A KNOWLEDGE - BASED APPROACH FOR ENGLISH- VIETNAMESE MACHINE TRANSLATION

DINH DIEN(*)

INTRODUCTION

The most difficult task in machine translation is the elimination of ambiguity in human language. A certain word in English as well as Vietnamese may have different parts of speech and meanings, which depend on its syntactical position in the sentence and the actual context. This paper will present a knowledge-based approach to solve the above-mentioned ambiguity and the application of the knowledge base in the English-Vietnamese Translation (EVT). We have experimented on this approach in the EVT system with the effective solutions to the lexical and structural ambiguities in Noun Compound.

BASIC STRATEGIES OF MACHINE TRANSLATION

Basing on the process in MT, current MT systems may be classified into the following 3 basic strategies[1]:

a. Direct translation

b. Transfer translation

c. Interlingual translation

Recently, there have also been new approaches. One such approach is:

- Knowledge-based machine translation (KBMT). This paper will describe this approach and apply it to our EVT system.

Now, we will review briefly the above-mentioned kinds of MT system and evaluate them.

Direct translation:

The Direct-translation system translates by replacing simply and automatically source language words/phrases with target language words/ phrases. This kind of MT is simple, fast, and involves no syntactic parsing. There is only 1 phase in this MT strategy, that is to convert directly from SL into TL without any intermediate representation. Such a translation system is appropriate for applications where the text has a limited vocabulary and sentence pattern. This system operates fairly well for translating languages which are in the same typology and have 1-1 correspondence in lexicon, syntax,....

(*)College of Natural Sciences, VNU - HCM City

[1] As John Hutchins and Harold L.Somer (1992:106)

Transfer translation:

The Transfer system translates by analyzing the input sentence and then applying linguistic and lexical rules, called *transfer rules* to map grammatical information from Source Language (SL) to Target Language (TL). There are 3 phases in this MT system, they are: analyzing the SL to create the SL intermediate representation (dependent on SL); converting SL representation into TL intermediate representation (dependent on TL); and generating the TL sentence from the TL intermediate representation.

Interlingual translation:

The interlingual system translates by using an intermediate central data - representation notation called *interlingua*. This representation is independent of all languages and is able to represent even the subtlest difference in meanings between every language in that MT system. Constructing an interlingua powerful enough to represent all of the information every language may require, with the appropriate analyzer and generator is an extremely complicated task and so far it has not been accomplished. There are 2 phases in this MT system, they are: analyzing the SL to create the *interlingua*; and generating the TL sentence from this *interlingua*.

Knowledge – based Machine Translation (KBMT):

The main argument of KBMT is: "we need to understand before translating". So, KBMT need a very sophisticated theory of : psychology, neuro-linguistics, socio-linguistics,...To understand a text, we need not only linguistics knowledge but also extra-linguistics knowledge (e.g. knowledge of real world, special fields,..). In this paper, we will make use of this extra-linguistic knowledge to disambiguate word senses in translating automatically English Noun Compound into Vietnamese ones.

KNOWLEDGE-BASED APPROACH FOR ENGLISH-VIETNAMESE TRANSLATION (KB-EVT)

The knowledge-based system is constructed on the hierarchical model consisting of nodes and relations. This model consists of 2 main types: concept and relation.

Concept

In fact, concept is the meaning of a word. These concepts are organized on the hyponym tree where child concepts are able to inherit their parents' relationships, properties and functions. These concepts may be a notion of the real world, e.g.: *biology, animal, botany, human, living, non-living,...* (pls. refer [A.Miller, 1996]).

Example for hyponym: concept "biology" contains the concept "animal", we symbolize *animal→biology,* similarly: *animal→ biology→living,....*

Relation

It expresses the relationship between concepts. There are many kinds of relation. Each relation expresses a proper sense. These relations are uni-direction.

There are main relations as follows: *Has_Attribute* (point to Adjectives), *Has_Function* (point to Verbs), *Has_part* (point to Nouns).

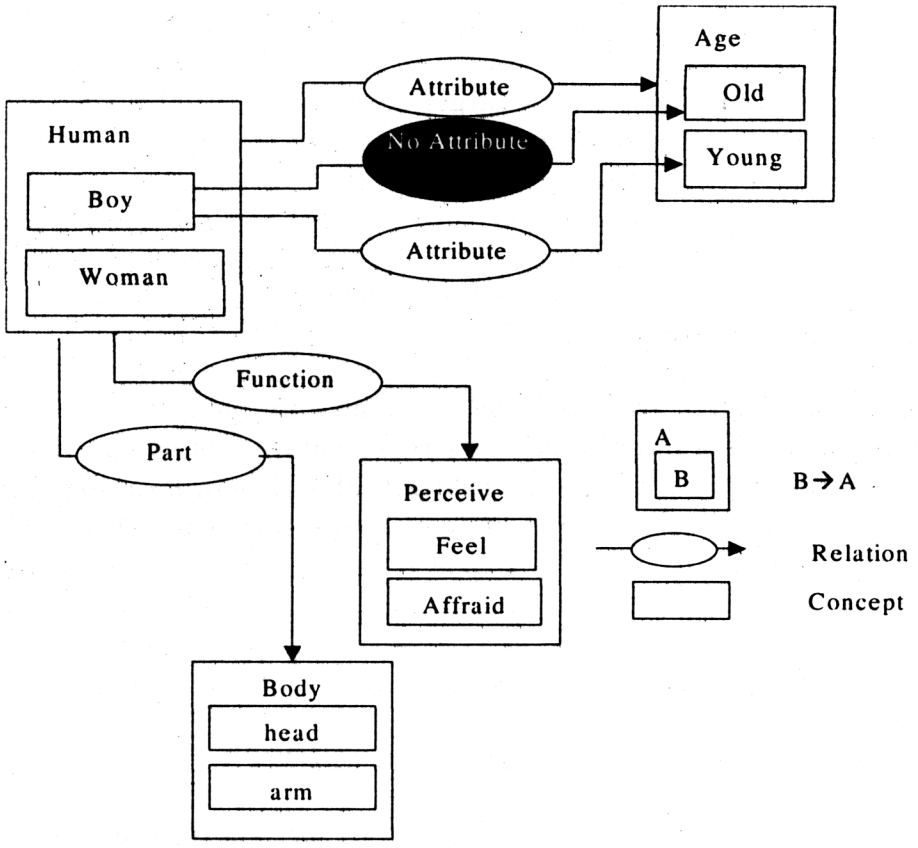

Figure 1: *Semantic network of concepts and relations.*

Example: Examining the above semantic network (figure 1), we see the concepts: *Human, Boy, Woman, Perceive, Feel, Afraid, Age, Old, Young,* in which:

+ Hyponym concepts:

- Boy → Human; Woman → Human; Feel → Perceive; Afraid → Perceive; Old → Age; Young → Age

+ Semantic Network has relationships between concepts as follows:

- The relationship between Human and Age is Has_Attribute
- The relationship between Human and Body is Has_Part
- The relationship between Human and Perceive is Has_Function
- The relationship between Boy and Young is Has_Attribute
- The relationship between Boy and Old is HasNo_Attribute

+And basing on above relationships, we can deduce the following relationships:

- The relationship between Woman and Age is Has_Attribute
- The relationship between Woman and Perceive is Has_Function
- The relationship between Woman and Afraid is Has_Function
- The relationship between Woman and Young is Has_Attribute

MODEL OF KNOWLEDGE-BASED EVT SYSTEM

In this model, we first organize the conceptual dictionary as follows: each word will be listed with all its possible parts-of-speech, each of which will consist of distinguished meanings which will be linked (pointed) to an object – its corresponding semantic frame structure. Then, following the fields in each object, we determine possible semantic relationships between 2 objects and then we choose the most reasonable relationship between them, from this reasonable relationship we are able to choose the correct meaning of a polysemous word.

SEMANTIC FRAME STRUCTURE OF OBJECTS

Firstly, we build objects with its semantic frame structure for each meaning of each word of each part-of-speech:

object of nouns:

Each meaning of a noun is an object with the following structure:

Name of field	**Example 1**	**Example 2**
Name of object	bird	animal
Hypernym	animal	biology
Hyponym	sparrow, dove,...	human, bird, ...
Holonym	flock of..	population, community
Meronym	swing, beak, ...	organs, ...
Function / ability	fly, sing, ...	eat, act, ...
Material attributes	small, light, color...	size, ...
Spiritual attributes	...	...
Properties (optional)	...	...

Table 1: Frame structure of Noun Object

- Hypernym field: indicating to its parent object. Based on this link, it will inherit their parents' attributes and functions (except prohibited cases mentioned in "Properties" field).
- Hyponym field: indicating child object.
- Holonym field: indicating "organization", "class",.. to which this object belongs.
- Meronym field: indicating parts that directly constitute this object.
- Function field: links to verb-network to show actions, abilities which this object is able to perform or traditional purposes of this object.
- Attribute field: linking verb-network to show attributes, materials, and states of this object. These fields are slots like OAV (Object-Attribute-Value) which require contextual words to fill them and are satisfied by its default value, too.
- Property field: showing abnormal or special attributes of this object (non-inherit, non-principle,..).

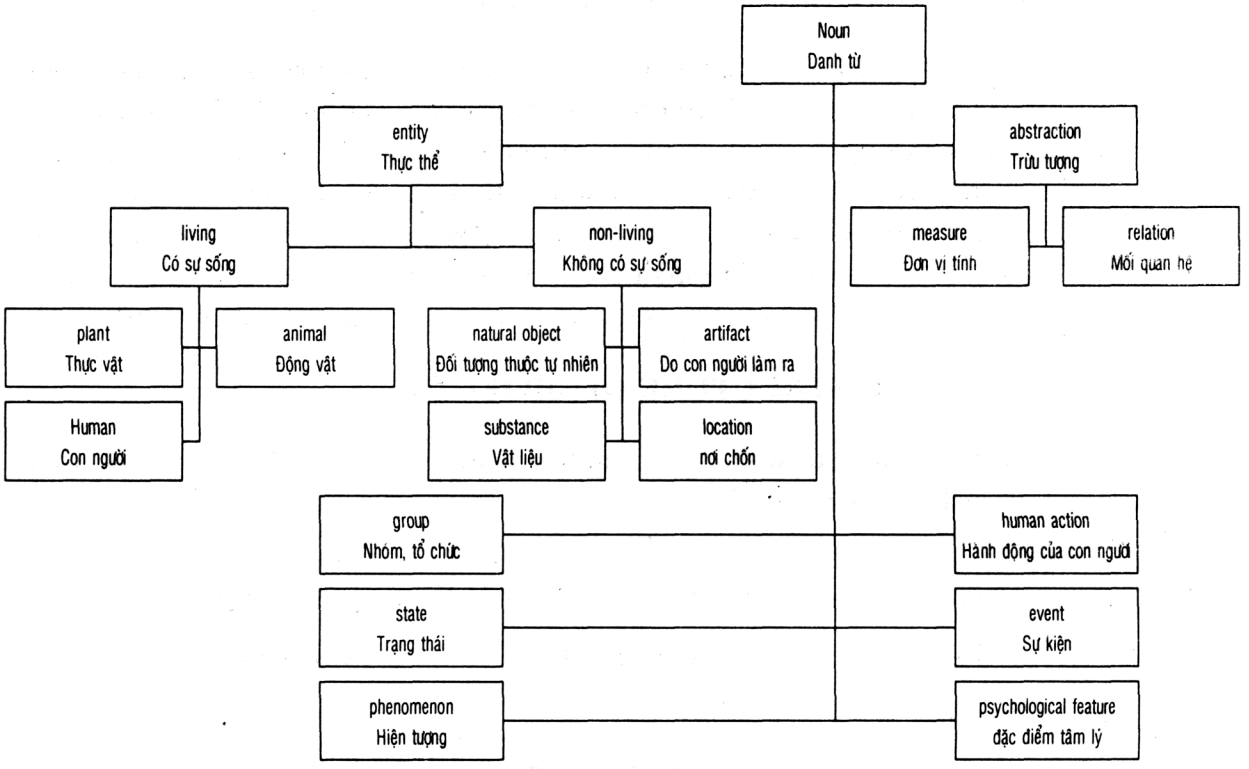

Figure 2: Schema of nouns

OBJECT OF VERBS:

Each meaning of a verb is an object with the following structure:

Name of field	**Example 1**	**Example 2**
Name of action	write	give
Hypernym	make words	abstract motion (possession)
Agentive role	writer	donor
Objective role	writings, works, ...	thing
Instrument role	pen, pencil,	facility
Patient role	to whom	receiver
Benefactive role	for whom	
Location, Spatial role		
Properties (optional)	...	...

Table 2: Frame structure of Verb Object.

Example: "Mr. Ba writes a letter to the teacher for me by a pen"

Agent V Obj Patient Beneficent Instrument

In the relation to other words in a sentence, the verb plays a central role and requires other nouns (subject, object,...) to meet its role (in the semantic aspect) which its semantic frame has specified.

OBJECT OF ADJECTIVES:

Each meaning of an adjective is an object with the following structure:

Name of field	Example 1	Example 2
Name of attribute	Long	long
Parent Class	shape	speed
Synonym cluster	size	time
Antonym cluster	short	fast
Properties (optional)	...	...

Table 3: Frame structure of Adjective Object.

Adjectives are organized into following classes:

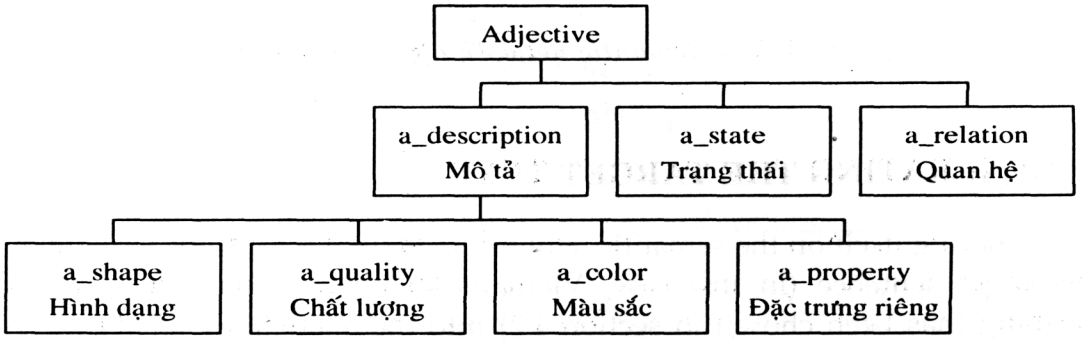

Figure 3. Schema of Adjectives.

SHALLOW SEMANTICS ANALYSIS

Basing itself on the Knowledge Base (in section 4.1), the Shallow Semantics Analyzer will analyze the English string "*A good book on the table*" into a semantic network as in following figure. In this case, the semantics analyzer must be able to choose appropriate meaning for following words:

- Good (adj.): *hay (interesting), tốt (kind), giỏi (talent),...* → choose "hay" because it is concerned with "content".
- On (prep): trên (on the surface), vào (at a time point/ in a duration), về (regarding to),.... → choose "trên" because it is concerned with "position".
- Table (noun): bàn (furniture), bảng (flat board on the wall) → choose "bàn" because it is concerned with "place to contain".

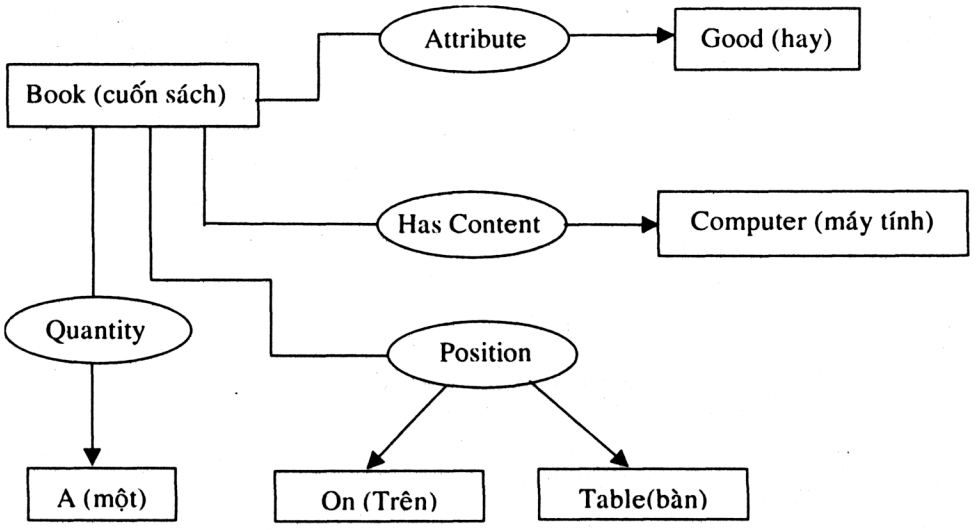

Figure 4: *The semantic network after analyzing.*

GENERATING THE TARGET TEXT

Basing itself on the semantic network in section 4.1, the system will generate the target sentence (in this case, Vietnamese) by assigning the corresponding meaning (has been chosen in section 4.2) into the Vietnamese syntactic tree (this syntactic tree has been converted from the English syntactic tree) and gives the following result:

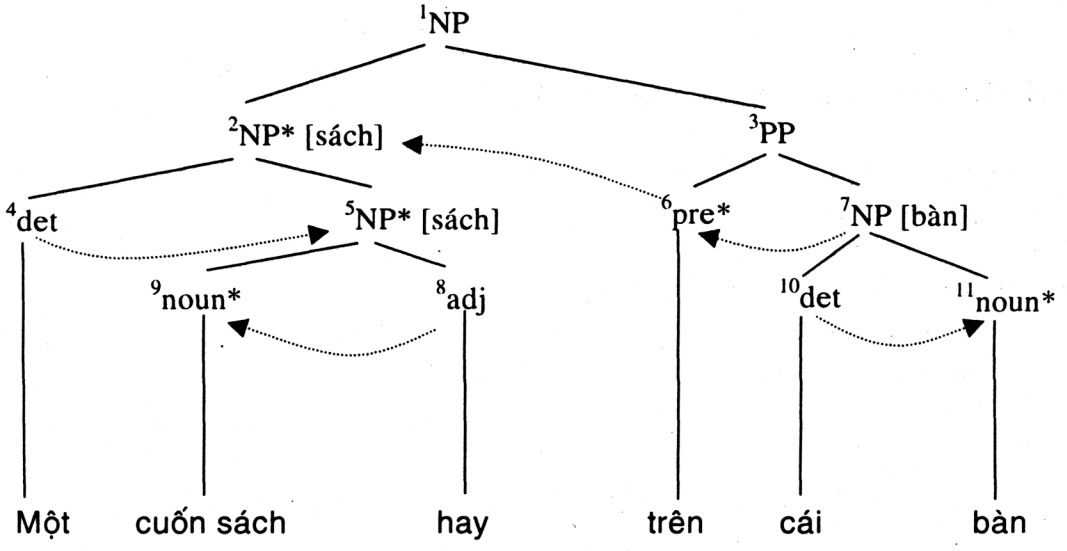

Figure 5: *The output in Target Language.*

CONCLUSION

The above-mentioned solution is just a suggestion and is still being implemented in the English to Vietnamese Translation (EVT). In the initial step, we

experimented with it on the Noun Compound and solved fairly well approximately 65% of ambiguous cases in technical and conventional scientific texts (computer science, electronics, physics,...)

In the future, however, the Knowledge Base System needs to be implemented with new concepts of the real world as well as their relationships. Additionally, this method needs to use the statistical results of the appearance probability of a word meaning in a certain context to help computers to know how to choose the correct meaning, or to help in cases which are not determined by existing knowledge in computer.

REFERENCE

1. Donald, Mac. (1982). *Understanding Noun Compounds*. Carnegie-Mellon Uni, Pittsburgh, PA, USA.
2. Hovy, Eduard. (1993): *How MT works*. Byte, Jan 1993.
3. Hutchins, John and Somer, Harold L. (1992). *An Introduction to Machine translation*. Academic Press.
4. Krulee, Gilbert K. (1991): *Computer Processing of Natural Language*. Prentice-Hall.
5. Lauer, Mark (1995). *Designing SLL: experiments on Noun Compounds*. Macquarie Uni NSW, Australia, 1995.
6. Miller, George A. (1996): *Introduction to WordNet*. http://www.cogsci.princeton.edu/~wn/. Princeton.

THE PHONOLOGY SYSTEM OF JUDU GELAO LANGUAGE

Zhou Guoyan(*)

Introduction

Gelao is one of the 55 ethnic minorities in China with a population of over 430,000 (1990). The Gelao people are mainly distributed in the Guizhou Province of Northern China, a small number of them live scattered in such neighboring areas as Yunnan Province and the Guangxi Zhuang Autonomous Region. In Guizhou, over 90% of the Geolao people inhabit three counties, i.e. Daozhen, Wuchuan, Zhen'an in the northern part, all the rest live in dispersed clusters of communities in about 20 counties in the western and southwestern part. The Gelao people have their own language the Gelao languages which is classified as belonging to a new branch of the Kam-Tai stock in the Sino-Tibetan language family. Western scholars tend to regard it as a language within the Kadai family. According to recent surveys, there are only about 150,000 people who still use the Gelao language as their native tongue, and the dialects of the language differ greatly as those Gelao language speakers live in separate region. That makes it difficult for these people to communicate among themselves in their language. Linguists in China and abroad usually divide the Gelao language into four dialects, naming each of them after the name of the branch, that is, Hagei /ha^{33}kei^{35}/, Gao /qau^{55}/, Doluo /to^{31}ʔlo^{35}/ and A'ou /ʔa^{33}ɣɯ^{55}/ (He, 1983). Some Chinese scholars name the Gelao dialects according to their geographical locations, such as the Central Guizhou dialect, the Central-North Guizhou dialect, the Southwestern Guizhou dialect and the Western Guizhou dialect (Zhang, 1993). As a matter of fact, the Gelao speakers from different dialects usually live in the same area, moreover, some of the Gelao people who live in Yunnan and Guangxi still use their native language, therefore, naming the dialects after their geographical location cannot properly reflect their distribution. In this paper, the present writer adopts the former naming method.

Judu is a Gelao village located in the North-western corner of Liuzhi county in Guizhou Province with over 90 households and nearly 400 people, all of them are Gelao. Near Judu, there are still some villages inhabited by such ethnic groups as Bouyei, Yi (Lolo), Gelao, Han Chinese and Miao (Hmong) respectively. The Judu Gelao have their own native language which is widely used among the villagers. On some occasions, the Gelao people from different villages nearby also use the Gelao language, but most of the time they prefer to use Chinese (the local Chinese dialect). Most of the Judu villagers can understand some simple words and sentences in the Bouyei and the Yi (Lolo) languages which are widely used in the nearby villages, but for longer and more complicated conversation, they have to use Chinese. Actually, Gelao language is only used among the Judu villagers. The Judu villagers call

(*)Linguistics Department, Central University for Nationalities, Beijing, China, 100081

themselves "to^{31}?lo^{35}" in their own language, therefore, the Gelao language here belongs to the "to^{31}?lo^{35}" dialect.

The present writer was in Judu for a few days in the summer of 1999 to investigate the Gelao language there and this article was written on the basis of the full and accurate first-hand materials obtained in the field work.

According to the traditional way of linguistic description, the phonology system of the Judu Gelao language can be described and analyzed according to the following four aspects, i.e. initials, finals, tones and syllable structures.

Initials

An Initial refers to an consonant or a cluster of consonants appearing at the beginning of syllables. In the phonology system of Judu Gelao language, there are altogether 53 initials, including 39 single consonant initials, 6 palatalized consonant initials, 3 labialized consonant initials and 6 consonant clusters. Samples of these initials are as follows:

p	puɯ^{24}	give	pu^{31}	four	po^{31}	cymbals	
ph	phɯ^{31}	a tree	phin^{24}	put together	phɑɯ^{35}	fly	
b	bɑŋ^{33}	rice field	bəŋ^{24}	spring, well	bei^{31}	split	
pj	pjɑu^{33}	float	pji^{31}	pass by	pje^{35}man^{31}	dizzle	
phj	phjo^{35}	wooden dipper	phin^{31}phja^{35}	flat land, plain			
bj	bjɑu^{31}	swallow (a kind of bird)	bja^{35}tin^{31}	plank			
	qə^{0}bji^{35}	a bundle wrapped in cloth					
pl	pla^{24}	blood	plɯ^{31}	alcohol	plaŋ^{31}	tooth	
phl	phlɯ^{24}	silver					
phlɑŋ^{31}qɑi^{31}	the name of young hen before laying eggs						
bl	ble^{31}	duck	bla^{24}	(of plough) damaged			
m	mɑɯ^{31}	dog	məɯ^{31}	tongue	man^{31}	rain	
m̥	m̥e^{35}	flea	m̥o^{31}	arm	m̥an^{31}	look at, see	
mj	mje^{33}	mother	mja^{35}ti^{31}	clogs	mji^{31}	hand	
ml	mlɑŋ^{31}	pond, pool	mlan^{31}	agree	mlɯ^{31}	low	
f	fən^{31}	empty-handed	fan^{31}	cent	fe^{24}	blow	
v	vei^{31}dzuŋ^{31}	emperor	vu^{31}	crazy, mad	vu^{24}	go	
ts	tsɯn^{31}	buy	tsɿ^{33}	one	tsɑŋ^{31}	bite	
tsh	tshe^{35}	generation	tshɯn^{31}n̥an^{31}	last year	qə^{0}tshu^{214}	people	
dz	dzai^{24}	chop, cut	dzuŋ^{24}	tired	dzɑu^{33}	chopsticks	
s	səɯ^{31}	two	so^{31}	shine upon	sɯn^{24}	urine	
z	zuŋ^{24}	cave in	vu^{31}zɯn^{35}	apron	zan^{31}	fill in	
t	ta^{214}	a bamboo basket	tu^{33}	break off	to^{214}	lose	
th	thuŋ^{31}	lead to	tho^{31}	carry on the back	than^{31}	dislike	
d	dɑŋ^{33}	gun	dɯn^{31}	edge	qə^{0}duŋ^{31}	egg	
tj	tje^{24}	liver	tjau^{24}	fish (v.)			

thj	thje^{31}thje^{35}	besiege	thjan^{31}	fill in							
dj	dje^{214}	load	vu^{31}dju^{31}	gallbladder							
n	qə^{0}nan^{31}	bait, rag	nu^{33}	bow	nei^{31}	sea					
n̥	n̥aŋ^{33}	steep, soak	n̥ɯn^{35}	fill up	n̥uŋ^{33}n̥a^{214}	voluntariness					
l	luŋ^{31}	socks	la^{35}	son	lu^{35}	field, plowland					
ɬ	ɬi^{33}	mouse	ɬa^{31}	take	ɬei^{31}ɬei^{31}	flow					
ʔl	ʔlei^{31}	house	ʔle^{24}	wet, damp	ʔluŋ^{31}	thick (soup)					
lj	lji^{24}	cry	qə^{0}lje^{31}	to swallow	vu^{31}ljɯ^{35}	face					
lw	lwan^{31}	kill	lwei^{31}	heaven	lwe^{31}lwe^{31}	rotate					
tɕ	tɕi^{35}	sleep	tɕa^{33}	grasp	tɕiu^{24}	rescue					
tɕh	tɕhiu^{31}	six	tɕhin^{33}n̥o^{31}	invade	tɕhaŋ^{24}	choke					
dʑ	dʑi^{31}dʑi^{31}	block	dʑin^{24}	entrust	dʑaɯ^{31}	believe					
ɕ	ɕe^{31}ɕe^{35}	brush	ɕi^{31}	speak	ɕiu^{214}	rust					
n̥	n̥u^{24}	straw, grass	n̥uŋ^{35}	hollow out	n̥in^{24}	oily					
ŋ̥	ŋ̥uŋ^{35}	ride (a horse)	ŋ̥aŋ^{33}	muddle							
k	ko^{35}	fast, quick	kuŋ^{31}kuŋ^{31}	heavy	ke^{24}	change					
kh	lə^{0}khuŋ^{31}	dustpan	khaŋ^{33}	compensate	khe^{35}	clean					
g	gəɯ^{214}	drought	gi^{33}	good	gei^{31}	ginger					
kj	kje^{24}	untie									
kw	kwan^{24}	be used to	kwe^{31}	country	kwa^{31}	shave					
ŋ	ŋu^{24}	sleep	ŋan^{24}	mortar	ŋau^{24}	dig (with stick)					
ŋ̊	*tin*^{31} *ŋ̊a*^{31}	*pine tree*	*ŋ̊a*^{35}	*maxilla*							
q	qɑ^{33}	no, without	qəɯ^{31}	before	qɑi^{35}	chicken					
qh	qhɯ^{24}	hot pepper	qhɑ^{24}	relative	qhɯn^{24}	(rope) snap					
h	huŋ^{35}	pus	hei^{31}	fume	hɯn^{24}	hate					
ɣ	ɣa^{31}	aux. of mood	ɣo^{35}	no							
ʔ	ʔo^{24}	loach	ʔɑɯ^{35}	sprout	ʔɑŋ^{31}	have, there be					
ʔj	ʔjɯ^{33}	dried up	ʔjai^{31}	lose	ʔjɑu^{214}	suffer from					
j	*juŋ*^{31}	*rattan*	*jin*^{33}	*lunar*	*jɑɯ*^{24}	*grow*					
w	wɑŋ^{42}	net	wei^{24}	stomach	wai^{33}	inclined					

Notes on initials:

1. When pronouncing voiced stops, there is an obvious vibration in the vocal cord, and no pre-glottalized stop before the articulation. The pronunciation of the semi-vowel /j/ is sometimes close to voiced fricative /ʑ/, but not as clear as it.
2. Sometimes the aspirated feature of devoiced nasals is very clear and strikingly contrasted to voiced nasals. But the frequency of all the devoiced nasals is very low except for n̥.
3. The difference between the pronunciation of palatalized initials and that of consonants followed by finals beginning with /i/ is quite clear. According to the existing data we have, the frequency of all the palatalized initials is very low, especially that of such initials as /phj/, /tj/, /thj/, /dj/, /kj/, etc.

4. When pronouncing a consonant cluster, the first consonant is usually pronounced weaker and quickly shift to the second one. This is more obvious when the first consonant in the cluster is voiceless.
5. Nasal /ŋ/ alone can form a syllable.
6. The sound value of dental affricates are close to /tθ/, /tθh/ and /dð/ respectively.
7. There are very few labialized initials, and most of them appear in Modern Chinese loan words.
8. Dorsal affricate /tɕh/ appears in modern Chinese loan words only.
9. There are semi-vowels /j/ and /w/, bilabial semi-vowel /w/ mainly appears in modern Chinese loan words.

Finals

There are altogether 29 finals, including 8 single vowels, 10 diphthongs, 2 triphthongs and 9 finals ended with nasal consonants. Samples of finals are as follows:

a	ɬa^{31}	give	ma^{42}	not	tsɿ^{31}pja^{31}	everyone					
ɑɯ		dʑɑɯ^{31}	love, like		bɑɯ^{214}	pull	pa^{31}ʔɑɯ^{35}	white			
ai		kai^{33}kai^{33}	rake up		qɑi^{31}	chicken	lai^{33}	plough			
ɑu		dʑau^{33}bɑŋ^{31}	cave		tshɑu^{42}	noisy	qə^{0}tsɑu^{42}	make up			
an		qə^{0}nan^{42}	press		ban^{42}	break off with both hands	han^{33}	don't			
ɑŋ		zɑŋ^{33}	call, cry		cɑŋ^{24}	elephant	lɑŋ^{31}dzɑŋ^{31}	stool			
e		he^{31}	eight		qhe^{24}	pot	me^{24}	white gourd			
ei		n̥ei^{35}	underlay		plei^{31}plei^{31}	steep	nei^{24}n̥ei^{33}	short			
əɯ		səɯ^{31}	two		pji^{31}ləɯ^{35}	get angry	qə^{0}təɯ^{35}	fat			
ɯ		pɯ^{24}	obtain		lə^{0}hɯ^{31}	wait	qɯ^{214}	to gore			
əu		zəu^{24}	go mouldy		ma^{42}ləu^{31}dəu^{31}	raft	gəu^{31}	escape			
ɯn		tɯn^{24}	to drop		n̥ɯn^{35}	push up, prop up	qhɯn^{24}	break			
ɯŋ/əŋ		bəŋ^{31}	satin		hɯŋ^{33}	river	ʔɯŋ^{31}dʑi^{31}	moon			
i		gi^{33}	good		li^{33}li^{33}	stroke	n̥i^{31}	cattle			
ɿ		tsɿ^{42}	one	sɿ^{31}	real	dzɿ^{35}	pull out a knife				
ia		ɬia^{31}me^{33}me^{33}	sheep		tɕia^{31}	the first of Tian Gan					
iɑu	ɬiɑu^{24}	follow, and	ɕiɑu^{42}man^{42}	one of the twenty^{n}four climate our climate phenomena in a year (a loan word from Chinese)							
ian		n̥ian^{214}	read aloud		tɕa^{31}tɕhian^{31}	forceps	tɕian^{24}	arrow			
iu		ʔiu^{24}	bamboo		n̥iu^{42}tsɿ^{42}	claspers	ɕiu^{24}	rust			
in		qə^{0}tin^{24}	remember		n̥in^{214}	oily	ɕin^{31}	pipe, tube			
o		m̥o^{31}	pig		tsho^{31}	be able to	tso^{35}	replace			
u		vu^{31}	mad		fu^{31}	buddha	ju^{31}	high			
ua		ho^{42}hua^{33}	sparkle		tsua^{31}mu^{31}kuan^{33}	wood pecker					
	khua^{33}	boast									

uai	khuai^{214}	hurry	tsuai^{33}	doze	tɕhi^{33}kuai^{24}	strange			
uan	kuan^{24}	thresh	tsuan^{24}	pore	tshuan^{42}	dig with a spade			
uɑŋ	huɑŋ^{33}	hurry	tsuɑŋ^{24}	bump, touch		suɑŋ^{33}	couple		
ui	wa^{42}hui^{35}	grey	tsui^{24}	most, adj+est	tɕian^{42}sui^{42}	buck			
uŋ	ɳuŋ^{35}	hold in both hands	duŋ^{24}	salty, salted	dzuŋ^{24}	tired			
y	jy^{31}tsha^{33}	fish fork	tɑu^{24}ɕy^{33}	agnail					

Notes on finals:

1. When used as a single vowel or followed by final /i/ and /n/, the sound value of /a/ is front /a/, written as /a/, but when appears as after /q/, /qh/ or followed by final /u/, /ɯ/, /ŋ/, it is a back /a/, written as /ɑ/.
2. /ɿ/ is always a single vowel, and appears after dental fricatives and dental affricates only, while /i/ never appears in this position.
3. /ɯ/ and /ə/ free variants of a phoneme. Generally speaking, when used as a single vowel, it is a front vowel which can be written as /ɯ/, but when followed by back vowel /u/ and /ɯ/, it is pronounced as central vowel /ə/, in other cases, sometimes it is pronounced as /ɯ/, sometimes as /ə/.
4. Finals begining with /i/ and /u/ usually appear in Modern Chinese loan words.

Tones

There are altogether six tones. Samples of tones are as follows.

1. middle level (33): suɑŋ^{33} couple tsuai^{33} doze zɑŋ^{33} cry, call
2. middle falling (31): vu^{31} mad fu^{31} Buddha ju^{31} high
3. low rising (24): duŋ^{24} salty, salted dzuŋ^{24} tired ʔiu^{24} bamboo
4. low falling-rising (214): khuai^{214} hurry ɳian^{24} read aloud qɯ^{214} butt, gore
5. middle rising (35): ɳuŋ^{35} hold with both hands ɳei^{35} fill up tso^{35} replace
6. high falling (42): tsɿ^{42} one ban^{42} break off with both hands khɑu^{42} stew tshɑu^{42} noisy

Notes on tones:

1. The middle level tone is usually as high as 44.
2. Sometimes the low falling-rising tone and the high rising tone can be mixed. For some words, it is usually difficult for both the informant and the investigator to determine whether it is 214 or 35. But in most cases, the tone value between these two can be clearly distinguished.
3. The tone of some noun prefixes are usually pronounced so weak that it is difficult to determine their tone values, therefore we can record them as zero (0), e.g.

vu^{0}no^{35}	bird	pə^{0}qu^{24}	skin
lə^{0}zo^{31}lə^{0}ze^{33}	speckle	qə^{0}tɕu^{31}	rimple.

4. There are no tone categories for modern Chinese loan words, the tones of modern Chinese loan words are classified under such tones as the mid-level (33) , the low-falling (31), the high-falling (42), and the low-rising (24) respectively.

Syllable Structures

A syllable in Judu Gelao language is usually formed by an initial consonant + a final + a tone. If we analyze initials and finals according to their exact sound value, that is, to divided them into consonants and vowels, then, the syllables in Judu Gelao can be divided into the following types (C = consonant, V = vowel, T = tone)

1. C+T This type of syllable is formed by a single consonant plus a tone. In the Judu Gelao language, only nasal /ŋ/ can form a C+T type syllable, e.g.

   $ŋ^{33}$ water

2. C+V+T This type of syllable is formed by a single consonant followed by a single vowel and plus a tone. In Judu Gelao language, almost all single consonants can combine with a single vowel, and the syllables formed by a single consonant plus a single vowel can be followed by any tones, therefore, in Judu Gelao language, this type of syllable makes up a large portion, e.g.

tso^{35}	replace	gi^{35}	good
$qɯ^{214}$	to gore, to butt	ju^{31}	high

3. C+V+V+T This type of syllable is formed by a single consonant followed by a diphthong and plus a tone. In Judu Gelao language, a diphthong makes up one third of all the finals, therefore, this type of syllable also makes up a large portion, e.g.

   $səɯ^{31}$ two $n̥ei^{35}$ prop up $dʑɑɯ^{31}$ like $khɑu^{42}$ stew

4. C+V+V+V+T This type of syllable is formed by a single consonant followed by a triphthong and then plus a tone. In Judu Gelao language, there are only two triphthongs, therefore, this type of syllable is very few and most of them are modern Chinese loan words, e.g.

   $ɬiɑu^{24}$ and, with khuai^{214} fast, quick $tsuai^{33}$ doze

5. C+V+C+T This type of syllable is formed by a single consonant followed by a final ended with a nasal (/n/ or /ŋ/) and then plus a tone. The Judu Gelao language has altogether 9 finals ending with nasal consonant, which make up one third of all the finals, therefore, this type of syllable also makes up a large portion, e.g.

   $tsɑŋ^{33}$ cry, call ban^{42} break off with both hands

   $ɳuŋ^{35}$ hold with both hands $tsɯn^{24}$ to drop

6. C+V+V+C+T This type of syllable is formed by a single consonant followed by a diphthong plus a nasal consonant and then plus a tone. According to the data we have, most of this type of syllable are modern Chinese loan words, e.g.

   $ɳian^{214}$ read aloud $suɑŋ^{33}$ couple $huɑŋ^{33}$ hurry $tɕian^{24}$ arrow

7. C+C+V+T This type of syllable is formed by a consonant cluster followed by a single vowel plus a tone. In Judu Gelao language, there are only a few consonant

clusters, and the single vowels which can combine with consonant clusters are limited, therefore, the number of this type of syllable is very small, e.g.

pla^{24} blood ble^{31} duck plɯ^{31} wine mlɯ^{31} low

8. C+C+V+V+T This type of syllable is formed by a consonant cluster followed by a diphthong and then plus a tone. According to the data we have, in Judu Gelao language, consonant /pl/ appears before a diphthongs, and only the diphthongs /ai/, /ei/ can follow a consonant cluster, therefore, very few samples can be found for this type of syllable, e.g.

plai^{214} carry on shoulder plai^{42}plai^{42} ask plei^{31} year

9. C+C+V+C+T This type of syllable is formed by a consonant cluster followed by a final ended with a nasal consonant (/n/ or /ŋ/) and then plus a tone. According to our data, in Judu Gelao language, only consonant clusters /pl/ and /ml/ can be followed by this kind of final, and only finals /an/ and /ɑŋ/ appear after these two consonant clusters, therefore, samples of this type of syllable are very few, e.g.

plan^{33} burn plɑŋ^{33} bamboo mat mlɑŋ^{31} oil

Features of the phonology system of Judu Gelao language

Amongs all of the Gelao dialects, the phonology system of Judu Gelao language is quite complicated, The main features of this dialectal point can be summarised as follows:

1. The initial system is quite integral in sound contrast. Except uvula which have only voiceless sound, each group of stop has a voiceless contrasting to a voiced sound, and all the fricatives and affricates have a voiceless vs. voiced correspondence:

   Stops: p → b, pl → bl, t → d, k → g, q → (no voiced sound corresponding to it)

   Fricatives: f → v, s → z, ɕ → ʑ (in the consonant chart, we regard this sound as a semi-vowel, but we can also regard it as a voiced fricative corresponding to voiceless /ɕ/), h → ɣ

   Affricates: ʦ → dz, ʧ → dʒ

   Nasals: m → m̥, n → n̥, ɲ → ɲ̊, ŋ → ŋ̊

2. There are palatalized initial consonants. Palatalization is very common in Judu Gelao language, and appears in labial, dental, velar and lateral consonants, except for /lj/, /bj/, /pj/, /dj/, all the palatalized aspirated stops appear in modern Chinese loan words. According to the general regularity of phonetic development, palatalization is derived from consonant cluster, but with the data we have in hand, it is difficult to determine that the palatalization in Judu Gelao has anything to do with consonant cluster.
3. There are two labialized initial consonants, i.e. /lw/ and /kw/, /lw/ which appear in native words, while /kw/ appears in modern Chinese loan words.
4. There are aspirated vs. unaspirated correspondence among stops (except /ʔ/) and affricates.

   Stops : p → ph, t → th, k → kh, q → qh;

Affricates : ts → tsh, tɕ → tɕh

Aspirated dorsal affricate /tɕh/ mainly appears in modern Chinese loan words.

5. There are consonants clusters, but most of them are labial consonant followed by a lateral (except /ʔl/) i.e. pl, phl, bl and ml. Velar and dental consonant clusters have already disappeared.
6. There are pre-glottal consonants /ʔl/ and /ʔj/, /ʔj/ can also be described as /ʔʑ/. This group of consonants in Judu Gelao ($to^{31}ʔlo^{33}$ dialect) are not as numerous as those in $ʔa^{33}ʔɯ^{33}$ dialect.
7. There is no pre-nasalized consonant. Two dialects, qau^{55} and $ʔa^{33}ɣɯ^{55}$, in Gelao language have this group of consonants, while in $ha^{35}kei^{55}$ and $to^{31}ʔlo^{33}$, none of have been found. It is still difficult to say whether pre-nasalized consonants were the intrinsic elements in the sound system of the proto language. As pre-nasalized consonants are distributed in the areas where the Gelao and the Miao (Hmong) live in compact communities, and the language of Miao (Hmong) people (being rich in pre-nasalized consonants) can almost serve as media language among different ethnic groups, the pre-nasalized consonants in qau^{55} and $ʔa^{33}ɣɯ^{55}$ dialects of Gelao may be regarded as the result of linguistic influence and culture contact. But in $ha^{33}ke^{35}$ and $to^{31}ʔlo^{33}$ dialects, we can also find that there are some words in which the initial consonant are correspondent to pre-nasalized consonant in qau^{55} and $ʔa^{33}ɣɯ^{55}$, e.g.

Chinese	PG	JG	ZG	LG
food	$mpə^{21}$	$mɑɯ^{33}$	$mɯŋ^{42}$	$muŋ^{53}$
pig	$mpɑ^{33}$	$m̥o^{31}$	mau^{35}	$mɑu^{55}$
cattle	$ntɛ^{55}$	$n̥i^{31}$	$n̥e^{42}$	$n̥ə^{31}$
bird	$ntuŋ^{21}$	no^{35}	$mo^{42}naŋ^{42}$	ne^{13}
snake	$ŋkau^{55}$	$ŋɯ^{31}$	$ŋo^{42}$	$ŋo^{55}$
flower	$ŋkau^{33}$	---	$kəu^{35}$	$ŋ^{33}$
salt	$n̥tɕɯ^{33}$	$n̥ɯ^{53}$	$n̥ɯ^{31}$	$n̥ɯ^{31}$

The samples above show that, in Gelao pre-nasalized consonants may have some genetic relationship with other nasals and stops, but now we still do not have enough data to demonstrate which group is the earlier form. It is interesting to find that, in the two dialects which have pre-nasalized consonants, one can seldom find words with pre-nasalized consonants in one dialect corresponding to the same group of consonants in another.

Chinese	PG	DG	**Chinese**	PG	DG
food	$mpə^{21}$	mo^{55}	**snake**	$ŋkau^{55}$	$ŋɯ^{31}$
cattle	$ntɛ^{55}$	$ŋau^{31}$	**salt**	$n̥tɕɯ^{33}$	$n̥au^{31}$
stool	thi^{44}	$ŋku^{31}$	**acid**	$vlɒ^{31}$	$mpiɑ^{33}$
flat	tu^{13}	nto^{55}	**coat**	$luŋ^{44}$	$ntsɯ^{53}$

(**Notes**: PG = Pingba Gelao [qau^{55} dialect]; JG = Judu Gelao [$to^{31}ʔlo^{33}$ dialect]; ZG = Zhenfeng Gelao and LG = Liangshuiying Gelao [$ha^{33}ke^{35}$ dialect]; DG = Dafang Gelao [$ʔa^{33}ɣɯ^{55}$ dialect]).

8. The final system is quite simple. Actually, there are 7 single vowel phonemes although we listed 8 finals in the chart, as /i/ and /ɿ/ are variants of one phoneme. Among the seven single vowels, /y/ appears in modern Chinese loan words only. All single vowels have no length contrast. Final vowels are /i/ and /u/, final nasals are /n/ and /ŋ/, there is no final stops.
9. There are six tones, more than the other three dialects. Pingba Gelao (qau^{55} dialect): 5 tones; Qinglong Gelao (ha^{33}ke^{35} dialect): 5 tones; Dafang Gelao (ʔa^{33}ʔɯ^{35} dialect): 4 tones.

References

1. Zhang Jimin, 1993, ***Studies on Gelao Language,*** Guizhou National publishing house.
2. He Jiashan, 1983, ***A Brief Introduction of Gelao language***, National publishing house.
3. Li Jinfang & Zhou Guoyan, 1999, ***A Study on Ge-Yang Languages***, Publishing House of the Central University for Nationalities.
4. Ni Dabai, 1990, ***An Introduction to Kam-Tai Languages***, Publishing House of the Central University for Nationalities.

HỆ THỐNG THAM CHIẾU ĐẠI TỪ TRONG TIẾNG DAAI- CÁC TẬP HỆ ĐẠI TỪ LUÂN PHIÊN VÀ TIÊU ĐIỂM

(TÓM TẮT)

Helga Hartmann

Daai là một ngôn ngữ Nam Chin được sử dụng trong các công đồng dân cư ở Kanpetlet, Mindat, Matupi và Paletwa ở phía Nam vùng đồi Chin thuộc Mianmar. Trong bài nghiên cứu này chúng tôi cố gắng hoàn thiện việc miêu tả sự tương tác giữa các đại từ trong phạm vi các danh ngữ hành chức ở cấp độ tiểu cú và các yếu tố được xem là đại từ của thành tố phù ứng tham chiếu trực thuộc động ngữ.

1. Tiểu cú nội động (Intransitive clause)

Ngữ đoạn phù ứng tham chiếu là một thành tố tương đối cần yếu, thường đứng trước động từ trong một tiểu cú trong tiếng Daai Chin.

2. Tiểu cú ngoại động (Transitive clause)

Việc đánh dấu chủ ngữ qua sự phù ứng tham chiếu trong tiểu cú ngoại động khá giống với việc đánh dấu chủ ngữ trong tiểu cú nội động. Hệ thống phù ứng tham chiếu bổ ngữ đồng nhất hơn so với hệ thống phù ứng tham chiếu chủ ngữ. Chẳng hạn như, không có sự phân biệt giữa các hình thức số đơn và hình thức số nhiều, cũng như không có sự khác biệt giữa hình thức số ít và số nhiều ở ngôi thứ nhất.

3. Đại từ sở hữu

Trong các danh ngữ có chứa đại từ sở hữu thì người ta thấy có hai tập hệ đại từ khác nhau: tập hệ đại từ tự do và tập hệ đại từ tham chiếu - được phân tích như một thành tố của động ngữ, cả hai được dùng như danh ngữ sở chủ ở cấp độ danh ngữ và tùy thuộc vào việc điểm nhấn rơi vào tâm của danh ngữ sở chủ hay tâm của danh ngữ biểu thị sở vật.

4. Tập hệ đại từ luân phiên và tâm điểm trong tiểu cú

Tâm điểm ở cấp độ tiểu cú cũng có thể được sử dụng thông qua việc chọn lựa tập hệ đại từ. Việc sử dụng đại từ tự do trong quan hệ kết hợp với đại từ phù ứng tham chiếu có thể được dùng để nhấn mạnh tham tố.