

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

Christopher,<sup>2</sup> M., Uma Maheshwar Rao<sup>1</sup>. G.

<sup>1</sup>CALTS, University of Hyderabad,

<sup>2</sup>LTRC, International Institute of Information Technology

Gachibowli, Hyderabad-46.

[guraohyd@gmail.com](mailto:guraohyd@gmail.com), [efthachris@gmail.com](mailto:efthachris@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 neutral in agreement in such case nouns are inflected for dative cases.

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

### **Key words:**

S=Subject, P=Predicate, Masculine=m, feminine=f, singular=sg, plural=p, human=h, no gender number person=!gnp, 1<sup>st</sup> person = 1p.

WX-Notation:

a A i I u U e V e E o V 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 I Y 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) 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) 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 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 Subject and Predicate in Hindi and the computational procedures for generating appropriate agreement in the Target Language (Telugu) in a Machine Translation System (IL-ILMT).

## 1.1 Definition of ‘Grammatical Agreement’:

*rAmudu*

*pAta*

*pAdAdu.*

*rAmudu(X)-{3,m,s}(X') pAta(Y){3,nm,s}(Y') pAdu(Z)-A(TAM){3,m,s}(Z')*

We could 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 of an NP (Subject i.e. X). The first condition laid out in the definition is very clear. There exists a syntactic relationship between the verb and the NP in the sentence. We will have to assume that there is an underlying relationship between the features (i.e. the PNG represented as is ‘x’ here) and the noun phrase in the sentence and this relationship is independent of the nature or kind of the verb. The at category Z' the inflection is a subcategory of the category ‘Z’ (the verbal inflection). The last condition suggests about the formation of a constituent (i.e. the VP) and this happens when the bound form(s) or the null marker of agreement feature appear(s) with the verb in the sentence. A few examples of the aforementioned definition and its explanation that is presented will help to see through the complex features and their function for agreement.

The verb (Predicate Z) displays its inflectional categories (Z") in terms of coindexing with the corresponding inflectional categories of (X') of the noun phrase (the Subject X). The following conditions apply:

- Only nominative NP qualify to be referential.
- Only a Finite verb (Predicate) qualifies to be coindexed.
- Only a predicate displays inflectional endings corresponding to the morphological categories of the subject NP
- If there is more than one NP in the nominative then ontological hierarchy decides what should be the candidate for subject NP.

## 2. Exponences of Agreemental in Hindi and Telugu

Agreement between Verb and its Compliments is expressed overtly in 12 ways in Hindi and 8 ways in Telugu. We may drop honorific as a distinct feature vale since plural agreement form is used instead of a distinct honorific form. Honorific forms do not show contract with the plural forms. Here are the possible Hindi and Telugu exponences of Agreement features gender, number and person:

### **Hindi :**

Total no. of feature values for the feature Genders: (2) Masculine(M), Feminie(F)

Total no. of feature values for the feature Numbers: (2) Singular(sg), Plular(pl)

Total no. of feature values for the feature Persons: (3) 1, 2, 3

These features and their feature values in various combinations result in 16 types of exponences of these gnp are as following:

S.No	Gender	Number	Person	Example
1	M	Sg	1	<i>mEM AwA hUM</i>
2	M	Pl	1	<i>hama Aw e hEM</i>
3	F	Sg	1	<i>mEM AwI hUM</i>
4	F	Pl	1	<i>hama AwI hEM</i>

5	M	Sg	2	wU Awe <b>ho</b>
	M	Sg	2h	Apa Awe <b>ho</b>
6	M	Pl	2	wuma(loga) Awe <b>hEM</b>
7	F	Sg	2	wU Aw <b>I ho</b>
	F	Sg	2h	Apa Aw <b>I ho</b>
8	F	Pl	2	wuma(loga) Aw <b>I hEM(ho)</b>
9	M	Sg	3	vaha Aw <b>A hE</b>
	M	Sg	3h	ve Awe <b>hEM</b>
10	M	Pl	3	ve Awe <b>hEM</b>
11	F	Sg	3	vaha Aw <b>I hE</b>
12	F	Sg	3h	ve Aw <b>I(M) hEM</b>
13	F	Pl	3	ve Aw <b>I(z) hEM</b>

### Telugu :

Total no. of feature values for the feature Gender: (4) Masculine(M), Non-Masculine(F) in singular and Human(MF) , Non-Human(N) in plural.

Total no. of feature values for the feature Numbers: (2) Singular(sg), Plural(pl)

Total no. of feature values for the feature Person: (3) 1, 2, 3

S.No	Gender	Number	Person	Example
1	m/f	Sg	1	nenu vaswA <b>nu</b>
2	m/f	Pl	1	memu vaswA <b>mu</b>
3	m/f	Sg	2	nuvvu vaswA <b>vu</b>
4	m/f	Pl	2	mIru vaswA <b>ru</b>
5	m	Sg	3	(a)wanu vaswA <b>du</b>
6	f/n	Sg	3	wanu/axi vaswu <b>Mxi</b>
7	n	Pl	3	avi vaswA <b>yi</b>
8	m/f	Pl	3	vAIYlu vaswA <b>ru</b>
	m/f	Sg	3h	vAru vaswA <b>ru</b>

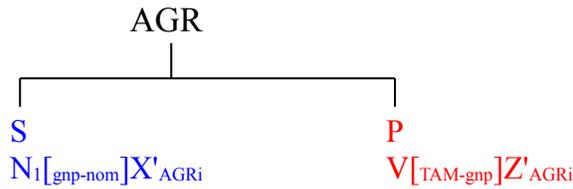
### 3. Types of Agreement:

In Hindi, the expression of agreement in the finite verb is realized in six distinct (discounting morphological identities in the exponence) morpho-syntactic categories. There are three **major types** of Agreement realization between the finite verb and its complement noun.

#### 1) Type:

In Hindi this type is characterized by the the realization of agreement by the finite

form of verb (other than the past perfect) with respect to the subject NP in the nominative.

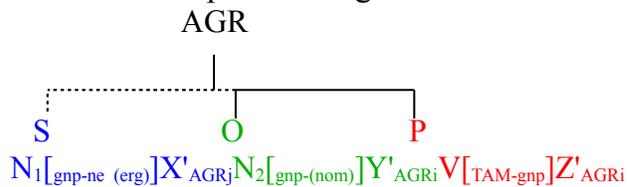


eg: Hin: **ladakA** Kela rah**A hE**. 'The boy is playing'  
 Boy play -ing is  
 Tel: **abbAyi** Aduwunn**Adu**.

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

## 2) Type:

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.



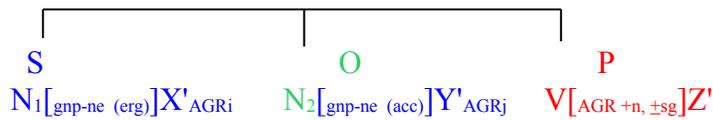
eg: Hin: **rAma** ne **rotI** **KAI**. 'Ram has eaten the bread'  
 rAma erg. bread eat[perf.pt,f.sg3]

Tel: **rAmudu** **roVtteV** winn**Adu**  
 rAmudu bread eat-[pt.fin.m.sg.3]

## 3) Type:

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 agent is NP is marked for ergative and object is marked for Accusative (Indirect object by dative) then the verb displays agreement with neutral form.

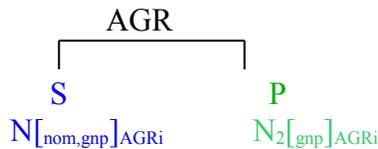
AGR



eg: Hin: *meM/hama ne usako xekA*  
 Tel: *nenu/ memu wanani cUsA mu/nu*.

#### 4) Type:

In Hindi often a verb is used to link a subject with a verb less predicate either a noun or an adjective. In cases where nouns are the predicates, the copula displays agreement with the subject. However in Telugu with the subjects as there is no copula in affirmative sentences.



eg: 1) Hin: *mEM dAktor hUM*.  
 Tel: *nenu dAktor nu*<sub>(any,sg,1)</sub>.

2) Hin: *wU dAktor ho*.  
 Tel: *nuvvu dAktor nu*<sub>(any,sg,2)</sub>.

3) Hin: *Apa dAktor hEM*.  
 Tel: *mIru dAktor lu*<sub>(any,sg,3)</sub>.

#### 5) Type:

Unlike in type 4, where nouns are used as non-verbal predicates, here adjectives are used as predicates which show agreement with the subject. Both the copula and the adjective show agreement with the subject.



eg: 1) Hin: *ladakA acCA hE*.  
 Tel: *abbAyi maMcivA du*<sub>(m,sg,3)</sub>.

2) Hin: *ladakI acCI hE*.  
 Tel: *ammAyi /axi maMcixi*<sub>(f/n,sg,3)</sub>.

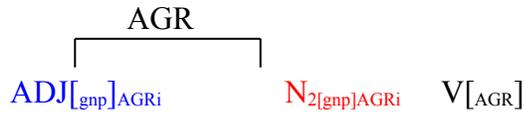
3) Hin: *ladakiyAM acCIz hEM*.  
 Tel: *ammAyilu/abbAyilu maMcivAIY lu*<sub>(f,pl,3)</sub>.

When the predicate of the clause is a non-verb (adjective) a special kind of Agreement 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.

### 6) Type:

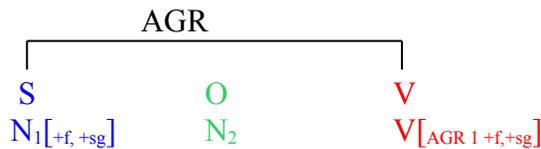
Agreement between an Adjective and its complement NP. An Adjective modifying a Noun (head) may show its agreement features with its head only in Hindi.



eg: Hin: acCA ladakA.  
Tel: maMci abbAyi.

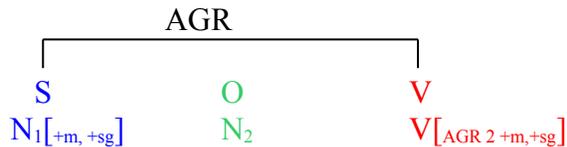
## 4. Realization of Exponence and its mapping on verbs:

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



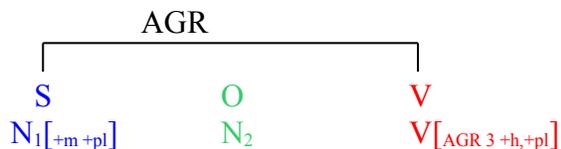
eg: Hin: vaha KAwi hE  
Tel: wanu vaswuMxi

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



eg: Hin: vaha AwA hE  
Tel: wanu vaswAdu

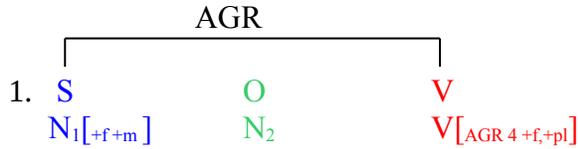
3. When the subject noun is marked for +masculine and the number is plural, the finite verb is marked for the same gender and number.



eg: 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.



eng: Hin: *ve AwI hEM*  
Tel: *vAru vaswAru*

5. In this type of Agreement displaying the finite verb, its complement subject can be of any gender, number and person.



eg: 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 exponents of realization 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 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 following examples.

1. Hin: *ladakA*<sub>{+m,+sg,3p}AGRi</sub> *kiwAba paDawA*<sub>{+m,+sg}ARGi</sub> *hE*<sub>{+sg,3p}AGRi</sub>  
Tel: *abbAyi*<sub>{+m,+sg,3p}AGRi</sub> *puswakaM caxuvuwAdu*<sub>{+m,+sg,3p}AGRi</sub>

1a. Hin: *ladake*<sub>{+m,+pl,3p}AGRi</sub> *kiwAba paDawe*<sub>{+m,+pl}AGRi</sub> *hEM*<sub>{+pl,3p}AGRi</sub>  
Tel: *abbAyilu*<sub>{+m,+pl,3p}AGRi</sub> *puswakaM caxuvuwAru*<sub>{+h,+pl,3p}AGRi</sub>

2. Hin: *ladaki*<sub>{+f,+sg,3p}AGRi</sub> *kiwAba paDawI*<sub>{+f,+sg}AGRi</sub> *hE*<sub>{+sg,3p}AGRi</sub>  
Tel: *ammAyi*<sub>{+f,+sg,3p}AGRi</sub> *puswakaM caxuvuwuMxi*<sub>{+f,+sg,3p}AGRi</sub>

2a. Hin: *ladakiyAz*<sub>{+f,+pl,3p}AGRi</sub> *kiwAba paDawIM*<sub>{+f,+pl}AGRi</sub> *hEM*<sub>{+pl}AGRi</sub>

Tel: *ammAyilu* <sub>{+f,+pl,3p}AGRi</sub> *puswakaM caxuvuwAru* <sub>{+pl,3p}AGRi</sub>

3. Hin: *meM* <sub>{!g(+m),+sg,1p}AGRi</sub> *ne usko xeKA* <sub>{!g(+m),+sg}AGRi</sub>

Tel: *nenu* <sub>{!g,+sg,1p}AGRi</sub> *wanani cUsAnu* <sub>{!g,+sg,1p}AGRi</sub>

4. Hin: *meM* <sub>{!g(+f),+sg,1p}AGRi</sub> *ne usko xeKA* <sub>{!g(+f),+sg}AGRi</sub>

Tel: *nenu* <sub>{!g,+sg,1p}AGRi</sub> *wanani cUsAnu* <sub>{!g,+sg,1p}AGRi</sub>

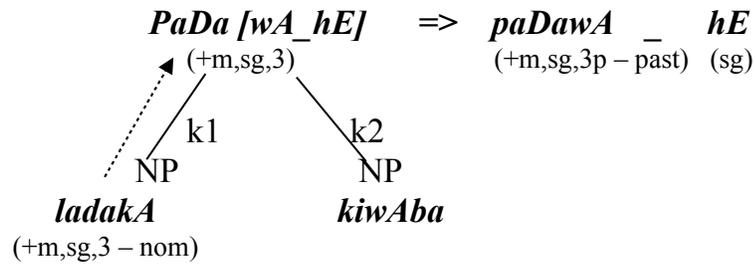
5. Hin: *hama* <sub>{+m,+pl/+hon,+1p}AGRi</sub> *ne usko xeKA* <sub>{!g,+sg}AGRi</sub>

Tel: *memu* <sub>{!g,+pl,1p}AGRi</sub> *wanani cUsAmu* <sub>{!g,+pl,1p}AGRi</sub>

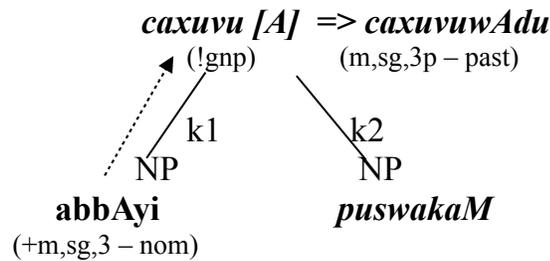
6. Hin: *hama* <sub>{+f,+pl,+1p}AGRi</sub> *ne usko xeKA* <sub>{!gnp}AGRi</sub>

Tel: *memu* <sub>{+h,+pl,+1p}AGRi</sub> *wanani cUsAmu* <sub>{!g,+pl,1p}AGRi</sub>

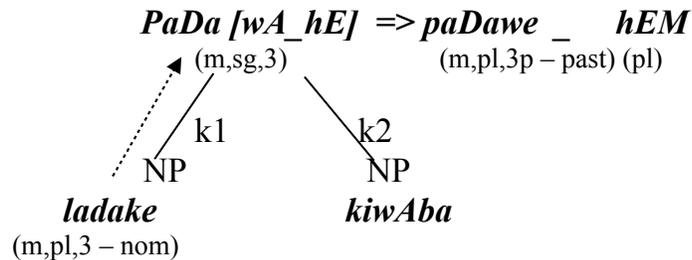
1.HIN:



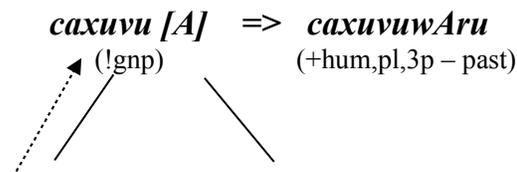
TEL:



1a.HIN:



TEL:



k1                      k2  
 NP                      NP  
**abbAyilu**              **puswakaM**  
 (+m,pl,3 – nom)

2.HIN:

**PaDa [wA\_hE]** => **paDawI \_ hE**  
 (+m,sg,3)                      (+f,sg,3p – past)    (sg)  
 k1                      k2  
 NP                      NP  
**ladakI**                      **kiwAba**  
 (+f,sg,3 – nom)

TEL:

**caxuvu [A]** => **caxuvuwuMxi**  
 (!gnp)                      (+f,sg,3p – past)  
 k1                      k2  
 NP                      NP  
**ammAyi**                      **puswakaM**  
 (+f,sg,3 – nom)

2a.

**PaDa [wA\_hE]** => **paDawIM \_ hEM**  
 (m,sg,3)                      (f,pl,3p – past)    (pl)  
 k1                      k2  
 NP                      NP  
**ladakiyAz**                      **kiwAba**  
 (f,pl,3 – nom)

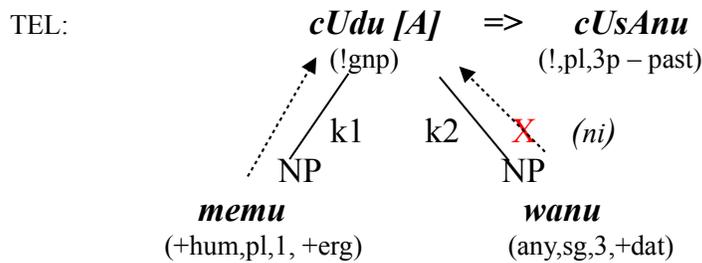
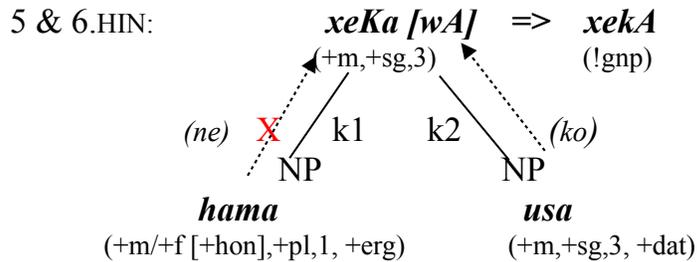
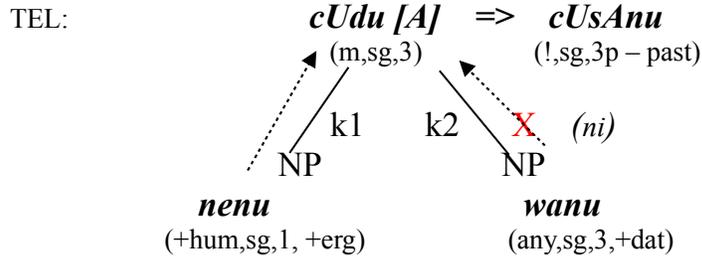
TEL:

**caxuvu [A]** => **caxuvuwAru**  
 (!gnp)                      (+hum,pl,3p – past)  
 k1                      k2  
 NP                      NP  
**ammAyilu**                      **puswakaM**  
 (f,pl,3 – nom)

3 & 4. HIN:

**xeKa [wA]** => **xekA**  
 (m,sg,3)                      (!,sg,3p – past)

(ne) **X** k1 k2 (ko)  
 NP NP  
**meM** **usa**  
 (m/f,sg,1, +erg) (m,sg,3,+dat)



## 6. Implementation of Agreement in Machine Translation System:

After the analysis of the source language text (Hindi) passes through the lexical substitution module where Hindi lexical items are substituted with 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 well formedness (faithfulness) and also increase the readability after the translation is completed. So when readability is increased the Machine translated text is very near to human translation text. Here are the Five Steps are followed to carry out the Agreement from Hindi to Telugu.

1. Find the Agent/karwa(k1) in the sentence and pass it GNP to the predicate inthe sentence.
2. Change the gender of lexicon according to the Target Language(Telugu).
3. If k1 is a pronoun (subject pronoun) then gender should be passed back, i.e from predicate to subject.

4. If Noun is followed with Adjective then its gender should be carried out in Adjective also.
5. If the Sentence is having copula as main verb then pass the gender number and person of the copula to the preceding non-verb.

Eg: rAma accA hE.

### **8. Conclusion:**

One of the major goals of this study is to demonstrate the efficacy of the design of a 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 developed at CALTS(HCU).

### **Bibliography:**

1. Boeckx, Cedric. 2008. *Aspect of the Syntax of Agreement*. New York: Routledge.
2. P. Mascia, Colin. 1993 *The Indo-Aryan Languages*. New York: University of Cambridge.
3. Agnihotri, Rama Kant. 2007. *Hindi An Essential Grammar*. New York: Routledge.
4. Krishnamurti, Bh. 1985. *A grammar of Modern Telugu*. Delhi, New York: Oxford University Press.
5. Krishnamurti Bh. 2003. *The Dravidian Languages*. Cambridge University Press, Cambridge, UK.
6. Beames, John. 1966. *A Comparative Grammar of Modern Aryan Languages of India*. Delhi.
7. Das, Pradeep Kumar 2006. *Grammatical Agreement in Hindi-Urdu and its major varieties*, Munchi, LinconEuropa.