Kashmiri to English Machine Translation: A Study in Translation Divergence Issues of Personal and Possessive Pronouns

Sajad Hussain Wani [a,*]

[a] Department of Linguistics, University of Kashmir, Srinagar, Jammu and Kashmir 190006, India
*Corresponding author Email: sajadw@uok.edu.in ; wani.sajad@gmail.com

Abstract: Machine translation (MT) as a sub-field of computational linguistics represents one of the most advanced and applied translation dimensions as a research field. Translation divergence occurs when structurally similar sentences of the source language do not translate into sentences that are similar in structure in the target language" (Dorr, 1993). The sophistication in the domain of MT depends mainly on the identification of divergence patterns in a language pair. Many researchers in MT field including Dorr (1990, 1994) have emphasized that the best quality in MT can be achieved when an individual language pair in a particular context is described in detail. This paper attempts to explore the divergence patterns that characterize the translation of Kashmiri pronouns into English. The analysis in this paper has been restricted to the class of personal and possessive pronouns. Kashmiri has rich inflections and pronouns are marked for case, number, tense and gender and show complex agreement patterns. The paper identifies and outlines a wide variety of divergence patterns that characterize the Kashmiri English language pair. These divergence patterns are identified and summarized in order to improve the quality of the MT system that may be developed for Kashmiri English language pair in the near future and can also be utilized for other language pairs that are similar in terms of their structure and typological features.

Keywords: Epic, Silappathikaram, Structure, Anthology, Contribution

Language: Tamil

About the Author

Dr Sajad Wani has a teaching experience of more than thirteen years at the department of Linguistics, University of Kashmir besides teaching at many other departments including IGNOU and Directorate of distance Education. Dr. Sajad Hussain Wani holds two doctorate degrees. He worked on *Kashmiri-English Machine Translation* for his PhD degree in Linguistics and his topic for PhD in English literature was on *Mysticism in Shamas Faqir and Ralph Waldo Emerson*. He has published a total of 30 papers in reputed and peer reviewed journals and has presented an equal number of papers in different national and international seminars. His M.Phil thesis on *Code Mixing Constraints* finds its place in international bibliography entitled "*Error Analysis in the World: A Bibliography*" compiled by Prof. Bernd Spillner in 2017. Dr. Sajad Wani has worked in many national and international projects with world reputed institutes like *NFLC, University of Maryland* and *C-DAC, Pune.* His research interests include Phonology, ELT, Stylistics and Computational Linguistics. He is also the Principal Investigator (PI) of a major research project on Language Contact in Kashmir which is sponsored by MHRD, Govt of India under RUSA 2.0 component in addition of being a *co-PI in another major research project* on developing an online Kashmiri-English-Hindi Trilingual dictionary. He is actively involved in the translation of software tools in Kashmiri language being developed by *CDAC,* Pune.
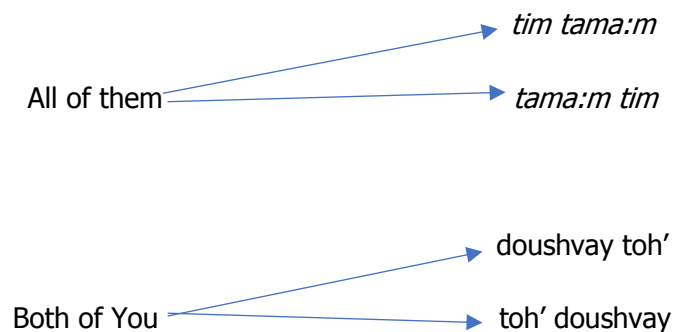
Introduction

Pronoun word class is one of the basic word classes and make communication very effective by their anaphoric and deictic function. Pronouns consist of many subtypes that are used in day-to-day speech. Like other noun phrases, Pronouns act as heads and function as the subject, object, and complement of a proposition/postposition. Pronouns act as heads like nouns but are not as flexible as nouns in taking various dependents. Pronouns substitute nouns and hence these are anaphoric in their functioning, which is their central function. Deixis can also be identified as a central property of pronoun class. English pronouns are special in the sense that they show more case forms when compared to other word classes. The person is a deictic category in pronouns and is involved in subject-verb agreement. Pronouns in English also vary as per the gender of the person involved, as is clear from pronouns like he, she and it. English does not show any rule for the agreement of gender and ordinary nouns have no category to apply for showing the gender.

One of the primary differences that cause divergence in Kashmiri English language pair is the rational nature of English. In contrast, Kashmiri is inflectional in nature and this is reflected in the pronominal system of the language as well. Wani (2013b) states that the freedom with which determiners combine with pronouns in Kashmiri is different from how English determiners combine with English pronouns. According to Wani (2017); divergence in Kashmiri English language pair occurs due to many structural differences between Kashmiri and English languages.

Determiners in English always precede the pronouns in English as in expressions like "All of them" and "both of you" have a fixed word order. In contrast, the translation in Kashmiri can place the determiner "tama:m"(all) and "doushvay" (both) before or after the pronoun which results in movement divergence as shown below:

```
                                      ┌──────►  tim tama:m
        All of them ──────────────────┤
                                      └──────►  tama:m tim


                                      ┌──────►  doushvay toh'
        Both of You ──────────────────┤
                                      └──────►  toh' doushvay
```

The first divergence pattern observed in the translation of pronouns enlisted here is movement divergence, which results from freedom of movement in one language, whereas such a movement is impossible in other language, as is the case with the English language.

English pronouns show a gender variation as is evident from pronouns like he, she, him, her, his but gender is of little grammatical significance compared to Kashmiri language which shows complex patterning concerning gender as gender agreement is a complex phenomenon in the Kashmiri language. This complex patterning of gender in the Kashmiri language compared to the English language results in gender agreement divergence, which will be discussed below.

Pronouns have many subclasses that show various types of properties and thus, each subclass of pronouns needs to be analyzed separately for the study of divergence patterns. In this paper, the focus of the study will be on the analysis of personal and possessive pronouns and the type of divergences that results when translating personal and possessive nouns from Kashmiri to English (Figure 1).
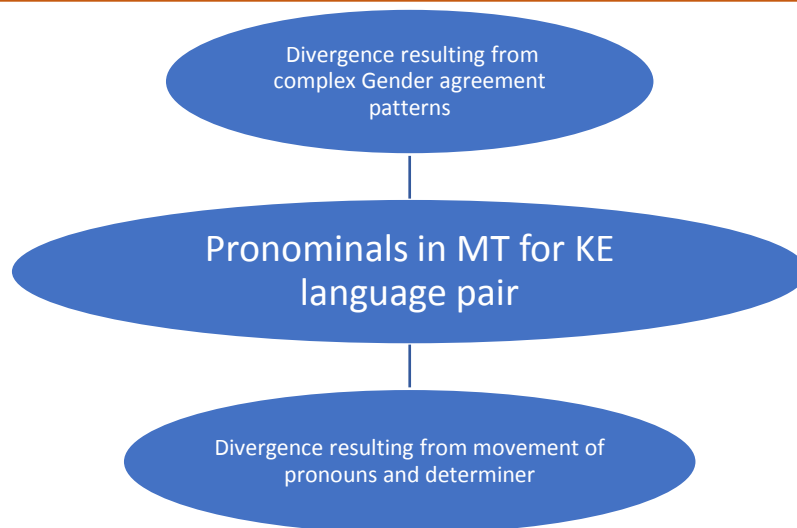
Figure 1 Showing Pronouns with divergence resulting from complex agreement patterns of gender and movement of pronouns and determiners**.**

Personal Pronouns and KEMT Divergence

English pronouns have retained many case forms compared to other word classes in the English language, which have become more analytic in the transition from Old English to Modern English. English personal pronouns show nominative case forms like *I, we, you, he, she,* whereas in accusative case, they have forms like *me, us, him, her, them.* However, the Kashmiri pronominal system is more complex and different forms of pronouns occur for the first, second and third persons inflected for marking gender, number, and case leading to multiple divergences of personal pronouns for Kashmiri English language pair.

The first and second person pronouns do not show any gender distinction in both English and Kashmiri languages. However, in case of third person personal pronouns; a three-term distinction is observed in English as is clear from forms like *he, she* and *it*. In case of Kashmiri language only a two-term distinction is found in third person pronouns. However, third person personal pronouns in Kashmiri show different forms in terms of their distancing in Kashmiri. Kashmiri third person personal pronouns have proximate (proximate), remote I (rem I, with in the sight) and remote II (rem II, out of the sight) forms which are not found in English language and hence leading to multiple divergence as shown in the below tables. Pronouns have more case forms than the nouns and therefore multiple divergence is expected naturally where one pronominal form in English corresponds to many pronominal forms in Kashmir and this has to be a part of any KEMT (Kashmiri English Machine Translation) system that will be developed in future.

The below given tables represent multiple divergences in translating Kashmiri pronouns into English into a tabulated form that can be used for developing a KEMT system (Tables A-D).

A single personal pronoun in English has many equivalent case forms in Kashmiri. It is easily observed that personal pronouns in English show a ratio of 6:15 in nominative case i.e., 6 personal pronouns of English have 15 equivalents in Kashmiri. It is clear that second person personal pronoun "you" has two different equivalents "tsI" and "toh'" in Kashmiri causing multiple divergence. It is clear from the below table that multiple divergence also results in translating English third person personal pronouns "he", "she" and "they" which have eleven (11) equivalents in Kashmiri accounting for third person pronouns in proximity, with in sight and out of sight. The divergence resulting from different case forms in dative and ablative cases is described below, which shows a ratio of 6:10.

Table A Showing Case Forms of Personal pronouns in Nominative Case for English Kashmiri Language Pair (6:15)

Person	masng	maplu	fesng	feplu
1st person	bI (I)	As' (we)	bI (I)	As' (we)
2nd person	tsI (you)	toh' (you)	tsI (you)	toh' (you)
3rd person (proximate)	yi (he)	yim (they)	yi (she)	yimI (they)
3rd person (remote I)	hu (he)	hum (they)	hO (she)	humI (they)
3rd person (remote II)	su (he)	tim (they)	sO (she)	timI (they)

Table B Showing Case Forms of Personal Pronouns in Dative Case for English Kashmiri Language Pair (6:10)

Person	masng	maplu	fesng	feplu
1st person	m'e (me)	asi (us)	m'e (me)	asi (us)
2nd person	tse (you)	tOhi (you)	tse (you)	tOhi (you)
3rd person (proximate)	yemis (him)	yiman (them)	yemis (her)	yiman (them)
3rd person (remote I)	homis (him)	human (them)	homis (her)	human (them)
3rd person (remote II)	tAmis (him)	timan (them)	tAmis (her)	timan (them)

Table C Showing Case Forms of Personal pronouns in Ablative Case for English Kashmiri Language Pair (6:10)

Person	masng	maplu	fesng	feplu
1st person	m'e (me)	asi (us)	m'e (me)	asi (us)
2nd person	tse (you)	tOhi (you)	tse (you)	tOhi (you)
3rd person (proximate)	yemi (him)	yimav (them)	yemi (her)	yimav (them)
3rd person (remote I)	homi (him)	humav (them)	homi (her)	humav (them)
3rd person (remote II)	tami (him)	timav (them)	tami (her)	timav (them)

Table D Showing Case Forms of Personal pronouns in Ergative Case for English Kashmiri Language Pair (6:13)

Person	masg	mapl	fesg	fepl
1st Person	m'e (I)	asi (we)	m'e (I)	asi (we)
2nd person	tse (you)	tOhi (you)	tse (you)	tOhi (you)
3rd person (proximate)	yem' (he)	yimav (they)	yemi (she)	yimav (they)
3rd person (remote I)	hom' (he)	humav (they)	homi (she)	humav (they)
3rd person (remote II)	tAm' (he)	timav (they)	tami (she)	timav (they)

The personal pronouns show a ratio of 6:13 in ergative case. This is a clear example of one to many or multiple divergence.

First, second- and third-person plural forms are used as honorifics when used for substituting first, second and third person singular forms resulting in honorific divergence. The following types of divergences can be outlined for personal pronouns in KEMT (Figure 2).
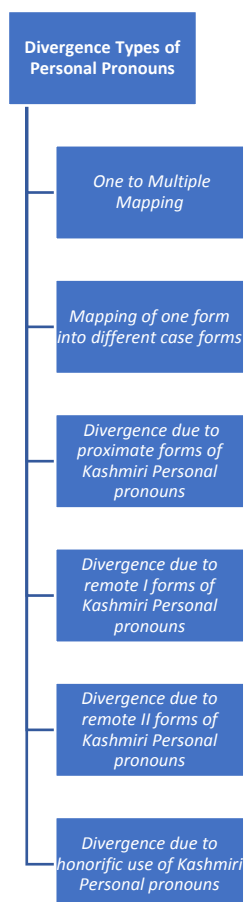

Figure 2 Different types of KEMT Divergence in various Personal Pronoun types

Divergence in KEMT due to Difference in Possessive Pronouns

Possessive pronouns in English include pronouns like *my, our, your, mine*. As discussed earlier, Kashmiri pronoun forms are inflected for person, gender, and case, which applies to possessive pronouns (Table E). This results in case and multiple divergence as a possessive pronoun in English has many case forms in Kashmiri. Kashmiri possessive pronouns agree with complement/modifier in an elaborate manner whereas such an agreement is not observed in English and this results in agreement divergence.

Another kind of multiple divergences that can be observed in the translation of possessive pronouns in KEMT is related to the functional potential of possessive pronouns. M*y, our, your, her, their* in English are called as possessive attributive pronouns and act as pre-head modifiers and are used attributively. Possessive pronouns like *mine, ours, yours, hers, theirs* are possessive absolute and occur as the predicative complements and are also called as predicative possessives.

Kashmiri does not recognize this difference and both possessive attributive pronouns and predicative possessive pronouns have a single translation equivalent in Kashmiri, leading to reverse multiple divergence. The following table represent the forms of personal possessive pronouns agreeing with complement/modifier in terms of number, gender and person in Kashmiri.

Table E Showing Forms of Personal Possessive Pronouns

Person	masng	maplu	fesng	feplu
1st person.sng	m'o:n (mine)	m'e:n (mine)	m'e:n (mine)	m'a:ni (mine)
1st person. plu	so:n (our)	sA:n' (our)	sA:n' (our)	sa:ni (our)
2nd person. sng	co:n (your)	cA:n' (your)	cA:n' (your)	ca:ni (your)
2nd person. plu	tuhund (your)	tuhInd' (your)	tuhInz (your)	tuhInzI (your)
3rd person .sng.proximate	yem'sund (his)	yem'sInd' (his)	yem'sInz (her)	yem'sInzI (her)
3rd person .plu.proximate	yihund (their)	yihInd' (their)	yihInz (their)	yihanzI (their)
3rd person .sng. remI	hom'sund (his)	hom'sInd' (his)	hom'sInz (her)	hom'sInzI (her)
3rd person .plu.remoteI	huhund (their)	huhInd' (their)	huhInz (their)	huhInzI (their)
3rd person .sng.remoteII	tAm'sund (his)	tAm'sInd' (his)	tAm'sInz (her)	tAm'sInzI (her)
3rd person .plu.remote II	tihund (their)	tihInd' (their)	tihInz (their)	tihInzI (their)

As is clear from the above table, possessive pronouns in KEMT represent an affluent area of multiple divergence as 7 personal pronouns in English have about 37 forms in Kashmiri inflecting for the distinction of proximate, remote, gender, number and person of the modifier/complement (Figure 3). Thus a ratio of 7:37 is observed in the above tables. This multiple divergence in possessive pronouns can be further expanded when consideration would be given to different case forms and honorifics as employed in Kashmiri, which in turn will lead to a higher ratio of multiple divergence (Wani 2013; Wani 2014).
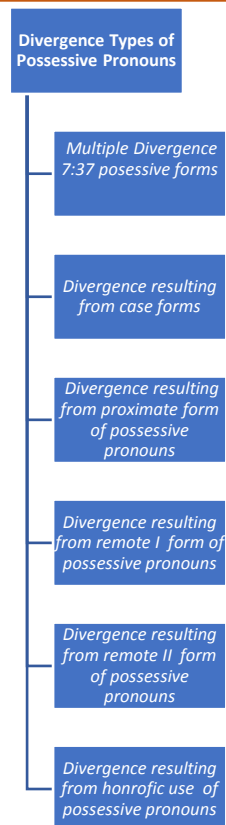
Figure 3 Different types of KEMT Divergence in various Possessive Pronouns

Conclusion

This paper presented a detailed overview of various divergence types that one encounters while translating personal and possessive pronouns from Kashmiri to the English language. It was made clear that the biggest question in translating these pronoun types is the issue of multiple divergence where a single English pronoun is mapped on to different forms in Kashmiri. Kashmiri has numerous equivalents for a single form in English. The change happens as Kashmiri pronouns show inflections for different cases, tenses, numbers and even gender, as evident from the above discussion. Inverse multiple divergence is observed where her and him are translated by a single form. The concept of case divergence, multiple divergence and gender divergence has also been demonstrated in other works on Kashmiri English machine translation divergence with respect to verb word class and translation across different tenses. Any MT system that may be developed in the future has to take this linguistic divergence in its base component. Whether one develops a hybrid system or example based machine translation system, this knowledge of linguistic divergence patterns will enrich the proficiency and efficiency of a KE Machine translation system.

List of Abbreviations

KEMT   -   *Kashmiri English Machine Translation*

masng   -   *Masculine singular*

maplu   -   *Masculine plural*

fesng   -   *Feminine singular*

feplu   -   *Feminine plural*

References

Dorr, B. J. (1990). Solving thematic divergences in machine translation. Proceedings of the 28th annual meeting on association for computational linguistics, 127-134. Pittsburgh, PA: University of Pittsburgh, 1990.127-134.

Dorr, B. J. (1993). Machine Translation: A View from the Lexicon. MIT Press: Cambridge.

Dorr, B.J. (1994). Machine translation divergences: A formal description and proposed solution. Computational linguistics 20(4), 597-633.

Wani, S. H. (2013). Divergence patterns in Kashmiri –English machine translation: A view from translation of tenses. Interdisciplinary journal of linguistics, Vol.VI, 201-222.

Wani, S. H.   (2013b). Kashmiri to English machine translation: A study of morpho-syntactic divergence patterns [Doctoral Dissertation, University of Kashmir].

Wani, S. H. (2014). Verb Word Class and Kashmiri –English Machine Translation Divergence. Interdisciplinary Journal of Linguistics, Vol.VII, 275-286.

Wani, S. H. (2017). Divergence Patterns for Kashmiri-English Machine Translation: Revisiting Dorr's Classification. Interdisciplinary journal of linguistics, Vol.X, 104-113.

Funding

Does this article screened for similarity?

Yes

Conflict of interest

The Author has no conflicts of interest to declare that they are relevant to the content of this article.

About the License

Cite this Article

Sajad Hussain Wani, Kashmiri to English Machine Translation: A Study in Translation Divergence Issues of Personal and Possessive Pronouns, Indian Journal of Multilingual Research and Development, Vol 2, Iss 1 (2021) 1-9. DOI: https://doi.org/10.34256/ijmrd2111