

Verb Word Class and Kashmiri English Machine Translation Divergence

Sajad Hussain Wani*

Abstract

This paper discusses the issue of Machine Translation with reference to Kashmiri and English language pair. The focus in this paper is on the study of verb word class and the divergence caused by different subclasses of verbs in the said language pair. The divergence is illustrated with reference to the case marking of verbs and properties of verb classes including causative verbs, conjunct verbs, dative verbs and compound verbs.

Key Words

Machine Translation, Translation Divergence, Verb Types, Dative verbs, Causative verbs, Compound verbs, Conjunct verbs.

Introduction

Kashmiri and English share many similarities but the differences are much more prominent. English is an analytical language relying mainly on the word order whereas Kashmiri with its rich inflectional system has a flexible word order and has a very well developed case system. This contrast between the morphological natures of the said languages forms the root cause of divergence which is so evident for this language pair. Kashmiri language has different types of inflections for expressing a wide variety of lexical and grammatical meanings like number, tense, gender, person, case which is very much different from English language which has evolved different mechanism for expression of the phenomena which are expressed with inflections in Kashmiri language. This different expression of the same grammatical phenomena is the cause of divergence as many of the morphological mechanisms in Kashmiri are mapped on to syntactic means in English language. A MT system for Kashmiri English language pair has to be well equipped for incorporating various kinds of morphological inflections and must be able to handle the divergence when the sentences from Kashmiri are translated into English involving a mapping of Kashmiri inflections to the syntax of the English language.

* Department of Linguistics, University of Kashmir , India

Analysis

Verb forms one of the most central members of the word classes in any language including English and Kashmiri. Verbs in English usually inflect for tense. English Verbs agree with the subject as observed in 3rd person singular. However, subject-verb agreement is not all-encompassing in English as most verbs do not show any inflection for number and gender in English. Only verb *be* has more person-number distinction than other verbs. Thus, inflections in English language are comparatively reduced to a minimum. This is, in fact, true in the light of the history of English language where a transition from old to modern English has proceeded with a great reduction in the number of inflections. Tensed verbs are prototypical verbs. Verbal nouns and adjectives are usually non-tensed. This is true in case of Kashmiri as well, as a prototypical verb carries the tense inflection whereas verbal nouns and adjectives carry inflections which are untensed. E.g., *asvun* 'smiling', *natsvun* 'dancing', *asun* 'smile', *parun* 'reading' are all non tensed and are different from those verbs which carry appropriate tense inflections for present, past, future in addition to the number and gender of the subject and object NPs. Verbs have tensed forms and most verbs in English usually have six inflectional forms. Morphologically, the base form is the lexical stem, and the general present tense is identical with it. However, there are some exceptions to the general inflection paradigm. E.g., verb *be* has extra person-number forms: *was* and *were* instead of a single past tense form, *am* and *are* instead of the normal general present tense form. In contrast, the modal auxiliaries *can*, *may*, *must*, etc., have no non-tensed forms and have no contrast in the present tense between 3rd person singular and general forms.

The processes used in the formation of verb stems are fewer and typically less productive than those used in the formation of nouns. Due to inflectional richness of Kashmiri language, verbs in Kashmiri exhibit different properties. Kashmiri verbs are highly inflectional in nature and inflect for number, tense, and gender.

1.1	<i>su</i>	<i>khey-i</i>	<i>amb</i>	K
	he.nom	eat.fut.fsg	mango	
'He will eat a mango.'				

1.2	<i>su</i>	<i>chu</i>	<i>khe-va:n</i>	<i>amb</i>	K
	he.nom	is	eat.pr	mango	
'He eats a mango./ He is eating a mango.'					

1.3	<i>tAm</i>	<i>kh'-Ay</i>	<i>mazda:r</i>	<i>amb</i>	K
	he.erg	eat.pst.fsg	tasty	mango	
'He ate a tasty mango.'					

The above examples (1.1, 1.2, 1.3) illustrate how the verb is inflected for the future, present and past tense respectively. In these examples, basic verb root *khe* 'eat' is inflected with different suffixes (*-yi, -van, -Ay*) to show the different tenses.

It can be easily observed that when the verb is in the past tense, the subject takes an ergative case (example 1.3). This is one of the differences between nominative-accusative English and ergative-absolutive Kashmiri language. The ergative case in Kashmiri occurs when the verb is in the past and past perfect tense. Thus case system of Kashmiri language leads to translation divergence as ergative subject in Kashmiri has an altogether different shape from the nominative. In comparison, English makes no such distinction and both the forms are represented by a single word as shown below:

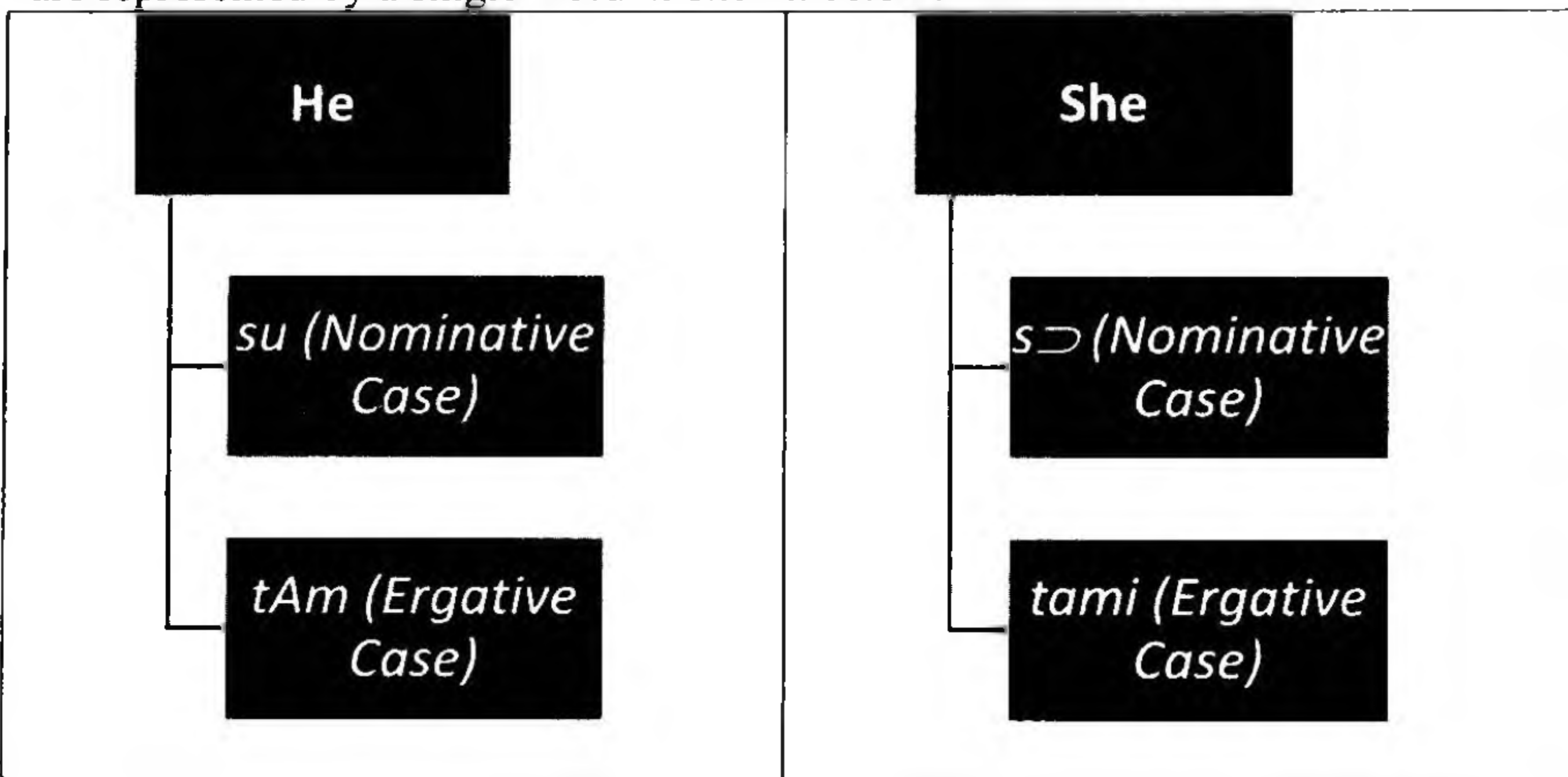


Fig. 3.1: Showing Divergence Resulting From Nominative and Ergative Case Forms of 'He' and 'She' in KEMT

Examples 1.4 and 1.5 represent the change that takes place in the verb with number. The suffixes *-yi* 'sg-fut' and *-yan* 'pl-fut' are added to the verbal root *khe* 'eat' for showing the singular and plural number.

1.4	<i>su</i> he.nom	<i>khe-yi</i> eat.fut.sg	<i>dach</i> grapes	K
'He will eat the grapes.'				
1.5	<i>tim</i>	<i>khe-yan</i>	<i>dach</i>	K

	they.nom	eat.fut.pl	grapes	
‘They will eat the grapes.’				

1.6	<i>tAm</i>	<i>kh’-Av</i>	<i>ayiskri:m</i>	K
	he.erg	eat.pst.fsg	ice cream	
‘He ate an ice cream.’				

1.7	<i>tAm</i>	<i>kh’-ov</i>	<i>Thu:l</i>	K
	he.erg	eat.pst.msg	Egg	
‘He ate an egg.’				

In the examples 1.6 and 1.7, the verb *khe* ‘eat’ takes *-Av* (feminine) and *-ov* (masculine) which directly correspond to the gender of the object *ice cream* (feminine) and *egg* ‘*Thu:l*’ (masculine). Such marking for number, gender and case is absent in English and hence translation divergence occurs at morpho-syntactic levels.

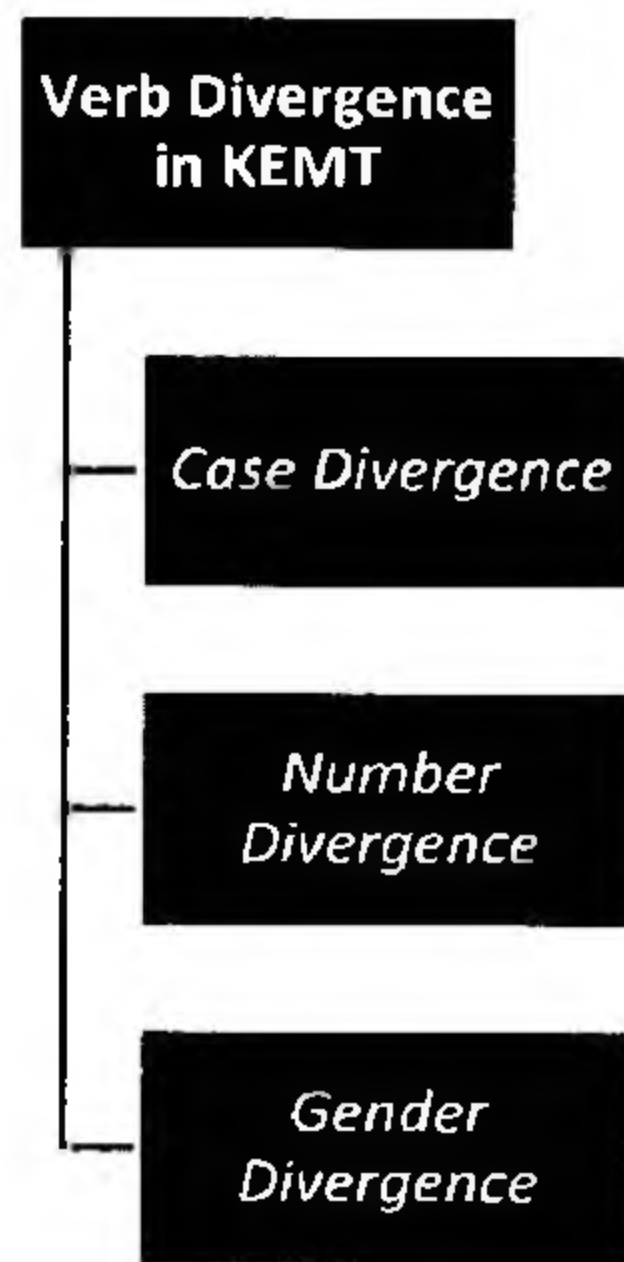


Fig. 3.2: Showing Divergence Resulting from Number, Gender and Case Marking of Verbs in KEMT

Traditionally, Kashmiri verbs are classified as intransitive, transitive, causative, dative, conjunct, or compound. Most intransitives mark their subjects in the nominative across all tenses. A few intransitives mark their subjects in the ergative in the past tense as is the case with transitives. Transitive subjects are marked nominative in the nonperfective. Direct objects are marked nominative or dative as dictated by the person hierarchy. In the perfective, transitive subjects

are marked ergative. Direct objects are in the nominative with the exception of *la:yun* 'to beat' which marks its direct object in the dative.

Kashmiri transitive verbs may be basic or derived from intransitives. Basic transitives in Kashmiri are only few which may be simple or ditransitive. All ditransitives are basic in Kashmiri. Simple transitives take a direct object. It may be marked nominative or dative depending on the context as in examples 1.8 and 1.9. In the example 1.8, the object *tsu:nT* 'apple' is marked nominative (unmarked in Kashmiri) whereas in the example 1.9, the object *Raja* is marked dative.

1.8	<i>tAm</i>	<i>kh'-ov</i>	<i>tsu:nT</i>	K
	he.erg	eat.pst.msg	apple	
'He ate an apple.'				

1.9	<i>tAm</i>	<i>lo:y</i>	<i>ra:j-as</i>	K
	he.erg	beat.pst.msg	raj.dat	
'He beat Raj.'				

Ditransitives take two objects: direct, and indirect. Direct objects follow the pattern in the simple transitive. Indirect objects are marked in the dative as in example 1.10.

1.10	<i>irsha:d-an</i>	<i>d'ut</i>	<i>shabir-as</i>	<i>tohfl</i>	K
	Irshad.erg	gave.msg	Shabir.dat	present	
'Irshad gave a present to Shabir.'					

Causative Verbs and KEMT

Causative verbs are derived from intransitives, transitives, or ditransitive verbs by changing vocalic length and suffixation. The derived causatives or causative I can be then changed into extended causatives or causativeII by further suffixation. All causatives are transitives in nature but all transitives are not causatives, as is the case with basic transitives. Causatives in Kashmiri are formed by following mechanisms:

Table 3.1: Showing Formation of Causatives by Changing Vocalic length

Intransitive	Transitive / Causative
<i>tar</i> 'cross'	<i>ta:r</i> 'make x cross'
<i>Dal</i> 'move'	<i>Da:l</i> 'make x move'

Table 3.2: Showing Formation of Causatives through Suffixation

Intransitive	Transitive / Causative
<i>ce</i> 'drink'	<i>ca:v</i> 'make x drink'
<i>ruk</i> 'stop'	<i>ruka:v</i> 'stop x'
<i>baD</i> 'grow'	<i>baDra:v</i> 'grow x'
<i>tsal</i> 'go away'	<i>tsalra:v</i> 'make x go away'

<i>as</i> 'smile'	<i>asna:v</i> 'make x smile'
<i>ruk</i> 'stop'	<i>rukna:v</i> 'make x stop'

Table 3.3: Showing Formation of Causatives through Suppletive forms

Intransitive	Transitive / Causative
<i>khas</i> 'climb'	<i>kha:r / kha:r na:v</i> 'raise x'
<i>vas</i> 'descend'	<i>va:l / va:l na:v</i> 'bring x down'
<i>daz</i> 'burn'	<i>za:l / za:l na:v</i> 'burn x'

Causal II or extended causatives are formed by suffixing Causal I forms with -*na:v* suffix as shown below

Table 3.4: Showing Formation of Extended Causatives through Suffixation

Causative I	Causative II
<i>h'a:v na:v</i> 'make x buy'	<i>h'a:v na:v na:v</i> 'make x buy through y'
<i>chal na:v</i> 'make x wash'	<i>chal na:v na:v</i> 'make x wash through y'
<i>n'a:v na:v</i> 'make x take'	<i>n'a:v na:v na:v</i> 'make x take through y'
<i>kh'a:v na:v</i> 'make x eat'	<i>kh'a:v na:v na:v</i> 'make x eat through y'

As is obvious from the above examples, translation of causatives from Kashmiri to English involves morpho-syntactic divergence. English has a syntactic mechanism for the expression of causativisation but in Kashmiri, same is expressed through different suffixes (morphological mechanisms) as shown above.

The existence of extended causatives or Causative II further increases the divergence in KEMT as doubled causative suffixes demand more words in English. Thus causatives involve a translation divergence where suffixes (morphological level) in Kashmiri are mapped on to syntactic constituents (syntactic level) in English and hence can be labelled as an instance of inflational divergence.

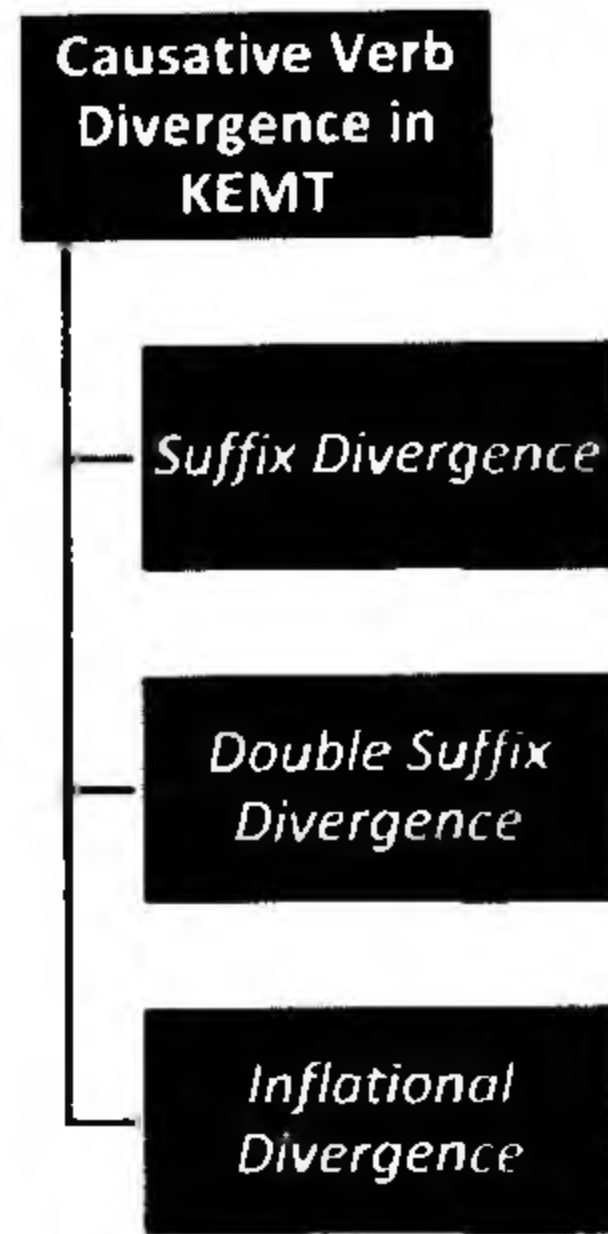


Fig. 3.3: Showing Divergence Resulting from Causatives Verbs in KEMT

Kashmiri has a variety of other verb types which have no corresponding equivalents in English and translation of these different verb types result in different types of morpho-syntactic divergences as follows:

Ingestive Verbs, Dative Verbs and KEMT Divergence

Verbs like *khe* 'eat', *ce* 'drink', *vuch* 'see', *par* 'read' have been classified as ingestive or affective verbs in traditional grammars. The case of the causee of an affective verb is either nominative or dative and is never oblique (Wali and Koul, 1997). The case of the causee of a non-affective causative may be oblique or dative. Dative or psyche verbs in Kashmiri have distinction of marking their subject as dative in all tenses and aspects. The thematic object is marked nominative and is in agreement with the verb.

Dative or Psyche verbs include verbs of perception, knowledge, belief system, and different physical and mental states. Verbs like *banun* 'ability to get something', *ba:sun* 'having a feeling', *nanun* 'coming to know', *tagun* 'skill to perform', *yun* 'skill to perform', *lagun* 'to be suitable', *tsetas py'on* 'to remember', *isa:lun* 'to bear/tolerate', *za:nun* 'to understand', *ma:nun* 'come to agree', etc.

The classification of ingestive and dative verbs is based on the case marking and case marking distinctions are not relevant for English verbs. Hence, the type of divergence which is observed while translating ingestive verbs and dative verbs in English can be labelled as case divergence. Many Dative or psyche verbs also exhibit Inflational divergence as these are translated with the help of more than one word in English.

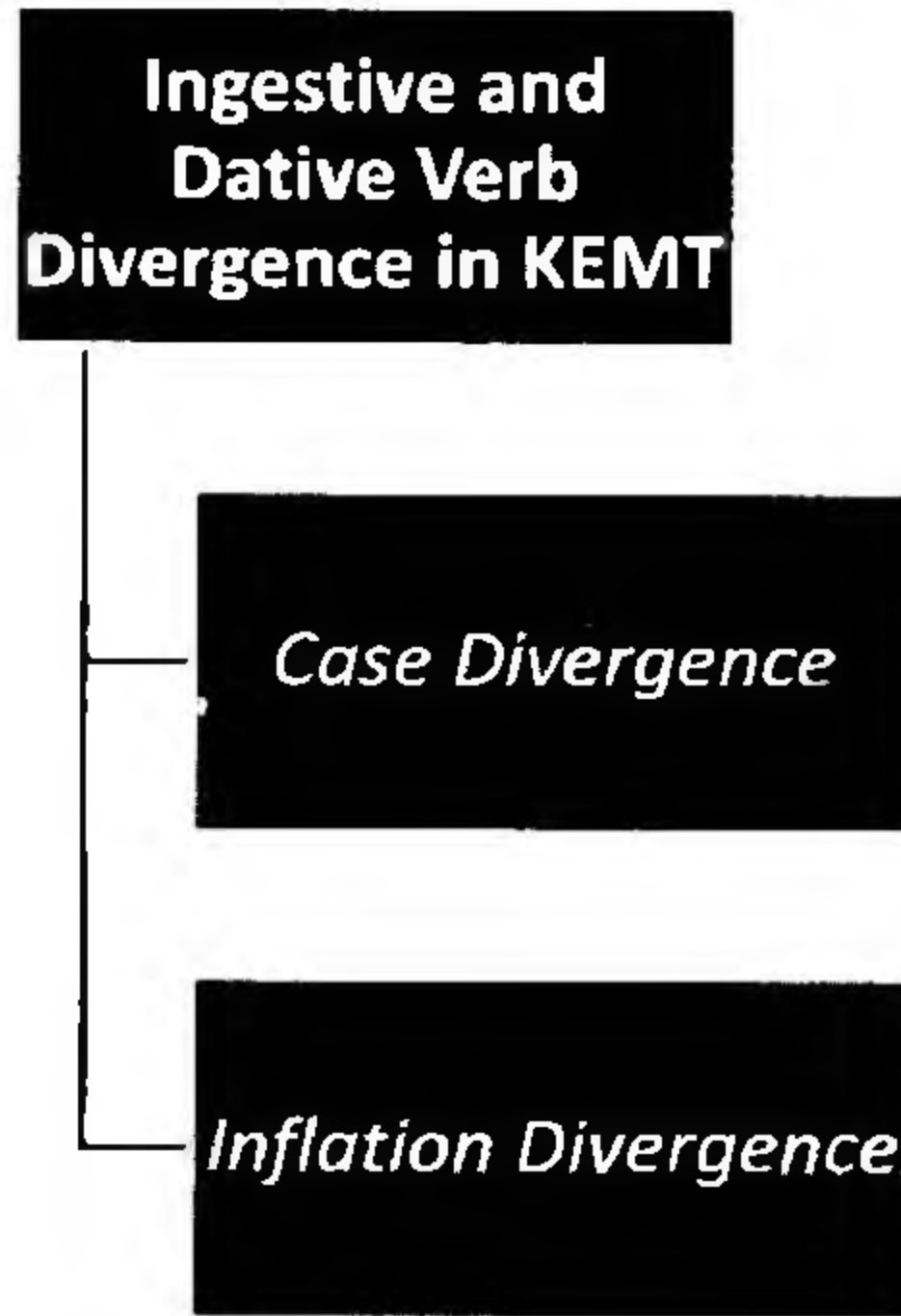


Fig. 3.4: Showing Divergence Resulting from Ingestive and Dative Verbs in KEMT

Conjunct Verbs and KEMT Divergence

Conjunct verbs are derived by adding a certain set of verbs to nouns, adjectives and adverbs. This set of verbs are often referred to as vectors. Vectors like *a:sun* 'to be', *ro:zun* 'to stay', *khasun* 'to rise'/'to increase'/'to climb', *d'un* 'to give', *karun* 'to do', *yun* 'to come', *p'on* 'to fall', etc belong to this group.

Table 3.5: Showing Formation of Conjunct Verbs in Kashmiri from Nouns and Vectors

Noun	Vector	Conjunct Verb
<i>na:d</i> 'call'	<i>d'un</i> 'give.inf.msg'	<i>na:d d'un</i> 'to call'
<i>aza:b</i> 'suffering'	<i>gatchun</i> 'happen.inf.msg'	<i>aza:b gatchun</i> 'to suffer'
<i>ya:d</i> 'memory'	<i>ro:zun</i> 'remain.inf.msg'	<i>ya:d ro:zun</i> 'to remember'

Table 3.6: Showing Formation of Conjunct Verbs in Kashmiri from Adjectives and Vectors

Adjective	Vector	Conjunct Verb
<i>thi:kh</i> 'correct'	<i>karun</i> 'do.inf.msg'	<i>thi:kh karun</i> 'to cure'

<i>rut</i> 'beautiful'	<i>bana:vun</i> 'make.inf.msg'	<i>rut bana:vun</i> 'to beautify'
<i>mlklr</i> 'dirty'	<i>karIn</i> 'do.inf.fsg'	<i>mlklr karIn</i> 'to pollute'

Vectors carry the inflections and the associated noun, adjective and adverb remains unaffected. Thus in case of most conjunct verbs a conflation divergence is observed as a conjunct verb in Kashmiri is translated into a simple verb in English. Besides, the inflections on the vector lead to number, gender and person divergence.

Compound Verbs and KEMT Divergence

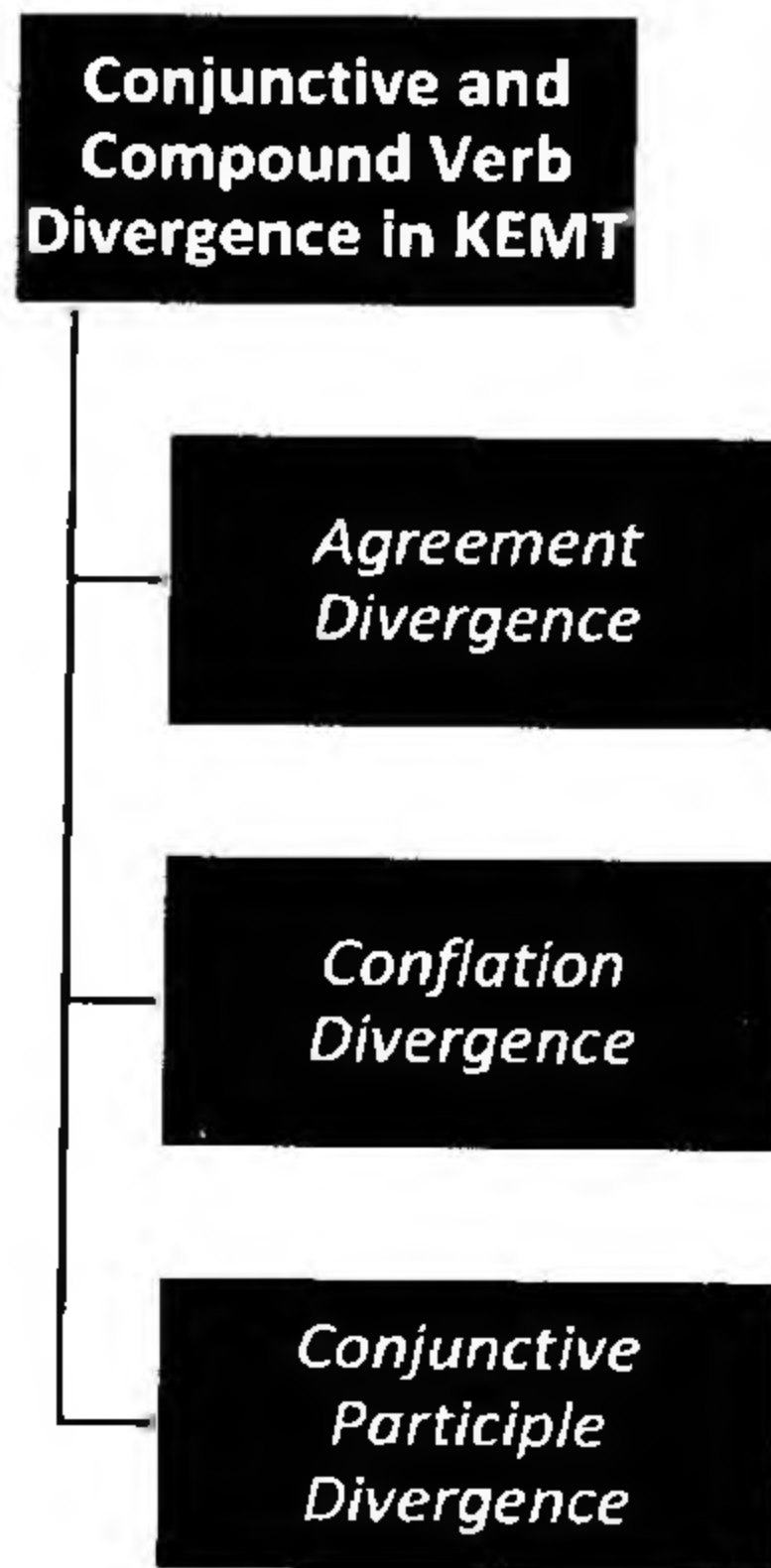
Compound verbs are formed by a combination of the conjunctive participle of the main verb with with certain auxiliary verbs. These auxiliaries are called explicators / operators. The list of explicators varies with the authors (Grierson 1911, Kachru 1969, Hook and Koul 1992).

The explicators have undergone a semantic shift as they do not possess their original meanings. These explicators/operators add certain aspectual values such as completion of an action, benefaction, or intensification, to the main verb. Compound verbs in Kashmiri are replaced by simple verbs much like conjunct verbs and hence can be regarded as the instances of conflation divergence where a combination of conjunctive participle of a main verb and operator is replaced by a simple verb in English .

Table 3.7: Showing Formation of Compound Verbs in Kashmiri

Conjunctive participle of the Verb	Operator	Compound Verb
<i>kh'e-ith</i> 'eat.cp'	<i>n'un</i> 'to snatch' / 'to take'	<i>kheth n'un</i> 'to eat'
<i>rAn-ith</i> 'cook.cp'	<i>tha:vun</i> 'to keep'	<i>rAnith tha:vun</i> 'to cook'
<i>dA:r-ith</i> 'throw.cp'	<i>din</i> 'to give'	<i>dA:rith din</i> 'to throw'

Fig. 3.5: Showing Divergence Resulting from Conjunctive and Compound Verbs in KEMT



Conclusion

Verbs in Kashmiri exhibit a much more complicated morphological behaviour and complex rules are observed for different agreement patterns. Verbs change their form and have different cases in different tenses. Kashmiri language is rich in terms of having a wide variety of compound, conjunct and dative verbs which differ in behaviour from the corresponding simple English verbs. The complex behaviour of these wide varieties of verbs along with their agreement and conjugation patterns has to be a part of Kashmiri English MT system. The divergence patterns outlined in this paper are intended to serve as a theoretical construct for the development of any MT system in the future by providing it with sound linguistic principles to increase the accuracy and proficiency of its output. The divergence patterns outlined here can be utilized for further advanced research in this direction.

List of abbreviations:

1p = 1st person.	2p = Second	3p = 3 rd Person.	nom = Nominative case
Gen = Genitive case.	erg = Ergative case.	dat = Dative case	pst = Past tense.
fut = Future tense	sng = Singular	pl = Plural.	fem = Feminine.
abl= Ablative case	inf = infinitive	K = Kashmiri	' = Platalization
int = interrogative	hon = honorific		

References

- Arnold, D. J. 1994. *An Introduction to Machine Translation*. Oxford: Blackwell.
- Bhat, Roopkrishen. 1987. *A Descriptive Study of Kashmiri*. Delhi: Amar Prakashan.
- Dorr, B. J. 1993. *Machine Translation: A View from the Lexicon*. Cambridge, MA: MIT Press.
- Dorr, B. J. 1994. Classification of Machine Translation Divergences and a Proposed Solution. *Computational Linguistics* 20(4):597-633.
- Dorr, B. J., Jordan, P. W. and Benoit, J. W. 1998. A Survey of Current Paradigms in Machine Translation. *Technical Report LAMP-TR-027, UMIACS-TR-98-72, CS-TR-3961*. University of Maryland, College Park: USA.
- Goddard, Cliff. 1998. *Semantic Analysis. A Practical Introduction*. Oxford: Oxford University Press.
- Gupta, D. and Chatterjee, N. 2002. Study of Similarity and its Measurement for English to Hindi EBMT. *Proceedings of STRANS-2002*: IIT Kanpur.
- Gupta, D. and Chatterjee, N. 2003b. Identification of Divergence for English to Hindi EBMT. *Proceedings of the MT SUMMIT IX, Orleans, LA*. 141-148.
- Huddleston, Rodney. 1988. *English Grammar: An Outline*. Cambridge: Cambridge University Press
- Jakobson, Roman. 1959. On Linguistic Aspects of Translation. *Translation Studies Reader. (2nd Edition)*. L. Venuti. New York: Routledge. 113-119.
- Kachru, Braj B. 1969 a. *A Reference Grammar of Kashmiri*. Urbana: University of Illinois.
- Kak, A. A . and S.H.Wani. 2006. Strategies of Neutrality and Code Mixing Grammar. In K. Rai, etal(ed.). *Recent Studies in Nepalese Linguistics. Proceedings of the 12th Himalyan Languages Symposium & 27th annual Conference of Linguistic Society of Nepal, November 26-28, 2006*. LSN: Nepal. 467-487.
- , 2009. Code Switching, Code Mixing and Teaching Process. In Manish A. Vyas and Yogeh L. Patel(etal). *Teaching English as a Second Language: A New Pedagogy for a New Century*. New Delhi: PHI Learning Private Limited. pp. 309-318.
- , 2009. An Evaluation of Kashmiri English Code Mixing in Myers Scotton Models with Special Reference to MLF and 4M Model. In A. R. Fatihi

- (ed.) *Language Vitality in South Asia*. Aligarh: Department of Linguistics, AMU. 227-237.
- Koul, O. N. and Wali, kashi. 2009. *Modern Kashmiri Grammar*. Delhi: Indian Institute of Language Studies.
- Leech, Geoffrey.1971. *Meaning and the English Verb*. London: Longman.
- Wani,S.H. 2004. Formal and Functional Aspects of English-Kashmiri Code Mixing. Diss. University of Kashmir.
- , 2009. Code Mixing Constraints from Grammatical to the Minimalist Framework: A Study in Kashmiri-English Mixed Code. *In Asian EFL Journal*.
- ,2013. Divergence Patterns in Kashmiri –English Machine Translation: A View from Translation of Tenses. *Interdisciplinary Journal of Linguistics, IJL vol.VI.p.201-222*.
- ,2013.*Kashmiri to English Machine Translation: A Study of Morpho-Syntactic Divergence Patterns*. Diss. University of Kashmir.

