng ~ | kháng thể (antibody) |
| 2. | bi | nhị ~ | nhị phân (binary) |
| 3. | co | đồng ~ | đồng tác giả (co-author) |
| 4. | de | khử ~, giải~ | giải mã (decode) |
| 5. | dis | xả ~ | xả điện (discharge) |
| 6. | hyper | siêu ~ | siêu văn bản (hypertext) |
| 7. | in, il, im, ir | bất ~, vô ~ | bất hợp pháp (illegal), bất quy tắc (ir-regular) |
| 8. | inter | liên ~ | liên ngành (interdiscipline) |
| 9. | meta | siêu ~ | siêu ngôn ngữ (metalanguage) |
| 10. | micro | vi ~ | vi xử lý (microprocessing), vi lệnh (microinstruction) |
| 11. | mid | trung ~ | trung thu (mid-autumn) |
| 12. | mono | đơn ~ | đơn điệu (monotone) |
| 13. | multi | đa ~ | đa phương tiện (multi-media) |
| 14. | non | phi ~ | phi chính phủ (non-government) |
| 15. | over | quá ~ | quá tải (overload) |
| 16. | para | cận ~ | cận ngôn ngữ (paralanguage) |
| 17. | photo | quang ~ | quang điện tử (photo-electronic) |
| 18. | post | hậu ~ | hậu xử lý (post-process) |
| 19. | pre | tiền ~ | tiền xử lý (pre-process) |
| 20. | pseudo | ~ giả | mã giả (pseudo-code) |
| 21. | re | tái ~ | tái chế (recycle) |
| 22. | self | tự ~ | tự kiểm (self-check), tự hành (self-acting) |
| 23. | semi | bán ~ | bán dẫn (semi-conductor) |
| 24. | sub | ~ con | chương trình con (subprogram) |
| 25. | super | siêu ~ | siêu dẫn (super-conductor) |
| 26. | tele | viễn ~ | viễn ấn (teletype) |
| 27. | tri | tam ~ | tam giác (triangle) |
| 28. | ultra | cực ~ | cực tím (ultraviolet) |
| 29. | un | bất ~ | bất hợp lý (unreasonable) |

B. **Derivational Suffixes**

| No | Suffix | Vietnamese meanings | Examples |
|---|---|---|---|
| 1. | able | khả ~ | khả biến (variable) |
| 2. | er | ~ viên, ~ sĩ, người ~, thợ ~ | teacher (giáo viên), thợ in (printer), lập trình viên (programmer) |
| 3. | er | máy ~, bộ ~ | printer (máy in), repeater (bộ lặp) |
| 4. | ian | ~ viên | kỹ thuật viên (technician) |
| 5. | ible | khả ~ | khả kiến (visible), khả thi (feasible) |
| 6. | ise/ize | ~ hóa | bình thường hoá (normalise), điện toán hoá (computerize) |
| 7. | ist | nhà ~, ~sĩ | nhà khoa học (scientist), nha sĩ (dentist) |
| 8. | less | bất ~ | bất cẩn (careless) |
| 9. | ology | ~ học | sinh vật học (biology), bản thể học (ontology) |
| 10. | or | người~ /máy ~/ bộ ~/ thiết bị ~/ đầu ~ | người biên tập (editor), máy phát (generator), bộ xử lý (processor), đầu nối (connector) |