

Implementation of Subject-Predicate Agreement in Hindi and Telugu: A Machine Translation Perspective

Christopher Mala
LTRC

International Institute of Information Technology
Hyderabad-500046, India
efthachris@gmail.com

G. Uma Maheshwar Rao
CALTS

University of Hyderabad
Hyderabad - 500046, India
guraohyd@gmail.com

Abstract

The study presented here provides a description of the transfer of the Grammatical Agreement from Hindi to Telugu in the context of Machine Translation. The quality of translation is extremely dependent on the appropriateness of the agreement transfer. The Relationship between syntax and semantics of agreement can best be captured through the investigations on the expression of Agreement realization on the finite verb. In Hindi, agreement is observed between the Finite verb and its complement noun in the nominative except when the verb expresses default agreement in such case nouns are inflected overtly for cases.

One of the major goals of this study is to come up with a design to implement a computational model of the agreement in Hindi-Telugu Machine Translation System that is being developed at CALTS-HCU.

0.1 Key words:

S=Subject, P=Predicate, m=Masculine, f=feminine, sg=singular, pl=plural, h=human, !gnp=no gender and number and person, 1p=1st person, g=gender, n=number, p=person, AGRi=agreement coindex.

0.2 WX-Notation

Notation used in this paper
a A i I u U eV e E oV o O M H
k K g G f, c C j J F, t T d D N,
w W x X n, p P b B m, y r l v S R s h

1 Introduction

The phenomenon of Grammatical Agreement, which functions like a bridge between the boundaries of morphology and syntax, is one of the interesting aspects of synchronic study of languages. Grammatical agreement mainly explains the phenomena that exhibit the specific morphological form of a word (Predicate) appearing in a sentence with respect to the presence of some other word (Subject or object in case of Hindi) elsewhere in the sentence. This is probably why Lehmann (1988:55) prefers to call agreement to be referential in nature. This referential in turn helps to retrieve the referent(s). Agreement does this by encoding the information of the grammatical properties of its referent(s) (i.e. the NP-Subject or Object) into the morphological elements that appear(s) with the word (Predicate) in the sentence. In other words, it deals with the selection of the appropriate inflectional categories of the word (i.e. the verb) with respect to the properties of other words (i.e. the PNG of the Subj/Obj NP) in the sentence. It is for this reason, grammatical agreement is said to be closely related to the inflectional morphology in nature. Grammatical agreement helps us to explore and explicate how languages are structured. Here we are trying to study the grammatical agreement between the Subject (or Object in Hindi) and Predicate in Hindi and the computational procedures for generating appropriate agreement in the

1.1 Preliminaries to Grammatical Agreement in Hindi and Telugu:

Predicate Agreement in Hindi and Telugu differs in various respects. One of the major differences being that Hindi is an ergative - accusative language whereas Telugu is nominative unaccusative language and has non-verbal predicates. The third major difference comes from the difference in the gender systems of the two languages. Hindi has two genders, viz. masculine and feminine both in singular and plural whereas Telugu has masculine vs. non-masculine in singular while human vs. non-human in plural. Beyond these differences the operation of Agreement is similar and word order doesn't effect in any way with agreement. Consider the following sentences in Hindi and Telugu for the exemplification of basic agreement facts about the agreement phenomena in Hindi and Telugu.

$rAma(X)_{3,m,sg,(X')} \quad gAna(Y)_{3,f,s,(Y')} \quad gAwa(Z)-$
 $hE_{(TAM),3,sg,(Z')}$
 $rAmudu(X)_{3,m,sg,(X')} \quad pAta(Y)_{3,nm,sg,(Y')} \quad pAdu(Z)-$
 $wAdu_{(TAM),3,m,sg,(Z')}$

We can interpret the above statement in the following way. The verb (i.e. Z) displays inflectional categories in terms of co-indexing with the inflectional categories (X') of an NP (Subject i.e. X). The first condition laid out in the statement is very clear. There exists a syntactic relationship between the verb and the nominative NP in the sentence. We will have to assume that there is an underlying relationship between the features (i.e. the PNG represented by X' and Z' here) corresponding to the noun phrase (subject) and the finite verb (predicate) in the sentence and this relationship is independent of the nature or kind of the verb. The category Z' i.e. the Agreement inflection is a subcategory of the category Z (the predicate). The last condition suggests the formation of a constituent and this happens when the markers of agreement feature appear(s) with the predicate in the sentence. A few examples of the aforementioned statement and its explanation that is presented will help to see through the complex features and their expression as agreement.

The verb (Predicate Z) displays its inflectional

categories (Z') in terms of co-indexing with the corresponding inflectional categories of (X') of the noun phrase (the Subject X). The following conditions apply:

1. Only a nominative NP qualifies to be referential.
2. Only Predicates or constituents of predicates qualify to be co-indexed (Adjectives, Nouns, Pronouns, Locative nouns, Number words beside verbs in Telugu).
3. Only a predicate displays inflectional endings corresponding to the morphological categories of the subject NP (or object NP in case of Hindi).
4. If there is more than one NP in the nominative then the ontological hierarchy decides what should be the candidate for subject NP.

2 Exponences of Agreement in Hindi and Telugu

Agreement between Verb and its Complements is expressed overtly in 12 ways in Hindi and 8 ways in Telugu. We have excluded honorifics from this list as it is not a distinct feature value since plural agreement form is used instead of a distinct honorific form. Honorific forms do not show contrasts with the plural forms. Here are the possible Hindi and Telugu exponences of Agreement features viz. gender, number and person:

2.1 Hindi

1. Total no. of feature values for the feature Gender: (2) Masculine(M), Feminine(F)
2. Total no. of feature values for the feature Number: (2) Singular(sg), Plural(pl)
3. Total no. of feature values for the feature Person: (3) 1, 2, 3

These features and their feature values in various combinations result in 12 types of exponence of agreement in Hindi as in the Table -1:

| Gen. | Num. | Per. | Example |
|------|------|------|-------------------------------|
| M | Sg | 1 | <i>mEM AwA hUM</i> |
| M | Pl | 1 | <i>hama Awe hEM</i> |
| F | Sg | 1 | <i>mEM AwI hUM</i> |
| F | Pl | 1 | <i>hama AwI hEM</i> |
| M | Sg | 2 | <i>wU Awe ho</i> |
| M | Sg | 2h | <i>Apa Awe ho</i> |
| M | Pl | 2 | <i>wuma(loga) Awe hEM</i> |
| F | Sg | 2 | <i>wU AwI ho</i> |
| F | Sg | 2h | <i>Apa AwI ho</i> |
| F | Pl | 2 | <i>wuma(loga) AwI hEM(ho)</i> |
| M | Sg | 3 | <i>vaha AwA hE</i> |
| M | Sg | 3h | <i>ve Awe hEM</i> |
| M | Pl | 3 | <i>ve Awe hEM</i> |
| F | Sg | 3 | <i>vaha AwI hE</i> |
| F | Sg | 3h | <i>ve AwI(M) hEM</i> |
| F | Pl | 3 | <i>ve AwI(z) hEM</i> |

Table 1: Exponences of GNP in Hindi

| Gen. | Num. | Per. | Example |
|--------|------|------|--------------------------|
| m/nm | Sg | 1 | nenu vaswAnu |
| hum/nh | Pl | 1 | memu vaswAmu |
| m/nm | Sg | 2 | nuvvu vaswAvu |
| hum/nh | Pl | 2 | mIru vaswAru |
| m | Sg | 3 | (a)wanu vaswAdu |
| nm | Sg | 3 | wanu/axi vaswuMxi |
| nh | Pl | 3 | avi vaswAyi |
| hum | Pl | 3 | vAIYlu vaswAru |

Table 2: Exponences of GNP in Telugu

2.2 Telugu

1. Total no. of feature values for the feature Gender: (4) Masculine(M), Non-Masculine(NM) in singular and Human(MF/HUM) , Non-Human(NH) in plural.
2. Total no. of feature values for the feature Numbers: (2) Singular(sg), Plural(pl)
3. Total no. of feature values for the feature Person: (3) 1, 2, 3

The exponences of these gnp are listed in Table -2

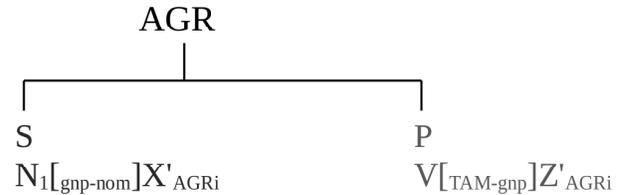
3 Types of Agreement:

With respect to the agreement operations, in Hindi, the expression of agreement in the finite verb

is realized in five distinct ways (discounting morphological identities in the exponence). There are **three major types** of Agreement realization between the finite verb and its complement noun and two types of agreement one involving copula verb and the other one with the prenominalized adjective.

3.1 Type-1

In Hindi this type is characterized by the realization of agreement by the finite form of verb (other than the past perfect) with respect to the subject NP in the nominative. ex.



Hin: ladakA Kela rahA hE. 'The boy is playing'

Boy play -ing is

Tel: abbAyi AduwunnAdu_(3,sg,m).

In this type, the gnp features of the subject noun(nom.) is expressed on the finite verb both in Hindi and in Telugu.

3.2 Type-2

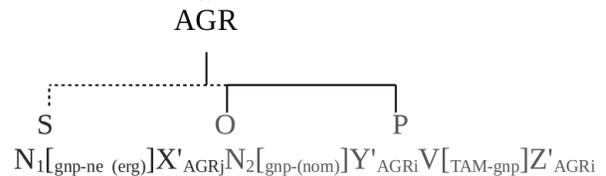
This type of agreement occurs in Hindi when the verb is inflected for perfective past and the agentive noun is marked for ergative. In this type, the external argument is in ergative, and the direct object expresses its gnp features on the finite verb. This type can be seen only in Ergative languages like Hindi but not in Telugu as described by the example of Telugu. ex.

Hin: rAma ne rotI KAI. 'Ram has eaten the bread'

rAma_{erg} bread eat_(perf.pt,f.sg3)

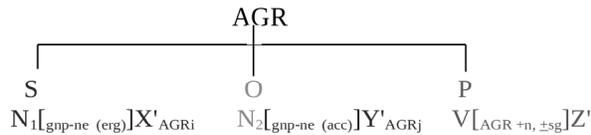
Tel: rAmudu roVtteV winnAdu

rAmudu bread eat_(pt.fin.m.sg.3)



3.3 Type-3

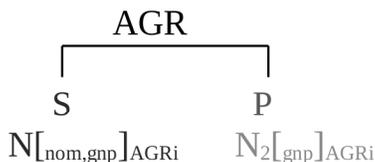
This type of agreement is a special kind found in Hindi. In this case the verb's gnp features are realized as the unmarked masculine singular irrespective of the gender, number and person of the arguments of the verb. In Hindi when the agent NP is marked for ergative and the object is marked for Accusative (Indirect object by dative) then the verb displays default agreement. ex.



Hin: meM/hama ne usako xekA. 'I saw him'
 I/We_{erg} him_{Acc} saw.
 Tel: nenu/memu wanani cUsAnu/mu^(!g,1/2,sg/pl).

3.4 Type-4

In Hindi often a verb is used to link the subject NP with a noun or an adjectival complement. In cases where nouns are complements equational construction, the copula displays agreement with the subject. However in Telugu the predicate nouns agree with the subjects as there is no the copula in the affirmative sentences. ex.



- Hin: mEM dAktar hUM. 'iam doctor'
 I doctor is
 Tel: nenu dAktarnu^(any,sg,1).
- Hin: wU dAktar ho. 'you are doctor'
 You doctor are
 Tel: nuvvu dAktarvu^(any,sg,2).
- Hin: Apa dAktar hEM. 'you are doctor'
 You doctor are
 Tel: mIru dAktarlu^(any,sg,3).

3.5 Type-5

Unlike in type 4, where nouns are used as non-verbal predicate nominals, in this type adjectives which occur as predicative adjectives show agreement with the subject. Both the copula and the adjective show agreement with the subject. ex.

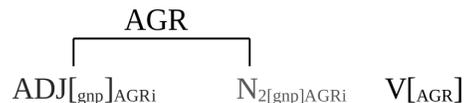


- Hin: ladakA acCA hE. 'boy is good'
 Tel: abbAyi maMciVAdu^(m,sg,3).
- Hin: ladakI acCI hE.
 Tel: ammAyi/axi maMciXi^(nm,sg,3).
- Hin: ladakiyAM acCIz hEM. 'girls are good'
 Tel: ammAyilu/abbAyilu
 maMciVAIYlu^(hum,pl,3).

When the predicate of the clause is a non-verb (adjective) a special kind of Agreement is found both in Telugu and Hindi. The predicate adjectivals show agreement with the subject of the clause. However Hindi shows a copula in the corresponding sentences.

3.6 Type-6

This type involves Agreement between an Adjective and its complement NP. An Adjective modifying a Noun (head) may show its agreement features with its head only in Hindi. This is not found in Telugu, i.e. Telugu prenominal adjectives usually don't show agreement with its complement nouns. ex.

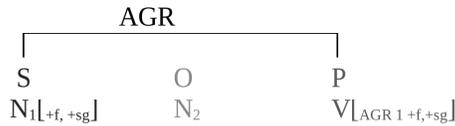


Hin: acCA ladakA. 'good boy'
 Tel: maMciV abbAyi.

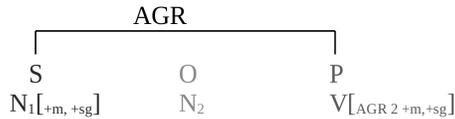
4 Realization of Exponence and its mapping on verbs

- When the subject noun (pronoun) is feminine and the number is singular, the finite verb also displays the same gender and number. ex.

Hin: vaha AwI hE
Tel: wanu vaswuMxi

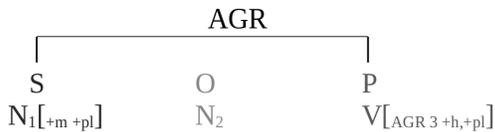


2. When the subject noun is of masculine and number is singular, the finite verb displays the same gender and number. ex.



Hin: vaha AwA hE
Tel: wanu vaswAdu

3. When the subject noun is marked for +masculine and the number is plural, the finite verb is marked for the same gender and number. ex.



Hin: ve Awe hEM
Tel: vAru vaswAru

4. When the subject noun is feminine and the number is plural the same features are expressed on the finite verb. ex.

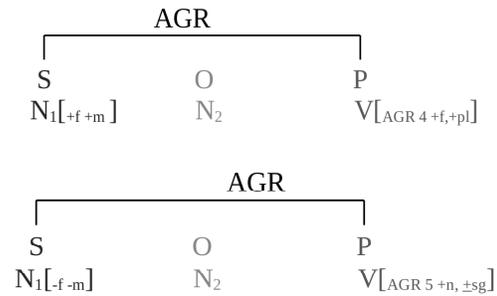
Hin: ve AwI hEM
Tel: vAru vaswAru

5. In this type of Agreement, the finite verb displays its agreement in a default. The default agreement is usually the masculine singular 3p exponence ex.

Hin: meM/hama ne usako xeka
Tel: nenu/memu wanani cUsAnu/mu.

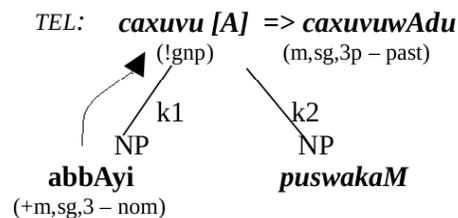
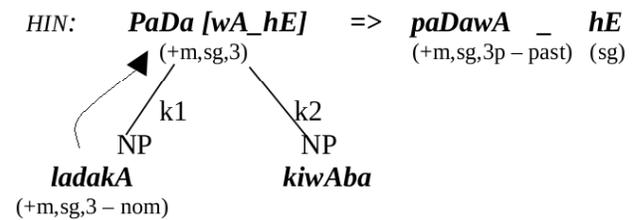
5 Agreement Transfer between Hindi Telugu

Though nouns in Hindi either belong to masculine or feminine, the expression of gender and number on the finite verb leaves space for unmarked masculine

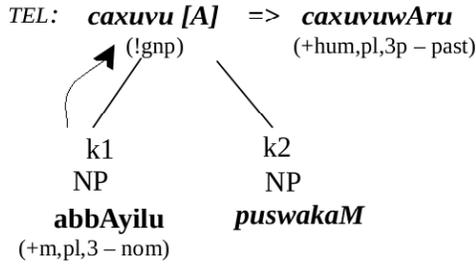
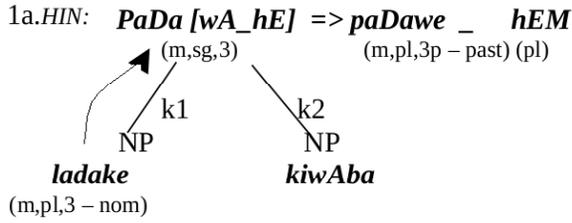


displayed as in the type 3 and the exponence of realization as in 5. Considering the above three types of expression of Agreement on the Finite verb and five distinct realizations of the morpho-syntactic functional categories on the finite verb in Hindi, one may get a total of 15 distinct categories of the expression of Agreement feature for transfer from Hindi to Telugu. However in Telugu, considering the kind of gnp mapping of noun into masculine Vs. non-masculine in singular and Human Vs. Non-Human in plural. There is only four ways of direct mapping on the finite verb are available. There is also a special case where the finite verb is marked for unmarked masculine singular (neutral) agreement feature. In the transfer of Agreement from Hindi to Telugu the gnp of subject noun(nom.) features or noun(erg.) features are replaced with the relevant gnp properties of Subject noun in Telugu and then passed on to the finite verb for mapping as in the following examples.

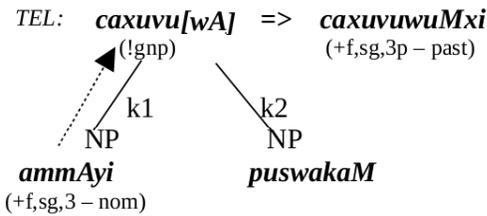
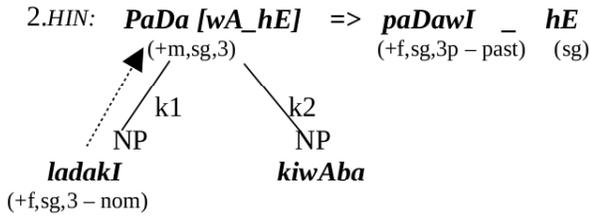
1. **Hin:** ladakA_{{+m,+sg,3p}ARGi} kiwAba
paDawA_{{+m,+sg}ARGi} hE_{{+sg,3p}ARGi}
Tel: abbAyi_{{+m,+sg,3p}ARGi} puswakaM
caxuvuwAdu_{{+m,+sg,3p}ARGi}



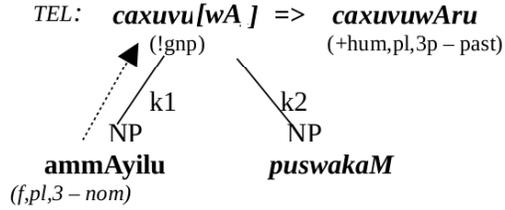
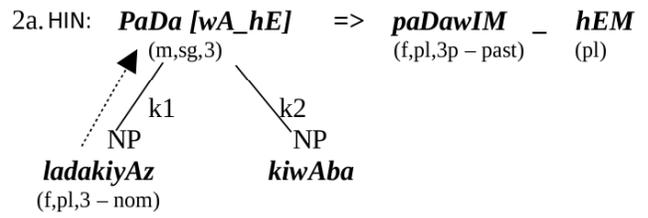
1a. **Hin:** ladake_{{+m,+pl,3p}AGRi} kiwAba
 paDawe_{{+m,+pl}AGRi} hEM_{{+pl,3p}AGRi}
Tel: abbAyilu_{{+m,+pl,3p}AGRi} puswakaM
 caxuvuwAru_{{+h,+pl,3p}AGRi}



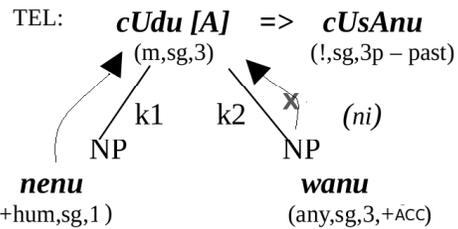
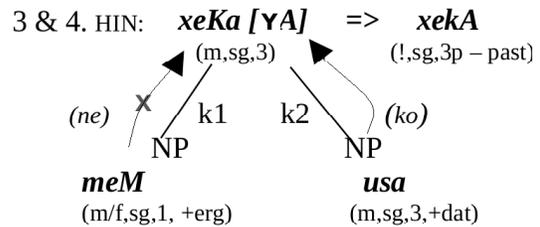
2. **Hin:** ladakI_{{+f,+sg,3p}AGRi} kiwAba
 paDawI_{{+f,+sg}AGRi} hE_{{+sg,3p}AGRi}
Tel: ammAyil_{{+f,+sg,3p}AGRi} puswakaM
 caxuvuwuMxi_{{+f,+sg,3p}AGRi}



2a. **Hin:** ladakiyAz_{{+f,+pl,3p}AGRi} kiwAba
 paDawIM_{{+f,+pl}AGRi} hEM_{{+pl}AGRi}
Tel: ammAyilu_{{+f,+pl,3p}AGRi} puswakaM
 caxuvuwAru_{{+pl,3p}AGRi}



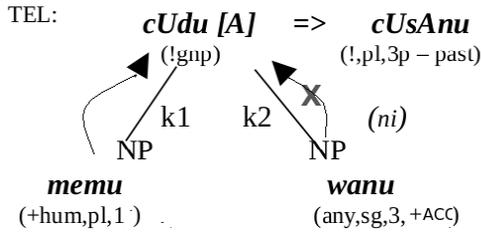
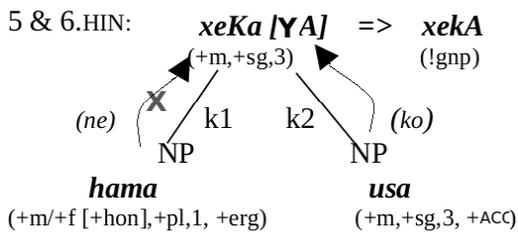
3. **Hin:** meM_{{!g(+m),+sg,1p}AGRi} ne usko
 xeKA_{{!g(+m),+sg}AGRi}
Tel: nenu_{{!g,+sg,1p}AGRi} wanani
 cUsAnu_{{!g,+sg,1p}AGRi}



4. **Hin:** meM_{{!g(+f),+sg,1p}AGRi} ne usko
 xeKA_{{!g(+f),+sg}AGRi}
Tel: nenu_{{!g,+sg,1p}AGRi} wanani
 cUsAnu_{{!g,+sg,1p}AGRi}

5. **Hin:** hama_{{+m,+pl/+hon,+1p}AGRi} ne usko
 xeKA_{{!g,+sg}AGRi}
Tel: memu_{{!g,+pl,1p}AGRi} wanani
 cUsAmu_{{!g,+pl,1p}AGRi}

6. **Hin:** hama_{{+f,+pl,+1p}AGRi} ne usko
 xeKA_{{!gnp}AGRi}



Tel: memu_{{+h,+pl,+1p}AGRi} wanani
 cUsAmu_{{!g,+pl,1p}AGRi}

6 Implementation of Agreement in Machine Translation System

It is assumed that the proposed agreement rules work with almost all sorts of finite structures in Hindi. The system identifies the boundaries of agreement source and the target. The proposed Agreement rules apply within these boundaries even in a complex or compound sentences identification of Agreement boundaries ensures correct agreement.

After the analysis of the source language, The text (Hindi) passes through the lexical substitution module where Hindi lexical items are substituted by the corresponding Telugu items in the Machine Translation System. It is now ready for the Agreement generation. In this process the gnp features of Telugu subject noun are passed on to the corresponding predicates first. This agreement tell us about the sentence wellformedness (faithfulness) and also increases the readability after the translation is completed. When readability is increased the Machine translated text is very close to human translation. Here are the Five Steps that are followed to carry out the Agreement from Hindi to Telugu.

1. Find the Agent/karwa (k1) in the sentence and pass it's GNP to the predicate in the sentence.
2. Retrieve the gender of the lexicon according to the Target Language (Telugu).

3. If k1 is a pronoun (subject pronoun) then the gender should be passed back, i.e. from predicate to its subject.
4. The gender of the nonverbal predicate (adjective and noun) should be passed on to the subject NP.
5. In case the predicate has a copula then pass the gender number and person of the copula to the preceding non-verb.

ex:

$vaha_{sg,3} acCA_{m,sg} hE_{m,sg,3} = vAdu maMci-vAdu.$

7 Conclusion

As one of the major goals of this study, we have demonstrated the efficacy of the design of a the computational model for the transfer of grammatical agreement which is implemented in the Machine Translation (MT) System of Hindi-Telugu and Telugu Hindi that is being developed at CALTS(HCU).

References

- Boeckx, Cedric 2008. Aspect of the Syntax of Agreement New York: Routledge
- G. Uma Maheshwar Rao. A Computational Grammar of Telugu: University of Hyderabad (memo).
- Agnihotri, Rama Kant. 2007. Hindi An Essential Grammar. New York: Routledge.
- Krishnamurti, Bh. 1985. A grammar of Modern Telugu. Delhi, New York: Oxford University Press.
- Krishnamurti, Bh. 2003. The Dravidian Languages. Cambridge University Press, Cambridge, UK.
- Beames, John. 1966. A Comparative Grammar of Modern Aryan Languages of India. Delhi.
- Das, Pradeep Kumar 2006. Grammatical Agreement in Hindi-Urdu and its major varieties, Munchi, LinconEuropa.
- Lehman, Thomas. 1989. Agreement if Mordern Tamil PILC, Pondicerry.