# Preliminaries for Digitizing the Subjective Personal Pronouns in English into Tamil (Distribution -Sensitive Machine Translation Aid -DSMTA) - A Demo Paper

# Dr. S. Kamakshi

Dept. of Tamil Language, University of Madras, Chennai

# **Introduction:**

The aim of this paper is to prepare a Distribution - Sensitive Machine Translation Aid (DSMTA) to translate the Subjective Personal Pronouns (SPP) in English into Tamil.

Linguistic Tools mentioned as **Preliminaries** in the title such as Micro Planner (MP), Modules etc, are being supported by various arbitrarily reordered dictionaries. They are named as ARDSPP, ARDBV, ARDMV, ARDLV and GR. These have been carefully compiled for this Machine Translation purpose. The tools and the relevant information are necessary to digitize the theoretical strings into numerical strings.

## Transfer Mechanisms adopted for this task:

1. The following Physical Structured input sentences in English have rudimentarily been taken on par with the distribution of Subjective Personal Pronouns:

Structure 1.

SPP. + 'be' auxiliaries + V4\*

e.g. She is paying.

Structure 2.

SPP. + 'be' auxiliaries + V3\*

e.g. She is paid.

Structure 3.

SPP + Modal +V1\*

- e.g a) She can pay. ('can /able to' type sentences in English muTi / iyalum type sentences in Tamil)
- b) She will pay (other than `can /able to' type sentences i.e. other modal types in English)
- 2. Micro Planner (MP) a Linguistic Module has been developed to parse the input sentences in Source Language, which is being assisted by six sub modules developed and named as:
- i) ARDSPP- Arbitrarily Reordered Dictionary of Subjective Personal Pronouns in English and Tamil (see appendix No 1.)
- ii) ARDBV Arbitrarily Reordered Dictionary of 'Be' auxiliaries in English and its equivalents in Tamil. (see appendix No.2)
- iii) ARDMV- Arbitrarily Reordered Dictionary of 'Modal' auxiliaries in English and its Equivalents in Tamil (see appendix No.3)

- iv) ARDLV Arbitrarily Reordered Dictionary of Lexical Verbs\* in English and its inflected forms (V1\*, V1s\*, V2\*, V3\*, V4\*) with Tamil meaning. ( see appendix No 4.)
- v) GR-Grammar rules given in this program have been prepared specifically for both English and Tamil, as English Personal Pronouns are less in number compared to Tamil.

For instance the Subjective Personal Pronoun (SPP) 'He' has twenty four different translation equivalents in Tamil (see appendix 1), as per its distribution with other grammatical units.

e.g. a). He is giving (his computer to X).

avan / ivan koTuttukkoNTu irukkiRaan/irukkinRaan.
avar / ivar koTuttukkoNTu / irukkiRaar/irukkinRaar
avarkaL / ivarkaL koTuttukkoNTu irukkiRaarkaL/irukkinRaarkaL.
Here `He' =` avan / ivan / avar / ivar / avarkaL / ivarkaL'

-----

V1\* – Lexical Verb or Root verb : V2\* – Past tense verb : V3\* – Past Participle Verb :

V4\* – Present Participle verb.

\* About thousand five hundred Lexical Verbs in English have been analysed into Intransitive, Transitive and Di-transitive verbs.

`koTu' (give) type verbs which are di – transitive, take two objects.

.....

b). He is given (a Computer).

avanukku / ivanukku koTukkap paTukiRatu/paTukinRatu. avarukku / ivarukku koTukkap paTukiRatu/ paTukinRatu. avarkaLuku / ivarkaLukku koTukkap paTukiRatu / paTukinRatu.

Here `He' = `avanukku / ivanukku / avarukku / ivarukku / avarkaLukku / ivarkaLukku '

c). He is known to me

avanai/ivanai enakkut teriyum avarai/ivarai enakkut teriyum avarkaLai/ivarkaLai enakkut teriyum

Here 'He' = avanai / ivanai /avarai /ivarai /avarkaLai /ivarkaLai

d). He can/could give (him a computer) /He is able to give (him a computer) avanaal / ivanaal koTukka iyalum / muTiyum.

avaraal/ ivaraal koTukka iyalum / muTiyum.

avarkaLaal / ivarkaLaal koTukka iyalum / muTiyum.

Here `He'=` avanaal / ivanaal/ avaraal/ivaraal/avarkaLaal/ivarkaLaal

vi) MP- Morphological Parser for Tamil is also incorporated to achieve this aim of Translation of Subjective Personal Pronouns at simple sentence level.

# Language tools to be provided to equip the Modules

English Subjective Personal Pronouns are seven in number whereas Tamil is populated in nature. So the knowledge about these optional entries (e.g. 'He' takes twenty four options-(see appendix 1)) is being given to the computer according to the distribution of the same.

For example, in the sentence 'She is reigning England' the inference that the Subjective Personal Pronoun 'She' is a superior female demands the selection of avar / ivar or avarkaL / ivarkaL instead of avaL / ivaL.

The argument structure of English verbs and their equivalents in Tamil have to be provided.

Collection of English sentences with modal auxiliary verbs and their counterparts in Tamil has been given completely. (see appendix 3)

Objective, Possessive, and Reflexive Pronouns and their distribution are also being analyzed as the present author aims at the Machine translation of Said - Pronouns in English into Tamil.

# Intricacies involved in translating the personal pronouns in English into Tamil

There are some inescapable solutions in providing language tools, as this program is highly linguistic tool dependent. They are:

1. Proximity and Distance:

This cannot be differentiated through Distribution Sensitive Machine Translation.

2. Honorific and Non-honorific sense:

The personal pronouns inferring honorific or non-honorific sense in Tamil can possibly be translated at the Post - Editing Stage of Machine translation through human interference.

3. The Transitive and Intransitive Verbs in English:

Verbs acting as both Transitive and Intransitive are translated into Tamil with semantic ambiguity, and this also can possibly be rectified at the Post - Editing Stage of Machine Translation through human assistance.

#### **Conclusion:**

Finding equivalents in the socio-cultural context is really a challenge for the computational linguist. The problem will multiply if we take into account the anaphoric reference into consideration.

## Appendix 1:

**ARDSPP** (Arbitrarily Reordered Dictionary of Subjective Personal Pronouns)

e.g He - 3<sup>rd</sup> Person SPP in English has the following Tamil forms: avan,ivan,avar,ivar,avarkaL.ivarkaL. avanukku,ivanukku, avarukku,ivaruku,avarkaLukku,ivarkaLukku. avanaal,ivanaal,avaraal,ivaraal,avarkaLaal,ivarkaLaal. avanai,ivanai,avarai,ivarai,avarkaLai,ivarkaLai.

# Appendix 2:

**ARDBV** – A sample of Arbitrarily Reordered Dictionary of 'Be' auxiliaries in English and their equivalents in Tamil. (here present participle marker is -ing)

### Appendix 3:

**ARDMV-** Arbitrarily Reordered Dictionary of 'Modal' auxiliaries in English and their Equivalents in Tamil (appendix No.3)

The auxiliaries such as aam, aakaatu, kuuTu, TTum, maaTTu, muTi, and veeNTu express mood in Tamil.

# Transfer of complex verbal forms denoting Tense, Aspect and Mood (TAM).

Both English and Tamil employ the complex process of combining both inflection and compounding in denoting Tense, Aspect, and Mood (TAM). We can find correspondences between English and Tamil for the purpose of translating one from the other. However, the correspondences are not always perfect. There are proper, improper and defective equivalents. The Tense, Aspect, and Mood systems of English and Tamil operate differently and finding equivalents is a tough task. But for the purpose of MT, we compromise with certain peripheral differences between them and try to capture the core of their systems with the view in mind that what is conveyed in English can be transferred to Tamil without much distortions as our idea is to translate linguistic text in English into Tamil. The emotive and attitudinal senses conveyed by the auxiliary system will not play a vital part in expressing linguistic concepts. So we ignore the emotive and attitudinal sense and try to capture the core aspectual and modal system. That is why we have ignored certain auxiliaries, which are used in Tamil to denote certain attitudinal and non-attitudinal senses. With this aim in mind, the aspectual and modal systems in both languages have been correlated for the purpose of preparing MTA (Machine Translation Aid). The following table correlates TAM (Tense, Aspect, Mood) system of English with that of Tamil.

TAM forms in Figure	Magning	Carrieralant TAM farmes in Tamil with	
TAM forms in English	Meaning	Equivalent TAM forms in Tamil with	
with examples		examples	
Sub+V + past tense	Past tense	Sub+V + past tense + PNG	
e.g. He wrote.	i dot telloc	e.g. avan ezhutinaan.	
Sub+V+present tense	Present Tense	Sub+V + present tense + PNG	
e.g.He writes.	i resent rense	e.g. avan ezhutukiraan.	
Sub+has / have + V3	Present perfect aspect	Sub+V + past participle +ru + present +	
	i recent perieut depect	PNG	
e.g. He has written.		e.g. avan ezhutiyirukkiRaan	
I have written.		ndaan ezhutiyirukkiReen.	
Sub+had + V3	Past perfect aspect	Sub+V + past participle +iru + past +	
		PNG	
e.g. He had written.		e.g. avan ezhutiyirundtaan.	
	Present progressive	Sub+V + past participle +koNTiru +	
tense + V-ing	aspect	present + PNG	
	•	e.g avan ezhutik koNTiukkiRaan	
e.g. He is writing.			
Sub+'Be' verb + past tense	Past progressive aspect	Sub+V + past participle *koNTiru+ past	
+ V-ing	. • .	+ PNG	
eg. He was writing.		e.g. avan ezhutik koNTirundtaan	
Sub+will be/shall be verb	Future progressive	Sub+V + past participle koNTiru	
+future tense + V-ing	aspect	+ future + PNG	
e.g. He will be writing a	1	e.g. avan kaTitam ezhutik koNTiruppaan	
letter.			
Sub+ can + V1	Ability	Sub+V + infinitive +muTiyum/ iyalum	
	= be able to	e.g. avanaal aangkilam peeca muTiyum,	
eg. He can speak English		aanaal ezhuta muTiyaatu.	
but he can't write it ver	y = know how to		
well.			
Sub+ Can + V1	I Permission	1.Sub+ V-al + aam + aa	
e.g.Can I smoke here ('Am		ndaan pukai piTikkalaamaa?	
I allowed to smoke here?).	= be permitted to		
	('can' is less formal that	1	
	'may' in this sense)		
4.0 h	I.B 9.99	40 1. 7.1.	
1. Sub+can + V1	I.Possibility	1.Sub+ V-al + aam	
e.g. He can make	= it is possible but / to	e.g. avan tavaRu ceyyalaam	
mistakes.	theoretical possibility		
0. Out 1 and 1 to 1 1/0	may = factual possibility	O Code IV i infinition to Tally and	
2. Sub+can + be + V3		2. Sub+V + infinitive +paTal + aam	
e.g.The road can be		caalai aTaikkappaTalaam.	
blocked.			
Sub+Could + 1/4	Λ hilifu	Sub-W + infinitive tour Tirered include	
Sub+Could + V1	Ability	Sub+V + infinitive +muTiyum/ iyalum	
e.g.		e.g. enaal caturangkam aaTa muTiyum.	
I could play the chess. Sub+Could + V1	II Dormicsion	Sub+Val + 22m + 22	
Sub-Could + VI	II. Permission	Sub+V-al + aam + aa	

TAM forms in English with examples	Meaning	Equivalent TAM forms in Tamil with examples
e.g.Could I smoke here?		e.g. ndaan pukai piTikkalaamaa?
1.Sub+ could be + C eg. That could be my train.	II. Possibility (theoretical or factual, cfmigh)	1.Sub+irukkkal+ aam e.g. atu ennuTaiya toTarvaNTiyaay irukkalaam.
2.Sub+ could be + V3 e.g.The road could be blocked.	"	2a.Sub+V+infinitive †paTTu + irukkal+ aam e.g caalai aTaikkap paTTu irukkalaam
		2b.Sub+ V + infinitive †paTTu + irukka+ kuuTum e.g.caalai aTaikkap paTTirukkakkuuTum
Sub+may + V1 e.g.He might leave tomorrow.	Future time with moda auxiliaries. In many contexts, modal auxiliaries have inheren future reference, both in their present and pastense form.	e.g.avan ndaalai pookalaam t
1. Sub+may + V1 e.g.You may borrow my ca if you like.	III. Permission r = be allowed to = be permitted to	1.Sub+V-al + aam e.g. ndii ennuTaiya kaarai kaTanaakap peRalaam.
2. Sub+may not + V1 e.g. You may not borrow my car. (=You are not allowed to borrow my car.)	mayn't, mustn'is often	2 Sub+.V + infinitive +maaTTu+ PNG e.g ndii ennuTaiya kaarai kaTanaakap peRamaaTTaay.
1.Sub+ may + V1 e.g.He may never succeed ('It is possible that he wi never succeed')	III. Possibility = it is possible that /to	1a.Sub+V-al + aam (for positive meaning) 1b.Sub+ V + infinitive maaTTu + PNG e.g. avan veRRi peRamaaTTaan
2. Sub+may + be + V3 e.g. The road may be blocked 'It is possible tha the road may be blocked'.		2 Sub+V + infinitive +paTal + aam e.g. caalai aTaikkappaTalaam
Sub+might + V1 e.g.He might leave tomorrow.	Future time with moda auxiliaries. In many contexts, modal auxiliaries have inheren	e.g avan ndaaLai pookalaam

TAM forms in English with examples	Meaning	Equivalent TAM forms in Tamil with examples		
	future reference, both in	1		
	their present and pas			
	tense form.	•		
Might +sub+V1	IV. Permission	Sub+V-al + aam + aa		
e.gmight I smoke here?		e.g ndaan pukai piTikkalaamaa?		
Sub+might + V1	IV. Possibility	Sub+V-al + aam		
e.g.He might succeed.	TVI I Goodbanky	e.g. avan veRRi peRalaam.		
Sub+Shall + V1	II. Willingness on the par	•		
e.g1.He shall get the	•	e.g.1.avan paNam peRalaam.		
money.	person and 3 <sup>d</sup>	o.g partam portama		
2.You shall do exactly as	•	2.ndii virumpuvatu pool ceyyalaam.		
you wish.	restricted use.			
1a.Sub +Shall + V1		f 1a. Sub+V + future tense + PNG		
		e.g. ndaangkaL ungkaLiTam engkaL		
our decision.		tiirmaanattait terivippoom		
We shall overcome.	volition').			
	,			
1b. Sub+shan't + V1		1b. aakaatu		
e.g. It shan't be long for me		e.g. enakku mandtiriyai paarkka ndiiNTa		
to meet the minister.		ndeeram aakaatu.		
Sub+Shall + V1	I.a. Insistence ('strong	Sub+V-al+aam		
e.g.1.You shall do as I say.	` -	. e.g 1.ndaan collukiRa paTi ndii		
2.He shall be punished.	•	ceyyalaam.		
	b. Legal and quasi-legal.	2.avan taNTikkappaTalaam.		
3.The vendor shall		s		
maintain the equipment in	1	3.teru viyaapaari tannuTaiya karuviyai		
good repair.		ceppam ceytu vaittirukkalaam.		
Sub+should + V1	I. Obligation and logica	I Sub+V + infinitive +veeNTum		
e.g 1.You should do as he	necessity (= ought to)	e.g 1.ndii avan colvatu pool ceyya		
says.		veeNTum.		
2.They should be home by	1	2.avarkaL ippootu viiTTil irukka		
now.		veeNTum.		
Sub+will/shall + V1	Future Tense	Sub+V + future + PNG		
e.g. 1.He will write.		e.g.1. avan ezhutuvaan.		
2.I shall write.		2.ndaan ezhutuveen		
Sub+will + V1		Sub+V + future tense + PNG		
e.g.1.I'll write as soon as	•	,1.epootu muTikiRatoo apootu ndaan		
can.	especially 2 <sup>d</sup> person.			
2.Will you have another	•	• • •		
cup of tea?	may be used to softer			
A 1 - 111 - 144	the tone in requests.			
Sub+will + V1	•	Sub+V+future tense+PNG		
e.g.1.l'll write as soon as		1.muTiyumpootu uTanee ezhutuveen.		
can.	contracted It; mainly 1st	2.ndaangkaL iraNTu maNi ndeerattiRku		
2.We won't stay longer	person.	meel tangka maaTToom.		
than two hours.	•	- I		

TAM forms in English with examples	Meaning	Equivalent TAM forms in Tamil with examples
Sub+will + V1  1.He 'will' do it, whateve you say ('He insists or doing it') (cf He 'shall' do it, whateve you say = 'I insist on his doing it')	r volition' = insist on Stressed, hence on 'l contraction.An r uncommon meaning.	Sub+V + future tense + PNG .e.g. ndiingkaL colvatai avan ceyvaan.
Would.+Sub+V1 e.g. Would you excuse me?	• •	Sub+V + future tense + PNG +aa ? e.g ndii ennai mannippaayaa?
Sub+would have+ V3	III. Insistence ('strong volition')	Sub+V + past participle form + iru + infinitive + veeNTum.
eg. It's your own fault, you would have taken the baby with you.	u ,	e.g. atu un tavaRu. ndii kuzhandtaiyai unnuTan eTuttuc cenRirukka veeNTum.
by 10 o'clock.  Sub+will, must, should - V1 e.g. The game	r compulsion in the present tense (= be obliged to, have to); except in reported speech. Onlyhad to (not must) is used in the past of the negative sentence needn't, don't have to not be obliged to are used (but notmust not, mustn't which = 'not be allowed to').  [Prediction of the similar meanings of other expressions for	2.Sub+V+pastParticiple +iru+infinitive+veeNTum+enRu e.g1.ndeeRRu pattu maNikku ndaan tirumpiyirukka veeNTum enRu ndii connaay.  Sub+V + infinitive *veeNTum e.g. viLaiyaaTTu indndeeram
will/must/should be finishe by now.	-	muTindtirukka veeNTum
Sub+will + V1, V1 e.g. Oil will float / floats o water.	Timeless Prediction n	Sub+V + future + PNG e.g. eNNai taNNiiril mitakkum
Sub+will, 'll +V1 e.g.He'll (always) talk for hours if you give him the chance.  1.Sub+ must, has to +V1		Sub+V + future +PNG e.g. avanukkuc candtarppam koTuttaal avan (eppozhutum) maNikkaNakkaakap peecuvaan. 1. Sub+irundtirukka+ veeNTum

TAM forms in English with examples	Meaning	Equivalent TAM forms in Tamil with examples
e.g. There must / has to be a mistake. (mustis not used in sentences with negative or interrogative meanings, can is being used instead.		e.g. tavaRu irundtirukka veeNTum.
2.Sub+ cannot +V1		2.Sub+irundtirukka+ muTiyaatu
e.g. There cannot be a		e.g
mistake.		tavaRu irundtirukka muTiyaatu.
Sub+ought to + V1	Obligation and logica	Sub+V + infinitive +veeNTum
e.g 1.You ought to start a	t necessity	e.g 1.ndii uTanee puRappaTa veeNTum
once.	•	2.avarkaL ipootu ingkee irukka veeNTum.
2.They ought to be here by		
now.		
Sub + used to + V1	A state of habit that	Sub+V-atu + vazhakkam + aay + iru +
e.g	existed in the past bu	t past + PNG
1.He used to fish for hours	has ceased. (cf:would	e.g 1.maNikkaNakkil miin piTippatu avan
2.He used to be an	and formerly or once +	vazhakkamaay irundtatu.
excellent cricketer.	past)	2.avan oru arumaiyaana kirikket aaTTakaaranaaka irundtaan

## Appendix 4

ARDLV – A sample of Arbitrarily Reordered Dictionary of Lexical Verbs in English and its inflected forms (V1\*, V1s\*, V2\*, V3\*, V4\*) with Tamil meaning (DT-Ditransitive; T-Transitive; IT-intransitive)

Present tense verb (V1)	3 <sup>rd</sup> person singular present tense verb(V1s)	Past tense Verb (V2)	Past participle Verb (V3)	Present participle (V4)	Tamil Meaning for the present tense verb
Pay (DT*)	Pays	Paid	Paid	Paying	Celuthu
Invite (T*)	Invites	Invited	Invited	Inviting	Azai
Go (IT*)	Goes	Went	Gone	Going	роо

The following points have to be noted while transferring TAM system of English into Tamil.

- 1. Both English and Tamil make use of inflection as well as compounding (i.e. combining main verbs with the auxiliary verbs) to express TAM.
- 2. The important point to be noted from the point of view of word order is that auxiliary verbs in English precede the main verb in English, whereas in Tamil they follow the main verb
- 3. In English auxiliary verbs are inflected for Tense, Person and Number; whereas in Tamil they are inflected for Tense, Person, Number and Gender.
- 4. Both English and Tamil undergo auxiliary reduction.
- 5. Identical auxiliary verbs in complex constructions get deleted both in English and Tamil.
- 6. Auxiliary verbs occur in a sequence to denote tense, mood, aspect, voice etc. in both English and Tamil. The modal auxiliary verb in English never occurs after a primary auxiliary verb in English, whereas, in Tamil primary auxiliary verb never occurs after a modal auxiliary verb (with the exception of few modal auxiliary verbs).
- 7. Both Tamil and English express perfective and progressive sense by auxiliary verbs. But Tamil makes use of this device to express the completive and reflexive sense too.
- 8. In English, 'Yes or No' interrogative sentences are derived from their respective affirmative sentences by shifting the relevant auxiliary verbs to the initial position and a question mark at the end of it. In Tamil, 'Yes or No' interrogative sentences are derived from their respective affirmative sentences by adding interrogative clitic 'aa' with question mark.
- 9. This may be true with other Dravidian languages

#### References

Deshpande, W.R. 1994. Machine Translation: State of the Art (manuscript). Department of Electronics, Government of India.

Edward, W.T. 1973. A Contrastive Analysis of Tamil and English. Ph.D. Thesis. Annamalainagar: Annamalai University.

Harris, Z. 1954 Transfer Grammar. In: International Journal of American Linguistics vol. 20, pp 259-270.

Hausser, R. 1999. Foundations of Computational Linguistics: Man-machine Communication in Natural Language. Springer.

Hornby, A.S. 1975. Guide to Patterns and Usage in English. London: Oxford University Press.

Hutchins, W.G. 1986. Machine Translation: Past Present and Future. Ellis Horwood.

Isabelle, P. 1993. Machine-Aided Human Translation and the Paradigm Shift (manuscript). Japan: MT Smmit IV.

Ritchie, G.D. et 1992. al. Computational Morphology: Practical Mechanisms for the English Lexicon. MIT.

Sproat, R.1992. Morphology and Computation. MIT.

Theivanatham Pillai, K. 1970. A Comparative Study of the English and Tamil Auxiliary Verb Systems and Prediction of Learning Problems for Tamil Students of English.

Thiyagarajan, K. 1881. Modal System of English and Tamil. Ph.D. Thesis. Madras: University of Madras.

## TRANCRIPTION FOR TAMIL ALPHABETS HAVE BEEN USED.

Roman	Tamil	Roman	Tamil
a	a	nj	<b>6</b>
aa	್ರಿ	T	L
i	<b>@</b>	N	ண
ii	正	t	Б
u	2_	nd	Б
uu	<u>ூள</u>	p	П
e	ឥ	m	Ю
ee	ஏ	y	Ш
ai	22	r	σ
0	99	1	ស
00	ଚ୍ଚ	V	อ
au	ஓள	zh	ſδ
k	க	L	ள
ng	Ш	R	р
С	ச	n	ன

#### **Acknowledgements:**

Miss. Kamakshi Shanmuganantham gratefully acknowledges her deep indebtedness to her doctoral guide Prof S.Rajendran , Dept.of Linguistics, Tamil University , Thanjavur for his constant encouragement, and Computer Programmer Mr. T. Ronald of Linguistics Studies Unit, University of Madras for his patient and adept listening in understanding the intricacies involved in Natural Languages while comparing them with artificial languages like Visual Basic etc., for digitizing my concept of this work..