

Divergence patterns for Urdu to English and English to Urdu Translation.

Shukla, V. (1) and R.M.K. Sinha (2)

(1) CDAC, Noida, India

(2) IIT, Kanpur, India

1 Introduction

Translation divergence is said to occur when two languages differ in their grammars. Thus divergence would occur when a sentence in the source language translates to a sentence in the target language in a very different manner. Divergence for English to Hindi and Hindi to English has been studied to quite an extent. The existing literatures are: [Dave et al, 2001]; [Gupta et al, 2001, 2003] and [Sinha et al, 2005a, 2005b].

Divergence for English to Urdu or Vice versa has not been explored so far barring an exception [Saboor et al, 2010] which has studied the divergence patterns for Urdu to English and not vice versa. Thus here we have tried to identify the various divergence patterns that exist for English to Urdu as well as Urdu to English.

The study is based on the findings of [[Sinha et al, 2005a, 2005b] for Hindi. Hindi and Urdu are structurally very similar. They use similar postpositions, verb morphology as well as complex predicate verb structure [9]. The broad categories of divergence are still based on Dorr classification [Dorr, 1994].

The types of divergences that have been considered are Promotional /Demotional, Structural, Lexical, Categorical, Conflational/ Inflational and Thematic Divergence. There can be many other types of divergences other than those mentioned above and these could occur due to use of Reduplication, Honorific usage of words and omission of a subject in a language.

Divergence Patterns:

A. Conflational and Inflational Divergence

A conflational divergence results when two or more words in one language are translated by one word in another language. Let us illustrate this with a Urdu sentence.

(U)- voh khaana nosh farmaa rahe hein.

The English translation for this Urdu sentence would be:

(E)- They are eating.

In this example, the verb “nosh farmaa rahe” of Urdu sentence is equivalent to one word verb of English (i.e. eat) upon translation. Another example that can be considered is:

(U)- Baraaye Maherbaani aap tashreef le jaayen.

(E)- Please go.

Here the word “please” is being referred to as “baraaye maharbaani” in urdu and the verb “go” is referred as “tashreef le jaayen” and also in the English sentence there is no mapping for the word that has been translated to “aap”.

The examples discussed above show Conflational divergence. The opposite case would be when one word in the source language is realized by two or more words in the target language and this is referred to as inflational divergence. We can illustrate this with few examples:

(U)-loo chal rahii hai.

(E)-Hot wind is blowing.

In this example, the word *loo* of Urdu sentence requires two words of English (i.e. hot wind) upon translation. If we consider the case for English to Urdu translation, we find that inflational divergence is very common.

(E)- One moment please!

(U)- Baraaye Maherbaani Kuch Der Intazaar Kijiye

In this example “one moment” has been translated to “Kuch Deyr Intizar Kijiye” and please as “baraaye maharbaani”. Another example that would illustrate the inflational divergence is:

(E)- Welcome!
(U)- khush aamdeed

B. Structural Divergence

Structural divergence occurs where an NP argument in one language is realized by a PP adjunct/oblique NP in another language. For example,

(U)-Mere valid lucknow ja rahe hein
(E)-My father is going to Lucknow.

In this example, *lucknow*, the Noun Phrase in the Urdu sentence is converted into PP “*to Lucknow*” in English sentence upon translation. If we consider the case of English to Urdu translation then the following English sentence “He entered the room” which is translated to “*vah kamare mein daakhil huaa*”, the verb ‘*enter*’ in English sentence takes an NP argument ‘*the room*’ whereas its Urdu counterpart “*daakhil honaa*” takes a PP adjunct *kamare mein*.

C. Promotional and Demotional Divergence

As mentioned by [Sinha et al, 2005a] Promotional and demotional divergences or Head-swapping divergences arise where the status (lower or higher) of a syntactic constituent in one language is affected in another language. For instance, when an adverbial element in one language is realized by a verbal element, it constitutes a case of promotional divergence and an opposite case will result in demotional divergence. In demotional divergence the role of the main verb of the source language sentence is demoted upon translation. Some examples of demotional divergence in Urdu-to-English translation are discussed below:

In the examples discussed below, the word realized as a main verb in source language (Urdu) is realized as an Adverbial Modifier or as an adjective in the target language (English). For example,

(U)- sangeet baj rahaa hai
(E)-The music is on.

Here the main verb in Urdu sentences “*baj raha* ” is realized as an Adverbial Modifier “*on*” in English sentences upon translation. If we

consider an English sentence “Life goes on” then it is translated to Urdu as – “*Zindagi chalti rahti hai*”.

Here “on” is a particle in English sentence and this is realized an auxiliary verb “rahti hai” in Urdu sentence upon translation.

(U)-Hum thak kar chur ho gaye

(E)-We were dead tired.

In this example, the verb of Urdu sentence “*thak kar chur hona*” is realized as an Adjective “*dead tired*” in English sentence upon translation. If we consider English to Urdu case then the example that can be considered for illustration of demotional divergence is:

(E)- It suffices.

There are two possible translations for the above English sentence:

(U1)- yah kaafi hai.

(U2)- yah zaroorat ke mutaabik hai.

The word “suffice” is realized as the main verb in English but as an adverbial modifier *kaafi hai/ zaroorat ke mutaabik hai* in Urdu. Here the role of suffice (verb) in English sentence is being demoted in Urdu sentence and realizes as an adverbial modifier.

D. Lexical Divergence

Lexical divergence arises out of the unavailability of an exact translation map for a construction in one language into another language. It also means that the choice taken for the target language word is not the literal translation for an English word. Consider the following English sentences:

(E)- Good luck!

(U)- Allah Ka Fazal Ho

The words used in Urdu translation are not the literal meanings of the English words in the source sentence. Another English sentence that can be considered is:

(E)- It is cold.

(U)- Maahaul Sard Hai

In this translation the word *Maahaul* in Urdu does not have a corresponding map in the English sentence.

(E)- Excuse me!... (to ask for something)

(U)- Maazirat Chaahta Hoon

Here the English sentence verb “excuse” has been translated as a different verb meaning “want”- chaahta hoon. If we consider the following Urdu sentence to be translated to English

(U)-Vah chayan ki niid so rahaa hai

(E)-He is enjoying a sound sleep.

In this example, the verb of Urdu sentence “*so rahaa hai*” is converted into a different verb “*enjoying*” in English sentence upon translation. The verb enjoying as such has no word whose literal meaning exists in the urdu translated sentence.

E. Categorical Divergence

Categorical divergences are located in the sentences where there is a mismatch between parts of speech of the pair of translation languages. It is observed that this is the most common type of divergence that occurs in any pair of language. Since it is concerned with the POS of source language and target language, this type of divergence arises if the lexical category of a word changes during the translation process. Let us illustrate this category of divergence with certain examples. Let us consider the following English sentence:

(E)- She is jealous of me

This sentence can be translated in Urdu in two ways:

(U1)- vah mujhse jalti hai.

(U2)- usako mujhse jalan hai

In English sentence the word, ‘jealous’ is an adjective and it is realized as a main verb in the first sentence of Urdu, whereas in the second Urdu translation it is being realized as a noun.

(E)- They are waiting.

(U)- ve intezaar kar rahe hein.

Here “waiting” is expressed as a verb in the English sentence whereas in the Urdu translation it is realized as a combination of Noun and verb (*intezaar kar*). It’s a very common form of divergence in English to Indian languages [Dave et al, 2001].

Some of the examples mentioned for this category by authors of [Saboor et al, 2010] are:

(U)- vah kitaabi kidaa hai
(E)-He is a bookworm.

In this example, the Adjective *kitaabi kida* of the Urdu sentence is realized as a Noun *bookworm*, in English sentence upon translation. For Urdu into English translation, this divergence also occurs when the subjective complement of SL upon translation is realized as a verb in TL [Gupta et al, 2001;2003].

(U)- murgi ande se rahii hai
(E)-The hen is hatching.

In this case, the Noun *ande / eggs* of Urdu sentence is missing in English sentence and is covered by the verb *hatching*. Another sentence that can be considered in this category is:

(U1)-usko zahar se maar diyaa gayaa
(U2)- usko zahar dekar maar diyaa gayaa

Both the above sentences can be translated to a single English sentence which is:

(E)- He was poisoned to death.

Here, the PP “zahar se” and “zahar dekar” in the Urdu sentence is converted into verb poison in English sentence upon translation.

F. Thematic divergence:

Thematic divergence refers to those divergences that arise from differences in the realization of the argument structure of a verb. Consider the following English sentence.

(E)- Where are you from?

This sentence can have the following three Urdu translations.

(U1)- Aap Ka Taaluq Kahan Se Hai?
(U2)- Aap kahan se Taaluq rakhate hein?
(U3)- Aap kahan se hein?

G. Reduplication:

Urdu, like Hindi and most of the South Asian languages, uses reduplication quite frequently (Abbi 1991). Content words can generally be reduplicated and the effect of the reduplication is to either strengthen/emphasize the original word or to express something like “and those kinds of things”. The English counterparts of these constructions do not resort to replicative structure. This distinction may often result into a change in the category of the relevant elements.

Consider an Urdu sentence “*Khudaa zarre- zarre mein basta hai*”- when translated to English this can be simply achieved by the simple English sentence “God is everywhere”. Another feature that is exhibited by Indian languages is occurrence of Echo words. The following examples illustrate this:

(U)- *Kya aap kucch thandaa vandaa lenge.*

(E)- Will you take some soft drink?

The echo words generally have no lexical status in the lexicon of the language. However these are used very commonly.

H. Honorific:

Like Hindi in Urdu also honorific features are expressed by several linguistic markers including the use of plural pronoun and plural verbal inflections. This feature is not available in a European language such as English in a similar way. This also causes a type of divergence during the translation process.

(U1) *unake vaalid aaye hein.*

(E)- His father has come.

(U2) *uskaa dost aayaa hai.*

(E)- His friend has come.

In (U1), the subject “valid”/ ‘father’ is an honorific noun which is reflected by the use of plural inflectional elements on the agreeing elements such as verb and the genitive noun. On the other hand, in (U2), dost ‘friend’ is a non-honorific noun and no plural inflectional element is used in the sentence.

I. Null subject Divergence:

In Hindi the subject of the sentence can be left implicit, which is not the case in English. Hindi allows dropping of the subject where the subject is obvious [Dave et al, 2001]. Similar situation has been observed for Urdu language as well. This can be illustrated with the following examples:

(E)- Long ago, there lived a king.

(U)- mudaton pahale, ek baadshah tha.

Here there is no explicit mapping for the word “there” of English sentence in the Urdu translation, it is assumed implicitly.

(U)- jaa rahaa hoon.

(E)-Iam going.

The subject “mein/ I” is missing and the presence of this missing subject is reflected in the morphology of the predicate. However the subject needs to be explicitly mentioned in the English sentence.

References:

- [Dave et al., 2001] S. Dave, J. Parikh, and P. Bhattacharyya, “Interlingua-based English-Hindi machine translation and language divergence,” *Machine Translation*. vol. 16(4), 2001, pp. 251-304.
- [Dorr, 1993] Dorr, Bonnie. *Machine Translation: A View from the Lexicon*. Cambridge, Mass: The MIT Press, 1993.
- [Dorr, 1994] B. J. Dorr, “Machine translation divergences: a formal description and proposed solution,” *ACL*, Vol. 20.(4), 1994, pp. 597-633.
- [Gupta et al, 2001] D. Gupta and N. Chatterjee, “Study of divergence for example based English-Hindi machine translation,” *STRANS 2001*. IIT Kanpur, 2001, pp. 132-139.
- [Gupta et al, 2003] D. Gupta and N. Chatterjee, “Identification of divergence for English-to-Hindi EBMT,” *MT Summit-IX*, Orleans. LA, 2003, pp. 141-148.
- [Gupta, 2005] D. Gupta, “Contributions to English to Hindi machine translation using example-based approach,” Ph. D. thesis, Indian Institute of Technology, New Delhi-110016. India, Jan 2005.
- [Sinha et al, 2005a] R. M. K. Sinha and A. Thakur, “Translation divergence in English-Hindi MT,” *EAMT 10th annual conference*, Budapest. Hungary, May 2005, pp. 245-254.
- [Sinha et al, 2005b] R. M. K. Sinha and A. Thakur, “Divergence patterns in machine translation between Hindi and English,” *MT Summit X*. Phuket. Thailand, Sept 2005, pp.346-353.
- [Sinha, 2009] R. Mahesh K. Sinha, *Developing English-Urdu Machine Translation via Hindi*, Third Workshop on Computational Approaches to Arabic Script-based Languages (CAASL3), *MT Summit XII*, Aug.26-30, 2009, Ottawa, Canada, pp. 89-95.
- [Saboor et al, 2010] Abdus Saboor, Md. Abid Khan, *Lexical-Semantic Divergence in Urdu-to-English Example Based Machine Translation*, 6th International Conference on Emerging Technologies (ICET), 2010, pp. 316-320.