AUSTROASIATIC AND TAI-KADAI LANGUAGES IN THE INTERCONTINENTAL DICTIONARY SERIES

Commrie Bernard , Periros Ilia

The study of the lexical diversity of human languages is, probably, one of the most exciting aspects of linguistics. However, it is also one of the most time consuming and complex parts of the discipline. Even very simple tasks require extraordinary efforts and could take a long time to complete.

Let us consider a typical example. In the process of researching into the cultural development of Mainland Southeast Asia, one would need to determine what is the word for a particular animal, for example - pig, in approximately two hundred local languages. To obtain this information, one could use several etymological or comparative dictionaries of these languages. These dictionaries, however, would only reveal details of words with reliable etymologies, while all other words would, naturally, be absent from this particular type of source. The rest of the words required for the research would, then, be obtained from various bilingual dictionaries, where forms are usually alphabetized according to the language in question, while the meanings are given in English, Chinese, Vietnamese, Thai, French, Dutch, Russian, German or Indonesian. Most of these dictionaries would use their own phonological representations with different notations for tones, vowel length and other features. Therefore, it would take approximately two weeks of intensive (and not very productive) work to get the information related to the meaning 'pig'.

The situation is quite different for the Tibeto-Burman languages of China. With the aid of "A Tibeto-Burman Lexicon" edited by Huang Bufan (1992), one could get the same information for fifty-one languages and dialects, in only a few minutes. This publication brought together words from different languages with identical meanings into a simply and concisely structured arrangement. Therefore, to establish the representations of the meaning 'pig' in some Tibeto-Burman languages, all one needs to do is simply open the dictionary and read the information given.

With no generally accepted term, one could call such a dictionary *'comparative synonym dictionary'*. It is comparative, as it provides data from various languages and it is also a dictionary of synonyms, as it lists forms from different languages with identical meanings.

Huang Bufan's book is not the only published comparative synonym dictionary (for the Tibeto-Burman languages see, for example Hale 1973 or Sun 1991) and is, in fact, a continuation of a long linguistic tradition. However, printed comparative-synonym dictionaries are complex to use, and their rich substance is often not fully exploited. Therefore, over the past years, general interest in such dictionaries has not been exceptionally strong.

Modern electronic databases have given a new direction to comparative synonym dictionaries and this opportunity is used in the Intercontinental Dictionary Series (IDS) project. Marie Ritchie Kay (University of California, Irvine) has developed this project with the aim of establishing a database, which would contain lexical material from

languages of the world organized in such a way as to allow comparisons to be made simply.

An entry on the IDS database will reveal how a particular meaning is conveyed in the selected languages. Therefore, an entry is, in fact, a list of synonyms found in the different languages, thus making the database a typical comparative synonym dictionary.

The final outcome of the project will be an electronic database representing the entire range of human languages. It will bring together information on languages of the world published in dozens of different languages and scripts and scattered in hundreds of publications and manuscripts, which are often not available to the linguistic community. This database will also include word-lists of less known languages collected especially for this project.

Currently the IDS project is predicted to be a set of interlinked electronic databases representing various language families or geographic regions. We are hopeful that eventually the IDS databases will contain most, if not all the languages in the world. Clearly, then, the IDS project is an enormous task, which may be achieved only through wide international cooperation among linguistic institutes and individual scholars.

Apart from being an archive of lexical data, the IDS database will also provide rich resources for various studies such as:

(i) evaluation of the linguistic diversity of regions and thus, for developing detailed and justifiable linguistic maps;

(ii) genetic and non-genetic classifications of languages;

(iii) typological studies based on lexicon, such as analysis of sound symbolism, word formation, phonological theory;

(iv) etymological and contact studies, such as identification of contact zones, an analysis of the spread of various cultural ideas, or proto-lexicons.

The IDS project is based on a list of meanings, which largely follows that used by Buck (1949) in his *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. Buck's dictionary contains approximately 1200 meanings, which, together with their numbering systems are incorporated into the IDS master list. Approximately 100 additional meanings have been added to it by M. R. Key.

Therefore this master list (1,310 entries) consists of:

1. universal meanings found in most human languages (HAND, SPEAK, DRY, etc.)
2. meanings related to certain geographical or environmental phenomena: EARTHQUAKE, TIDE, PARROT, etc.
3. meanings reflecting certain cultural ideas: MEAD, TATTOO, COBBLER, etc.

Naturally, not all meanings from groups (ii) and (iii) will be found in all the languages.

In some cases, it may be important to add extra meanings to represent information relevant to a particular language group or a region. Therefore, in the process of data collection, the list of the meanings will be expanded, but under no circumstances will the meanings be lost from the master list. If the corresponding meaning is not known in a particular language, the entry will simply remain blank. The reason for this strict rule is quite clear: by adding meanings, the main body of the databases is still kept compatible, while deleting meanings may create databases which are no longer compatible.

Several dictionaries based on the general principles of IDS have already been prepared, or are in preparation, including the *Comparative Austronesian Dictionary* edited by D. Tryon et al. (1995) and the *South American Dictionary* edited by M. R. Key.

One of the IDS projects which is currently in preparation is the "Austraoasiatic and Tai-Kadai comparative synonym dictionary". Its aim is to represent in the IDS format all the languages of these two families. At this stage, however, we are not including the Munda and Nicobarese branches of Austroasiatic.

The languages of these two groups are often spoken in the same areas of China, Thailand, Vietnam, Laos, Cambodia, Malaysia, Burma and India. As these languages also often reveal traces of intensive and sometimes old mutual contacts it is preferable to study the two families in the same project.

The preliminary work on the project includes a creation of the lists of the languages to be investigated and some additions to the IDS list of meanings. This modified list of meanings is to be found in the appendix.

A much more challenging task has been to compile a list of languages or dialects to be included in the investigation. With the exception of Myanmar (Burma), territories where Austroasiatic and Tai-Kadai languages are spoken are now open to linguistic research. This means, that new languages are constantly being brought to light, and that there is a significant chance that new discoveries will be made in all of these countries. On the other hand, these two linguistic families are not very popular among linguists and in some cases, a language has not been included in the project simply for the reason that no linguist was found to conduct the research.

The list of languages already included in the project is given below. This list has been compiled with the inestimable help of colleagues: G. Diffloth, M. Ferlus, Li Jinfang, Luo Yongxian, S. Morey, Nguyen Van Loi, Sun Hongkai, Suwilai Premsrirat, Theraphan L. Thongkum, Wilaiwan Khanittanan and Zhang Qiusheng.

Most of the Tai-Kadai languages are known relatively well, with the basic information lacking only for some members of the Ge-Yang branch. Therefore, the main goal of the Tai-Kadai part of this project is to focus on the finer details of the bigger picture.

The central role in the Tai-Kadai family is played by the Zhuang-Tai languages, which are divided into Shan-Tai (South-Western), Nung (Central) and Zhuang (Northern) branches.

The largest area associated with the Shan-Tai languages covers Thailand and Laos. Both of these countries have their own official languages (Thai and Lao), which will be included in the project. The linguistic situation here, however, is much more complex, as some dialects of Thailand have differences comparable to those observed between Thai and Lao. Therefore, the project will include six dialects from Thailand, which, hopefully, will represent the linguistic diversity of this country:

1. Central = Thai
2. Khamuang (Chiang Mai)
3. Southern Tai
4. Dialect of Nongkhai
5. Dialect of Khorat
6. Eastern Thai (Trat)

At this stage the Lao language is represented only by

7. the dialect of Luang Prabang.

The dialect of Vientiane is probably very close to that of Nongkhai.

The Shan-Tai languages of India are represented in the project by:

8. Khamti
9. Aiton

Aiton is a representative of four quite similar and mutually intelligible dialects spoken in Assam.

The Shan-Tai of Burma are represented by:

- two varieties of Shan:

10. Northern Shan and
11. Southern Shan, which is known in Thailand as Tai Yai
12. Khuen.

The Shan-Tai languages of China are spoken only in Yunnan. For this region, it has been decided that the data should be collected in the same way as in Thailand: covering the main locations where these languages / dialects are spoken:

13. Dehong (Chinese Dai)
14. Lue (Sipsonpanna Dai)
15. Jinsha Dai (probably the most northern dialect of Shan-Tao)
16. Jingping Dai
17. Tai Ya

Currently it is rather difficult to evaluate the linguistic diversity of these languages / *dialects.*

The list of the Shan-Tai languages of Vietnam to be included in the project has not been finalized yet. So far, it consists of:

18. White Tai
19. Black Tai
20. Red Tai
21. Tai Muong
22. Tai Nam

The Nung or Central dialects of Zhuang-Tai are spoken mainly in Vietnam and China, where they are now called "Southern Zhuang" languages. Five varieties (three from Guangxi, China and two from Vietnam) have been selected for the project:

23. Longzhou Nung, which is, probably, the best know representative of this group;
24. Debao Nung
25. Lazhai Nung
26. Nung Fan Slihng
27. Tay Nung

We are hopeful that these five dialects will represent the entire diversity of the Nung group.

Three dialects (languages) have been chosen for the Zhuang group:

28. Saek of Thailand / Laos border

29. Fengshan dialect of Guangxi, which according to the Chinese classification belongs to Northern Zhuang

30. Zhenning dialect, which according to the Chinese classification belongs to the Bouyi language.

Altogether, this project consists of twenty-nine representatives of the Zhuang-Tai family, some of which, however, may be very similar or even identical.

The project includes all main languages of the Kam-Sui family:

31. Southern Kam
32. Northern Kam
33. Mulam
34. Maonan
35. Mak
36. Yanghuang
37. Chadong

Tai-Kadai languages of China also include:

38. Ong Be (LinGao)
39. Lakkia
40. Biao

Three dialects (or languages) of the Li group:

41. Baoding
42. Jia Mao
43. Cun

The languages, which are provisionally attributed to the Ge -Yang (Kelao) group, are represented in the project as follows:

Buyang languages (China):

44. Ecun (Yunnan)
45. Langja (Yunnan)
46. Baha(Yunnan)
47. Yalhong (Guangxi)

Gelao languages (China):

48. Anshun Kelao
49. Moji Kelao
50. Sanchong Kelao
51. Qau Kelao

It is possible, however, that some other varieties of Kelao may be discovered.

52. Mulao (China)
53. Laji (China)
54 Bubiao (China) called Laqua in Vietnam
55. Laqi (Vietnam)
56. Laha (Vietnam)
57. Nung Ven (Vietnam)

The internal structure of the Ge –Yang languages as well as their complete list remain to be investigated.

The Austroasiatic languages are much less well known and for this family, the tendency is to include all languages or major dialects, which may be investigated. Their list is:

Khmer:

1. Standard Khmer (Cambodia)
2. Surin Khmer (Thailand)

Monic languages:

3. Mon (Thailand)
4. one of the Nyakur dialects (Thailand)

Bahnaric languages:

of Vietnam:

5. Bahnar
6. Jeh
7. Halang
8. Hre
9. Sedang
10. Ka Dong =KaYoeng
11. Cua = Takua = Kor
12. Todrah
13. Rengao
14. Mo'Na\m =Po'noong
15. Stieng
16. Keho, Maa, Sre
17. Chrau
18. Eastern Mnong

of Cambodia:

19. Mnong
20. Brao- Krung = Laveh
21. Tampuon
22. Kaco'

of Laos:

23. Alak
24. Loven = Jru
25. Suk
26. Oi
27. Nhaheun

Katuic languages:

of Thailand

28. Kuai
29. Kuai Yoe
30. So = Tri
31. Bruu

of Vietnam:

32. Bruu Van Kieu
33. Pakoh
34. Katu
35. Taoi

of Cambodia:

36. Kuai Ndroe

of Laos:

37. Ngeq = Kriang

Aslian languages are represented in the project only by:

38. Kensiw (Malaysia, Thailand)
39. Semelai (Malaysia)

as so far contributors for other languages of this branch have not been found.

Vietic languages:

of Vietnam:

40. Vietnamese
41. Northern Mu'o'ng
42. Central Mu'o'ng
43. Southern Mu'o'ng
44. Nguon
45. Arem
46. Sach
47. Ruc
48. May
49. Maleng
50. Pong
51. Cuoi
52. Tho
53. Li Ha

of Laos:

54. Thavung (So)
55. Aho or Phon Soung
56. Kri
57. Atel
58. Bru-Maleng

Palaung-Wa languages:

of China:

59. Wa
60. Plang
61. Bangpin Angku
62. Kuang Angku
63. U Angku
64. De'ang
65. Rumai

of Laos:

66. Lamet = Kamet
67. Amok

Khmuic languages:

of China and Laos:

68. Khabit = Pusing

of Thailand:

69. Mlabri
70. Prai = Thin
71. Mal = Thin

of Vietnam:

72. Ksinmul
73. Khang

74. Iduh

of Laos:

75. Khmu
76. Kaniang = Pong
77. Mang (Vietnam, China)
78. Bugan (China)
79. Paliu / Lai (China)

Pearic languages:

of Thailand:

80. Chong
81. Kasong
82. Samre

of Cambodia:

83. Sa'oc
84. $Samre_2$
85. Pear

So far only one representative of the Khasi dialects (India) is included in the project:

86. Khynriam or Nonglum dialect

This preliminary list consists only of languages and dialects that in our point of view are likely to be studied in the project. In most cases, the potential contributors have already been found. Further additions to the list as well as help in data collection are very welcome.

The final outcome of this project will bean electronic database consisting of language data publicly available on the Internet.

It is also planned to compile a CD with a collection of the digital recording of the data suitable for various instrumental studies.

It is clear, however, that the success of the project depends totally on the broad international cooperation.

REFERENCES

1. Buck, C.D. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*, Chicago: University of Chicago Press, 1949.
2. Hale, A. *Clause, Sentence, and Discourse Patterns in selected languages of Nepal*, part 4. Norman: Summer Institute of Linguistics, 1973.
3. Huang Bufan, ed. *A Tibeto-Burman Lexicon*. Beijing: Central Institute of Nationalities 1992.
4. Key, M. R., ed. *South American Dictionary*, MS.
5. *Languages and Cultures of the Kam-Tai (Zhuang-Dong) Group: a word list*. Bangkok: Mahidol University 1996.
6. Sun, Hongkai et. al., compilers. *Tibeto-Burman phonetics and word-lists*. Beijing: Chinese Social Science Press. 1991.
7. 920 Tryon, D. T., ed. *Comparative Austronesian Dictionary*. Berlin: Mouton DeGruyter, 1995.

APPENDIX

The list of meanings adopted for the Austroasiatic and Kadai IDS dictionary

Meanings added for the Austroasiatic and Tai-Kadai project are marked as [ADD]

1.100	world
01.210	earth, land
01.212	earth = ground, soil
01.213	dust
01.214	mud
01.215	sand
01.220	hill, mountain
01.221	anthill [ADD]
01.222	cliff, precipice
01.230	plain, field
01.240	valley
01.250	island
01.260	mainland
01.270	bank (of river), shore
01.280	cave
01.310	water
01.320	sea
01.322	calm (of sea)
01.323	rough (of sea)
01.324	foam
01.329	ocean
01.330	lake
01.331	pond, pool [ADD]
01.340	bay, gulf
01.341	lagoon
01.342	reef
01.343	headland, point
01.350	wave
01.352	tide
01.353	low tide
01.354	high tide
01.360	river, stream, brook
01.362	whirlpool
01.370	spring, well
01.380	swamp
01.390	waterfall
01.410	forest, woods
01.420	tree
01.430	wood
01.440	stone, rock
01.441	gravel [ADD]
01.450	earthquake
01.510	sky
01.520	sun
01.530	moon
01.540	star
01.550	lightning
01.560	thunder
01.561	thunder (v.)
01.570	lightning(as striking)
01.580	storm
01.590	rainbow
01.610	light,sunlight (noun)
01.620	darkness
01.630	shade, shadow
01.640	dew
01.710	air
01.720	wind
01.730	cloud
01.740	fog
01.750	rain (noun)
01.760	snow (noun)
01.770	ice
01.771	hail [ADD]
01.780	weather
01.810	fire
01.820	flame (noun)
01.830	smoke (noun)
01.840	ashes
01.841	hot embers

01.851	burn (vb trans)
01.852	burn (vb intrans)
01.860	ignite, light, kindle
01.861	extinguish
01.870	match (noun)
01.880	firewood
01.890	charcoal
02.100	human being, person
02.210	man (vs.woman)
02.220	woman
02.230	male
02.240	female
02.250	boy
02.251	young man (adolescent)
02.260	girl
02.261	young woman (adolescent)
02.270	child
02.280	infant, baby
02.310	husband
02.320	wife
02.330	to marry
02.340	marriage, wedding
02.341	divorce
02350	father
02360	mother
02370	parents
02.380	married man
02.390	married woman
02.410	son
02.420	daughter
02.430	offspring (son or daughter)
02.440	brother
02.444	older brother
02.445	younger brother
02.450	sister
02.454	older sister
02.455	younger sister
02.456	sibling
02.458	twins
02.460	grandfather
02.461	old man
02.470	grandmother
02.471	old woman
02.480	grandson
02.490	granddaughter
02.510	uncle
02.511	mother's brother
02.512	father's brother
02.520	aunt
02.521	mother's sister
02.522	father's sister
02.530	nephew
02.540	niece
02.550	cousin
02.560	ancestors
02.570	descendants
02.610	father-in-law(of a man)
02.611	father-in-law (of a woman)
02.620	mother-in-law (of a man)
02.621	mother-in-law (of a woman)
02.630	son-in-law (of a man)
02.631	son-in-law (of a woman)
02.640	daughter-in-law(of a man)
02.641	daughter-in-law (of a woman)
02.710	stepfather
02.720	stepmother
02.730	stepson
02.740	stepdaughter
02.750	orphan
02.760	widow
02.770	widower
02.810	kinsmen, relatives
02.820	family
02.910	I

02.920 you (singular)
02.930 it/he/she
02.940 we
02.941 we (inclusive)
02.942 we (exclusive)
02.943 we two (inclusive)
02.943 we two (exclusive)
02.950 you (plural)
02.951 you Dual
02.960 they
03.110 animal
03.120 male (adj)
03.130 female (adj)
03.150 livestock
03.160 pasture
03.180 herdsman
03.190 stable, stall
03.200 cattle (bovine)
03.210 bull
03.220 ox
03.230 cow
03.240 calf
03.241 water buffalo
03.250 sheep
03.260 ram
03.280 ewe
03.290 lamb
03.320 boar
03.340 sow
03.350 pig
03.360 goat
03.370 he-goat
03.380 kid
03.410 horse
03.420 stallion
03.440 mare
03.450 colt, foal
03.450 foal, colt
03.460 ass, donkey
03.470 mule
03.500 fowl
03.520 rooster, cock
03.540 hen
03.550 chicken
03.560 goose
03.570 duck
03.580 nest
03.581 bird
03.582 seagull
03.583 heron
03.584 eagle
03.585 hawk
03.586 vulture
03.587 kite [ADD]
03.591 bat
03.592 parrot
03.593 crow
03.594 dove
03.595 pigeon
03.596 owl
03.597 quail [ADD]
03.598 swallow [ADD]
03.599 peacock [ADD]
03.600 pheasant [ADD]
03.610 dog
03.614 rabbit
03.615 pangolin [ADD]
03.618 panther [ADD]
03.619 tiger [ADD]
03.620 cat
03.621 civet cat [ADD]
03.622 opossum
03.630 mouse, rat
03.650 fish

03.652 fin (dorsal)
03.653 fishscale
03.654 gill
03.655 shell
03.656 crab
03.657 shrimp [ADD]
03.661 shark
03.662 dolphin, porpoise
03.663 whale
03.664 stingray
03.665 eel (freshwater)
03.710 wolf
03.720 lion
03.730 bear
03.740 fox
03.741 squirrel [ADD]
03.742 flying squirrel[ADD]
03.743 porcupine [ADD]
03.744 otter [ADD]
03.750 deer
03.751 barking deer [ADD]
03.752 wild ox, gaur [ADD]
03.760 monkey
03.770 elephant
03.771 trunk of elephant[ADD]
03.780 camel
03.810 insect
03.811 louse (1)
03.811 louse (2)
03.812 louse egg (nit)
03.813 flea
03.814 centipede
03.815 scorpion
03.816 cockroach
03.817 ant
03.818 spider
03.819 spider web
03.820 bee
03.821 wax (bees)
03.822 beehive
03.823 wasp
03.824 termite [ADD]
03.830 fly (noun)
03.831 gnat,sandfly (midge)
03.832 mosquito
03.833 woodborer [ADD]
03.839 caterpillar
03.840 worm
03.841 maggot [ADD]
03.842 cicada [ADD]
03.843 cricket [ADD]
03.844 water leech [ADD]
03.845 land leech [ADD]
03.850 snake
03.851 python [ADD]
03.910 firefly
03.920 butterfly
03.930 grasshopper
03.940 snail
03.950 frog
03.960 lizard
03.961 monitor [ADD]
03.962 gecko [ADD]
03.970 alligator, crocodile
03.980 turtle
04.110 body
04.120 hide, skin
04.130 flesh
04.140 hair (head)
04.142 beard
04.144 body hair
04.145 pubic hair
04.146 dandruff
04.147 sweat [ADD]

04.150 blood

04.151 artery, vein

04.152 tendon,sinew [ADD]

04.160 bone

04.161 marrow

04.162 rib

04.170 horn

04.180 tail

04.190 back

04.191 spine

04.200 head

04.201 side of head, temple

04.202 skull

04.203 brain

04.204 face

04.205 forehead

04.207 jaw

04.208 cheek

04.209 chin

04.210 eye

04.212 eyebrow

04.213 eyelid

04.214 eyelash

04.215 blink

04.215 blink

04.220 ear

04.221 earlobe

04.222 earwax

04.230 nose

04.231 nostril

04.232 mucus (nasal)

04.240 mouth

04.241 beak

04.250 lip

04.260 tongue

04.270 tooth

04.271 tusk(of elephant)[ADD]

04.271 gums

04.272 molar tooth

04.280 neck

04.281 nape of neck

04.290 throat

04.300 shoulder

04.301 shoulder blade

04.302 collarbone

04.310 arm

04.312 armpit

04.320 elbow

04.321 wrist

04.330 hand

04.331 palm of hand

04.340 finger

04.342 thumb

04.344 fingernail

04.345 claw

04.350 leg

04.351 thigh

04.352 calf of leg

04.360 knee

04.370 foot

04.371 ankle

04.372 heel

04.374 footprint

04.380 toe

04.392 wing

04.393 feather

04.400 chest

04.410 breast (of woman)

04.412 nipple, teat

04.420 udder

04.430 navel

04.440 heart

04.441 lung

04.450 liver

04.451 kidney
04.452 spleen
04.453 gall, bile [ADD]
04.459 belly [ADD]
04.460 stomach
04.461 guts, intestines
04.462 waist
04.463 hip
04.464 buttocks
04.465 anus [ADD]
04.470 womb
04.490 testicle
04.492 penis
04.493 vagina [ADD]
04.510 breathe, breath
04.520 gape, yawn
04.521 hiccough
04.530 cough
04.540 sneeze
04.550 perspire
04.560 saliva, spit
04.561 spit (v.)[ADD]
04.570 vomit
04.580 bite
04.589 stick out (tongue)[ADD]
04.590 lick
04.591 dribble
04.610 sleep
04.612 snore
04.620 dream
04.630 wake up
04.631 awake [ADD]
04.640 flatulence, break wind
04.650 urinate
04.660 defecate
04.670 have sexual intercourse
04.680 shiver
04.690 bathe
04.710 beget (of father)
04.720 birth, born (to be)
04.721 to give birth [ADD]
04.730 pregnant
04.732 conceive
04.740 alive, living, life
04.750 dead, die
04.751 drowned
04.760 kill
04.770 corpse
04.780 bury (the dead)
04.790 grave, tomb
04.810 mighty, strong, powerful
04.820 weak
04.830 health, well
04.840 ill, sick, sickness
04.841 fever
04.842 goiter
04.843 cold (catarrh)
04.844 diarrhea [ADD]
04.850 wound, sore
04.852 bruise
04.853 swelling
04.854 swell (v.)[ADD]
04.854 itch
04.855 blister
04.856 boil (noun)
04.857 pus
04.858 scar (1)
04.860 cure, heal
04.870 doctor, physician
04.880 medicine, drug
04.890 poison
04.891 poison (verb)[ADD]
04.910 tired, weary
04.912 rest

04.920 lazy
04.930 bald
04.940 lame
04.950 deaf
04.960 mute
04.970 blind
04.980 intoxicated, drunk
04.990 bare, naked
05.110 eat
05.120 food
05.121 cooked
05.122 raw
05.123 ripe
05.124 green, unripe
05.125 spoiled, rotten
05.130 drink
05.140 hunger
05.141 famine
05.150 thirst
05.160 suck
05.180 chew
05.181 swallow
05.190 choke
05.210 cook
05.220 boil (vb)
05.230 roast, fry
05.240 bake
05.250 oven
05.260 cooking vessel, pot
05.270 kettle
05.280 pan
05.310 dish
05.320 plate
05.330 bowl
05.340 jug, pitcher
05.350 cup, drinking vessel
05.360 saucer
05.370 spoon
05.380 knife
05.390 fork
05.391 tongs
05.410 meal (a)
05.420 breakfast
05.430 lunch
05.440 dinner
05.450 supper
05.460 peel (v.)
05.470 sieve, strain
05.480 scrape
05.490 mix, stir
05.510 bread
05.530 dough
05.540 knead
05.550 flour, meal
05.559 pound v.[ADD]
05.560 crush, grind
05.561 to winnow
05.570 mill
05.580 mortar
05.590 pestle
05.610 meat
05.630 sausage
05.640 broth, soup
05.650 vegetables
05.660 bean
05.661 turmeric Curcuma[ADD]
05.662 garlic [ADD]
05.663 onion [ADD]
05.664 chili pepper [ADD]
05.665 sesame [ADD]
05.666 ginger
05.667 cucumber[ADD]
05.668 egg plant
05.700 potato

05.710 fruit
05.712 bunch
05.750 fig
05.760 grape
05.770 nut
05.780 olive
05.790 oil
05.791 fat, grease
05.792 fat (ad.)[ADD]
05.793 be lean(not fat)ADD]
05.810 salt
05.820 pepper
05.821 chili pepper
05.840 honey
05.850 sugar
05.860 milk (noun)
05.870 milk (vb)
05.880 cheese
05.890 butter
05.900 beverage, drink
05.910 mead
05.920 wine
05.930 beer
05.940 drink (fermented)
05.970 egg
05.971 egg yolk
06.110 dress (vb), put on
06.120 clothing, clothes
06.130 tailor
06.210 cloth
06.220 wool
06.230 flax, linen
06.240 cotton
06.250 silk
06.270 felt
06.280 fur
06.290 leather
06.310 spin
06.320 spindle
06.330 weave
06.340 loom
06.350 sew
06.360 needle
06.370 awl
06.380 thread
06.390 dye
06.410 cloak
06.411 poncho
06.420 woman's dress
06.430 coat
06.440 shirt
06.450 collar
06.460 skirt
06.461 grass-skirt
06.480 trousers
06.481 loincloth[ADD]
06.490 sock, stocking
06.510 shoe
06.520 boot
06.540 shoemaker,cobbler
06.550 hat, cap
06.570 belt, girdle
06.580 glove
06.590 veil
06.610 pocket
06.620 button
06.630 pin
06.710 adornment,ornament
06.720 jewel
06.730 ring (for finger)
06.740 bracelet
06.750 necklace
06.760 bead
06.770 earring

06.780 headband,headdress
06.790 tattoo
06.810 handkerchief, rag
06.820 towel
06.910 comb
06.911 comb (v.) [ADD]
06.920 brush
06.921 plait, braid
06.930 razor
06.940 ointment
06.950 soap
06.960 mirror
07.110 live, dwell
07.120 house
07.130 hut
07.131 garden-house
07.140 tent
07.150 court, yard
07.160 men's house
07.170 cookhouse
07.180 meeting house
07.210 room
07.220 door, gate
07.221 doorpost, jamb
07.230 lock (noun)
07.231 door-bolt, latch
07.240 key
07.250 window
07.260 floor
07.270 wall
07.310 fireplace
07.320 stove
07.330 chimney
07.370 ladder
07.420 bed
07.421 pillow
07.422 blanket
07.430 chair
07.440 table
07.450 lamp, torch
07.460 candle
07.470 shelf
07.480 trough
07.510 roof
07.520 thatch
07.521 thatch grass [ADD]
07.530 ridgepole
07.540 rafter
07.550 beam
07.560 pole, post
07.570 board
07.580 arch
07.610 mason
07.620 brick
07.630 mortar, cement
07.640 adobe
08.110 farmer
08.120 field
08.121 field, dry [ADD]
08.130 garden
08.150 cultivate, till
08.160 fence
08.170 ditch
08.210 plow
08.212 furrow
08.220 to dig
08.230 spade
08.240 shovel
08.250 hoe
08.260 fork
08.270 rake
08.310 sow (seed)
08.311 seed
08.320 mow, reap

08.330 scythe, sickle
08.340 to thresh
08.350 threshing-floor
08.410 crop, harvest
08.420 grain
08.430 wheat
08.440 barley
08.450 rye
08.460 oats
08.470 corn, maize
08.480 rice (cooked)
08.481 paddy [ADD]
08.482 husked rice [ADD]
08.482 rice seedlings[ADD]
08.483 bran; husk[ADD]
08.485 rice straw [ADD]
08.486 millet [ADD]
08.510 grass
08.520 hay
08.530 plant (noun)
08.531 plant (vb)
08.540 root
08.541 tree stump[ADD]
08.550 branch
08.560 leaf
08.561 thorn [ADD]
08.570 flower
08.600 tree
08.610 oak
08.620 beech
08.630 birch
08.640 pine
08.650 fir
08.660 acorn
08.670 vine
08.680 tobacco
08.690 smoke (tobacco)
08.690 tobacco (to smoke)
08.691 pipe
08.692 areca, betel [ADD]
08.720 tree stump
08.730 tree trunk
08.740 forked branch
08.750 bark
08.760 sap
08.810 palm tree
08.820 coconut
08.830 citrus fruit
08.831 mango [ADD]
08.840 banana tree
08.841 banana [ADD]
08.850 banyan
08.851 tamarind [ADD]
08.852 papaya [ADD]
08.853 guava [ADD]
08.854 kapok tree[ADD]
08.910 potato (sweet)
08.912 yam [ADD]
08.913 taro [ADD]
08.920 manioc, tapioca, cassava
08.930 gourd
08.931 pumpkin, squash
08.938 rattan [ADD]
08.939 bamboo shoots
08.940 bamboo
08.941 cane (sugar)
08.960 poison (root) fish
08.970 nettle
08.980 mushroom
09.110 make, do
09.120 work (v.)
09.140 bend
09.150 fold
09.160 bind, tie

09.161	untie
09.180	chain
09.190	cord, rope
09.192	knot
09.210	beat, strike, hit
09.210	beat, strike, hit
09.211	pound with fist
09.220	cut
09.221	cut down
09.222	chop, hew
09.223	pierce, stab
09.230	knife
09.240	scissors, shears
09.250	axe
09.251	adze
09.260	break
09.261	broken
09.270	to split
09.280	tear (vb)
09.290	flay, skin
09.310	rub
09.311	wipe [ADD]
09.320	stretch
09.330	pull
09.331	pull out [ADD]
09.340	spread out
09.341	hang up
09.342	press
09.343	squeeze, wring
09.350	pour
09.360	wash
09.370	sweep
09.380	broom
09.422	tool
09.430	carpenter
09.440	build
09.460	pierce, bore
09.461	hollow out
09.480	saw
09.490	hammer
09.500	nail
09.560	glue
09.600	blacksmith
09.610	forge
09.620	anvil
09.630	cast (metals)
09.640	gold
09.650	silver
09.660	bronze, copper
09.670	iron
09.680	lead (noun)
09.690	tin, tinplate
09.710	potter
09.720	mold (clay etc)
09.730	clay
09.740	glass
09.750	plait, weave
09.760	basket
09.761	winnowing basket[ADD]
09.770	mat
09.771	rug
09.780	bag (net)
09.790	fan (noun)
09.791	fan (vb)
09.810	carve
09.820	sculptor
09.830	statue
09.840	chisel
09.880	paint (noun)
09.890	paint (vb)
10.110	move
10.120	turn over
10.130	turn around
10.140	wind, wrap

10.150 roll

10.160 drop (vb)

10.170 twist

10.170 twist

10.210 rise

10.220 lift, raise

10.230 fall

10.240 drip

10.250 throw

10.251 throw out, throw away [ADD]

10.252 catch (ball)

10.260 shake

10.320 flow

10.330 sink

10.331 flood (v.)[ADD]

10.340 float

10.350 swim

10.351 dive

10.352 splash

10.360 sail (vb)

10.370 to fly

10.380 blow

10.410 crawl, creep

10.412 kneel

10.413 crouch

10.420 slide, slip

10.430 jump, leap

10.431 kick

10.440 dance

10.450 walk

10.451 limp

10.460 run

10.470 go

10.471 go up ascend

10.472 climb

10.473 go down descend

10.474 go out

10.480 come

10.479 step on [ADD]

10.481 come back

10.481 return, come back

10.490 depart, go away

10.491 to disappear

10.510 flee

10.520 follow

10.530 pursue

10.550 arrive, reach

10.560 approach

10.570 enter

10.610 carry (bear)

10.612 carry-in-hand

10.613 carry-on-shoulder

10.614 carry-on-head

10.615 carry-underarm

10.616 carry on back[ADD

10.617 carry with pole[ADD]

10.620 bring

10.630 to send

10.640 lead (vb)

10.650 drive

10.660 ride

10.670 push, shove

10.710 road

10.720 path

10.740 bridge

10.750 carriage,wagon, cart

10.760 wheel

10.770 axle

10.780 yoke

10.810 ship

10.830 boat

10.831 canoe

10.832 outrigger

10840 balsa, raft

10.850	oar
10.851	paddle (noun)
10852	row (vb)
10.860	rudder
10.870	mast
10.880	sail (noun)
10.890	anchor
10.910	harbor, port
10.920	land (vb)
11.110	have
11.120	own, possess
11.130	take
11.140	grasp, seize
11.150	hold
11151	hold (in mouth)[ADD]
11.160	get, obtain
11.180	thing
11.210	give
11.220	return, give back
11.240	look after, preserve
11.250	rescue, save
11.270	destroy
11.280	damage, harm, injure
11.310	look for, seek
11.320	find
11.330	lose
11340	let go, release
11.430	money
11.440	coin
11.510	rich
11.520	poor
11.530	beggar
11.540	avaricious, stingy
11.610	lend
11.620	borrow
11.620	borrow (2)
11.630	owe
11.640	debt
11.650	pay (vb)
11.660	account, reckoning
11.690	tax, tribute
11.770	hire
11.780	wages
11.790	earn
11.810	buy
11.820	sell
11.830	barter, trade
11.840	merchant
11.850	market (place)
11.860	store, shop
11.870	price
11.880	dear (costly, expensive)
11.890	cheap
11.910	distribute, share
11.920	weigh
12.010	after
12.011	behind
12.020	beside
12.030	below, down
12.040	before
12.041	front
12.050	in, inside
12.060	outside
12.070	under
12.080	up, above
12.110	place
12.120	put
12.130	sit
12.140	lie down
12.150	stand
12.160	stay, remain
12.161	wait [ADD]
12.170	left overs, remains
12.210	collect, gather

12.212	pick up
12.213	pile up
12.220	join, unite
12.230	separate
12.232	divide
12.240	open
12.241	to open one's eyes[ADD]
12.250	close, shut
12.251	close(the eyes)[ADD]
12.260	cover
12.270	hide, conceal
12.310	high
12.320	low
12.330	top
12.340	bottom
12.350	end
12.352	pointed
12.353	border, edge
12.360	side
12.370	center, middle
12.371	between [ADD]
12.410	right (side)
12.420	left (side)
12.430	near (adv)
12.440	far (adv)
12.450	east
12.460	west
12.470	north
12.480	south
12.530	grow
12.540	measure
12.541	fathom, span
12.550	big, large
12.560	little,small(of size)
12.570	long
12.580	tall
12.590	short
12.610	broad, wide
12.620	narrow
12.630	thick (in dimension)
12.631	thick(liquid)[ADD]
12.650	thin (in dimension)
12.670	deep
12.680	shallow
12.710	flat
12.730	straight
12.740	crooked
12.750	hook
12.760	corner
12.770	cross
12.780	square
12.810	round
12.820	circle
12.830	ball, sphere
12.840	line
12.850	hole
12.920	like, similar
12.930	change
13.000	nothing, zero
13.010	one
13.020	two
13.030	three
13.040	four
13.050	five
13.060	six
13.070	seven
13.080	eight
13.090	nine
13.100	ten
13.101	eleven
13.102	twelve
13.103	fifteen
13.104	twenty
13.105	hundred

13.106 thousand
13.107 count
13.140 all, every
13.150 many, much
13.160 more
13.170 few, little
13.180 enough
13.181 some
13.190 crowd, multitude
13.210 full
13.211 full(from eating)[ADD]
13.220 empty
13.230 part, piece
13.240 half
13.330 alone, only
13.340 first
13.350 last
13.360 second
13.370 pair
13.380 two times
13.420 third
13.440 three times
14.110 time
14.120 age
14.130 new
14.140 young
14.149 old (not new) [ADD]
14.150 old, not young
14.151 grey (hair)[ADD]
14.160 early
14.170 late
14.180 now
14.190 immediately
14.210 fast, swift, quick
14.220 slow
14.230 hasten, hurry
14.240 delay, retard
14.250 begin, beginning
14.252 endure, last
14.260 end (temporal)
14.270 finish
14.280 cease, stop
14.290 ready
14.310 always
14.320 often
14.330 sometimes
14.331 soon
14.332 long-time (for a)
14.340 never
14.350 again
14.410 day
14.420 night
14.430 dawn
14.440 morning
14.450 midday, noon
14.451 afternoon
14.460 evening
14.470 today
14.480 tomorrow
14.481 day-after-tomorrow
14.490 yesterday
14.491 day-before-yesterday
14.510 hour
14.530 clock, timepiece
14.610 week
14.620 Sunday
14.630 Monday
14.640 Tuesday
14.650 Wednesday
14.660 Thursday
14.670 Friday
14.680 Saturday
14.710 month
14.730 year

14.740 winter
14.750 spring (season)
14.760 summer
14.770 fall, autumn
14.771 dry season[ADD]
14.772 rainy season [ADD]
14.780 season
15.210 smell (vb intrans)
15.212 sniff
15.220 smell (vb trans)
15.250 fragrant,good smelling
15.260 odor, stinking, bad smelling
15.310 taste
15.311 bland,tasteless[ADD]
15.350 sweet
15.360 salty
15.370 bitter
15.380 sour, acid
15.390 brackish
15.410 hear
15.420 listen
15.440 noise, sound
15.450 loud
15.460 quiet, silence
15.510 see
15.520 look, look at
15.550 show
15.560 shine
15.570 bright
15.610 color
15.620 light (in color)
15.630 dark (in color)
15.640 white
15.650 black
15.660 red
15.670 blue
15.680 green
15.690 yellow
15.710 touch
15.712 pinch
15.720 feel
15.740 hard
15.750 soft (to the touch)
15.751 soft,flexible[ADD]
15.760 rough
15.770 smooth
15.780 sharp
15.790 blunt, dull
15.810 heavy
15.820 light (in weight)
15.830 damp, wet
15.840 dry, as firewood
15.840 dry(of weather)[ADD]
15.841 dry (verb)[ADD]
15.850 hot
15.851 warm
15.860 cold
15.870 clean
15.880 dirty, soiled
15.890 wrinkled
16.110 soul, spirit
16.150 astonished, surprise
16.150 surprised, astonished
16.180 good fortune, luck
16.190 bad luck, misfortune
16.230 glad, joyful, happy
16.250 laugh
16.251 smile
16.260 play
16.270 love
16.290 kiss
16.300 embrace
16.310 pain
16.320 to be sad, grief, sorrow, sadness

16.330 anxiety, worry
16.340 regret, be sorry
16.350 compassion, pity
16.370 cry, weep
16.380 tear (noun)
16.390 groan
16.410 hate
16.420 anger, angry
16.440 envy, jealousy
16.450 shame (noun)
16.480 proud
16.510 dare
16.511 shy,be ashamed [ADD]
16.520 brave
16.530 fear, fright
16.540 danger
16.620 desire, want
16.622 choose
16.630 hope
16.650 faithful
16.660 true
16.670 tell lies, lie
16.680 deceit
16.690 forgive
16.710 good
16.720 bad
16.730 correct, right
16.740 wrong
16.760 fault
16.770 error, mistake
16.780 blame
16.790 praise
16.810 beautiful
16.820 ugly
16.830 greedy
16.840 clever
17.110 mind
17.130 think (= reflect)
17.140 think that
17.150 believe
17.160 understand
17.170 know
17.171 guess
17.172 imitate
17.180 seem
17.190 idea, notion
17.210 wise
17.220 foolish, stupid
17.230 insane, crazy
17.240 learn
17.242 study
17.250 teach
17.260 pupil
17.270 teacher
17.280 school
17.310 remember
17.320 forget
17.340 clear, plain
17.350 obscure
17.360 secret
17.370 certain, sure
17.380 explain
17.410 intention, purpose
17.420 cause
17.430 doubt
17.440 suspect
17.441 betray
17.450 need, necessity
17.460 easy
17.470 difficult
17.470 difficult (2)
17.480 attempt, try
17.490 manner, way
17.510 and

17.520 because
17.530 if
17.540 or
17.550 affirmative, yes
17.560 no, negative
17.610 how?
17.620 how many?
17.630 how much?
17.640 what?
17.650 when?
17.660 where?
17.670 which?
17.680 who?
17.690 why?
17.691 this [ADD]
17.692 here [ADD]
17.693 that
17.694 there[ADD]
18.110 voice
18.120 sing
18.121 song [ADD]
18.130 cry out, shout
18.150 whisper
18.160 mumble
18.170 whistle
18.180 screech, shriek
18.190 howl
18.210 speak, talk
18.211 stammer, stutter
18.220 say
18.221 tell story
18.222 speech (make a)
18.230 silent (be)
18.240 language
18.260 word
18.280 name
18.310 ask (question, inquire)
18.320 answer
18.330 admit, confess
18.340 deny
18.350 ask, request
18.360 promise
18.370 refuse
18.380 forbid
18.390 rebuke, scold
18.410 call (=summon)
18.420 call (=name)
18.430 announce
18.440 threaten
18.450 boast
18.510 write
18.520 read
18.560 paper
18.570 pen
18.610 book
18.670 poet
18.710 flute
18.720 drum
18.730 horn, trumpet
18.740 rattle
19.110 country
19.120 one's native country
19.150 city, town
19.160 village
19.170 boundary
19.210 people (populace)
19.230 clan, tribe
19.240 chief, chieftain
19.250 staff, walking stick
19.310 govern, rule
19.320 king, ruler
19.330 queen
19.360 noble, nobleman
19.370 citizen, subject

19.410 master
19.420 slave
19.430 servant
19.440 freeman
19.450 command, order
19.460 obey
19.470 allow, let, permit
19.510 friend, companion
19.520 enemy
19.540 neighbor
19.550 stranger
19.560 guest
19.570 host
19.580 help, aid
19.590 hinder, prevent
19.610 custom
19.620 quarrel, strife
19.630 conspiracy, plot
19.650 meet
19.720 prostitute
20.110 fight (vb)
20.130 war, battle
20.140 peace
20.150 army
20.170 soldier
20.210 arms, weapons
20.220 club
20.222 battle-axe
20.230 sling
20.240 bow
20.250 arrow
20.260 spear
20.270 sword
20.280 gun, cannon
20.310 armor (defensive)
20.330 helmet
20.340 shield
20.350 fortress
20.360 tower
20.410 victory
20.420 defeat
20.430 attack
20.440 defend
20.450 retreat
20.460 surrender
20.470 captive, prisoner
20.471 guard, sentinel
20.480 booty, spoils
20.490 ambush
20.510 fisherman
20.511 fish (v.)
20.520 fishhook
20.530 fishing line
20.540 fishnet
20.550 fish trap
20.560 bait
20.610 hunt
20.611 hunter [ADD]
20.620 to shoot
20.621 to aim (a target)[ADD]
20.630 miss (target)
20.640 trap (noun)
20.650 trap (vb)
21.110 law
21.150 court
21.160 judge (vb)
21.170 judgment
21.180 judge (noun)
21.210 plaintiff
21.220 defendant
21.230 witness
21.240 swear
21.250 oath
21.310 accuse

21.320	condemn
21.330	convict (vb)
21.340	acquit
21.350	guilty
21.360	innocent
21.370	penalty, punishment
21.380	fine
21.390	jail, prison
21.420	murder
21.430	adultery
21.440	rape
21.460	fire, arson
21.470	perjury
21.510	steal
21.520	robber
21.521	thief [ADD]
22.110	religion
22.120	God
22.130	church, temple
22.140	altar
22.150	offering, sacrifice
22.160	worship
22.170	pray
22.180	priest
22.190	sacred, holy
22.220	preach
22.230	bless
22.240	curse
22.260	fast (vb)
22.310	heaven
22.320	hell
22.350	demon (evil spirit)
22.370	idol
22.420	magic, witchcraft
22.430	sorcerer, witch
22.440	elf, fairy
22.450	ghost, phantom
22.470	portent,omen

CÁCH TIẾP CẬN DỰA TRÊN CƠ SỞ TRI THỨC CHO VIỆC DỊCH MÁY ANH – VIỆT

(TÓM TẮT)

Đinh Điền

Điểm khó khăn nhất trong việc dịch tự động chính là việc loại bỏ tính nhập nhằng của ngôn ngữ tự nhiên. Bài báo này sẽ trình bày một cách tiếp cận dựa trên cơ sở tri thức để giải quyết sự nhập nhằng nói trên và áp dụng cơ sở tri thức đó vào hệ dịch máy Anh – Việt. Hệ cơ sở tri thức này bao gồm những ý niệm như : *đồ vật, hành động, quan hệ, thuộc tính,...* và được tổ chức trên cơ sở kiến trúc tầng bậc có tính kế thừa.

EVT là một chương trình chạy trên máy vi tính nhằm dịch ngôn ngữ tự nhiên từ tiếng Anh thành tiếng Việt. Đầu tiên, chương trình sẽ nhập các câu hay đoạn văn đúng ngữ pháp tiếng Anh, kế đó nó phân tích chặt chẽ về mặt từ vựng, cú pháp, ngữ nghĩa,.. Nhờ vào những luật dẫn xuất, tự điển nội tại và bộ phân tích ngữ nghĩa nông dựa trên cơ sở tri thức, chương trình sẽ tạo ra câu/đoạn văn tiếng Việt có nghĩa tương đương.

Từ điển điện tử Anh – Việt của hệ gồm khoảng 30.000 từ gốc và 2.000 thành ngữ thông dụng nhất (dựa trên từ điển tần số) cùng với khoảng 1.000 luật phân tích cú pháp và ngữ nghĩa. Tất cả các mục từ này được phân loại theo đặc tính cú pháp (như: từ loại, mẫu câu, vị trí và chức năng trong câu), thuộc tính ý niệm thích hợp (như: *người, vật, thời gian, không gian, nhanh chậm, tốt xấu, tích cực/ tiêu cực*,...và lĩnh vực liên quan).

Những ý niệm này được coi như là tập của những thực thể và có thể phân chia thành các lớp có tầng bậc khác nhau. Những thể hiện của mỗi lớp cùng chia sẻ những đặc tính tương tự vốn được kế thừa từ lớp cao hơn. Ngoài ra, tất cả các thể hiện của cùng một lớp thì có quan hệ logic (như: tương đương, đối lập, phụ thuộc, tác nhân,...) với nhau.