Particles in Odia, Polish and French: a Machine Translation Perspective

Panchanan Mohanty[1], Zygmunt Vetulani[2], Grażyna Vetulani[3]

[1] Maulana Azad National Urdu University in Hydarabad, India
School of Languages, Linguistics and Indiology
Hydarabad-500032 Telangana, India
panchananmohanty@gmail.com
https://orcid.org/0000-0001-8294-0941
[2]Adam Mickiewicz University in Poznań, Poland
Faculty of Mathematics and Computer Science
Poznań, ul. Wieniawskiego 1
vetulani@amu.edu.pl
https://orcid.org/0000-0003-4833-8601
[3]Adam Mickiewicz University in Poznań, Poland
Faculty of Modern Languages
gravet@amu.edu.pl
https://orcid.org/0000-0002-2138-3704

Abstract

It is difficult to expect a high quality translation from a machine. It was believed earlier that a good bilingual dictionary and a transfer-grammar comprising the similar and different rules of the source and the target languages are required to develop a machine translation (MT) system. The goal of such an activity was to create a text in the target language that is linguistically equivalent to the source language text. Scholars involved in this activity soon realized that lot more was required in order to get a more acceptable translation and also to develop a more effective MT system. This idea became more prominent especially in the case of translating a discourse as no sentence-grammar could deliver the desired results. Particles play a crucial role in joining sentences to form a discourse and that is why these are called 'discourse particles'. These particles do not occupy a significant place in most grammars, but their pragmatic role is important and use of these particles conveys the intended message of the speaker to the hearer. We wish to discuss selected pragmatically significant particles in three languages, i.e. Odia, a scheduled Indo-Aran language spoken in India, Polish and French, two major European languages in this paper.

Keywords: discourse, ornamental particles, Odia language, Machine Translation

1. Prolog

Particles are in general small words which are non-referential semantically. But they play an important role in discourse by providing specific information to the addressee about the addresser's intention. In other words, their function is predominantly metalinguistic and pragmatic. Translation is a higher order skill, and human translators often find it difficult to translate a text from one language to another. So understandably it is hard to expect a high quality translation from a machine. King (1987:25) has listed as many as fourteen MT systems developed all over the world. Out of these, only one is Interlingua, four are direct, and seven are transfer systems. Two of these have not been classified under any category. This clearly indicates the popularity of the transfer systems over others. Earlier it was believed that the requirements for human as well as machine translation were an exhaustive bilingual dictionary and a grammar comprising the matching rules between the source and target languages. In other words, the goal of such an enterprise was to create a text in the target language that was linguistically equivalent to the source language text. But later everybody involved in this activity realized that a lot more was required in order to get a more acceptable translation and also to develop an effective MT system.

A categorical statement in this regard comes from Lefevere (1992:59) who states: "Rule-giving in translation is a thankless undertaking precisely because translation involves much more than the search for the best linguistic equivalent". Here 'much more' refers to pragmatic and cultural equivalence between the source and the target language texts. This is clearly reflected in Beaugrande's (1978:13) following statement: "The focus of translation studies would be shifted away from the incidental incompatibility ties among languages toward the systematic communicative factors shared by languages". In the same vein, Widdowson (1979:105) also states, "We cannot, of course, by definition establish pragmatic equivalence by considering isolated sentences but only by considering what utterance count as in context". Thus, it was argued that the goal of grammar must not be restricted to handle sentences in Isolation, rather it should be extended to account for supra-sentential units like text and discourse. As a result, a lot of enthusiasm was noticed to elaborate text and discourse grammars which worked out the modalities to link micro-structures like clauses and sentences with micro-structures like a text or a discourse. In MT also it was considered necessary to incorporate a knowledge base

Panchanan Mohanty, Zygmunt Vetulani, Grażyna Vetulani, *Particles in Odia, Polish and French: a Machine Translation Perspective*. In: Z. Vetulani, P. Paroubek, I. Kardava (Eds.), Human Language Technologies as a Challenge for Computer Science and Linguistics – 2025, pp. 53–56. Poznań: Adam Mickiewicz University Press, 2025. © Authors, 2025. https://doi.org/10.14746/amup.9788323245414.10. Open Access chapter, distributed under the terms of the CC licence (BY-NC--ND, https://creativecommons.org/licenses/by-nc-nd/4.0/).

whose aim was to "...represent the expert knowledge of a human translator" (Tucker 1987:38).

However, one way of establishing the above said links is by using connectives which are better known as particles in English and *avyaya*s in Indian languages. These particles are an Interesting category because "Most languages have particles whose use seems to be related to gluing the parts of discourses together, but which are never easy to pin down" (Grimes1975:93). This kind of linking cannot be explained without referring to world knowledge used by the concerned speech-communities to construct macro-structures from micro-structures. As it stands today, world knowledge is something which has not been worked out in detail for most languages, and we know that particles are carriers of world knowledge as they demand explication in terms of the speaker's and the hearer's intentions and expectations. Keeping this is view. We would like to discuss some such particles of Odia in this paper emphasizing the world knowledge they convey when they are used to construct a macro-structure so that all these can be formalized for MT purposes.

2. Discourse particles in Odia

Odia is a low-resourced language belonging to the Indo-Aryan family. It is the mail language of the state of Odisha as well as a scheduled language recognized by the Constitution of India. It has been stated in Odia grammars that there are as many particles in Odia as there are in Sanskrit; but they have not received due importance in the works of either the traditional grammarians or the Odia linguists. The case is not very different in most other Indian languages. The probable cause of this is that particles do not play a prominent role in either morphological or syntactic analysis of any language, Out of all the analyses available on the Odia particles, Nanda Sharma (1968) is the most exhaustive one. He has listed more than 120 particles in Odia along with some Sanskrit particles used in this language. Though he mentions that these particles express the emotions of the speakers and different styles of speaking, his analysis is more concerned with the etymology of the concerned particles than their functions. Another account of the Odia particles is found in MOB (1973) which puts them in different classes out of which one is termed ‘ornamental’. These ornamental particles are described to be enhancing the beauty and grace of the sentences in which they are used, In fact, we want to concentrate on the seven commonly used Odia discourse particles, i.e. /ţi/, /tɔ/, /ki/ and /kɔṇɔ/, /pɔra/, /sina/ and /pɔche/ in this paper.

In this section, we review a selection of 7 examples of *Odia ornamental particles* that (to varying degrees) pose a challenge to Machine Translation between Odia on the one hand and Polish and French on the other.

We will illustrate some of the discussed cases with examples of sentences that may pose difficulties for current machine translation techniques for all three languages.

2.1. /ţi/

This is an enclitic classifier which also expresses definiteness and affection in Odia.

For example:

1. bileiţi sundɔrɔ
   cat+ţi beautiful
   The cat is beautiful.
2. ghɔrɔţi choţɔ
   house+ţi small
   The house is small

The particle /ţi/ must not be confused with this homophonous classifier /ţi/. The former is always used after finite verbs in order to forms questions. It should be mentioned here that thought use of this /ţi/ creates a yes-no question. It conveys something more than a normal yes-no' question.

Consider the following sentences (3) and (4):

3. tu bhatɔ khaibu ?
   you rice will eat
   Will you eat rice?
4. tu bhatɔ khaibuţi ?
   you rice will eat+ti
   Will you eat rice? (confirmation request)

Both (3) and (4) seem to be conveying the same meaning in the English renderings. The former is a normal ‘yes-no' question whereas the latter conveys a strong confirmatory sense. In other words, by using the particle /ţi/ in (4), the speaker wants to get confirmation from the hearer whether the latter is going to eat or not. But this sense is completely missing in (3).

In Polish and French, there is no enclitic used after a verb as a suffix to confirm the intention to perform a specific action, e.g. eating. To achieve the effect as in example 4 (above), both Polish and French use various syntactic, lexical and stylistic means to express *confirmation of the intention* to perform an action. At the same time, Polish offers more extra-lexical means of expression at the pragmatic level than many other languages (including Oriya). In particular, in Polish and other Slavic languages, nuances of meaning can be expressed by using *flexible word order*. In languages such as French, English or German, the order of words in a sentence is more or less fixed than in Slavic languages. Nevertheless, even in these languages particles of a purely pragmatic nature are not commonly used.

The Polish sentence corresponding to example (4) /*in the **confirmatory** function*/) may have the following form:

(4a) **Czy na pewno** (ty) zjesz ryż? /**Is it sure** that you will eat rice?/

(4b) **Czy *aby* na pewno** (ty) zjesz ryż? /**Is it *really* sure** that you will eat rice?/

In this case, the use of a compound adverb at the beginning of the sentence expresses the questioner's doubt. When translating from Odia into Polish without context, it is common for the translation to be ambiguous or inaccurate. This observation also applies to French and many other characteristic example for the French language will be the equivalents of questions 4a and 4b:

Est-ce que c’est sur que tu mangera du ris?

and

Est-ce que c’est <u>vraiment sur</u> que tu mangera du ris**?** respectively.

2.2. /tɔ/

The particle /tɔ/ is used for different purposes. It can be used as a sentence conjunctive as in the following:

5. tu asibu tɔ

You will come-tɔ
Come if you want. (en)
P Przyjdź jeśli chcesz. (pl)
Viens si tu veux. (fr)

It can al/so used in b.oth interrogative well as declarative sentences. When it is used in an interrogative sentence creating a yes-no question, it is obligatorily used after the finite verb. Let us take a normal yes-no question and contrast it with its counterpart that has /tɔ/.

6. se ghɔrɔku jibɔ?
he home-to will-go
Will he go home?
Czy pójdzie do domu? (pl)
Ira-t-il à la maison? (fr)
7. se ghɔrɔku jibɔ tɔ?
he home-to will go-tɔ
Will he go home? (doubt) (en)
Zdaje się, że do domu pójdzie. (pl)
Il parait qu'il ira à la maison. (fr)

Though from the English renderings one cannot make out any difference between (6) and (7), the former is a normal yes-no question whereas the latter expresses something more. In (7) the speaker wants or expects the subject to go home, and it has been possible due to the addition of the particle /tɔ/. This /tɔ/ can be use after any constituent, like subject, object, adverb, and event the finite verb of a declarative sentence, and its presence adds emphasis to the element after which it is used.

Let us first consider its use after a finite verb:

8. se ghɔrɔku jibɔ
he home-to will-go
He will go home. (en)
Pójdzie do domu. (pl)
Il ira à la maison. (fr)
9. se ghɔrɔku jibɔ tɔ
he home-to will-go tɔ
He will go home. (certainty) (en)
Z pewnością pójdzie do domu. (pl)
Il ira à la maison **certainement.** (fr)

A comparison between (8) and (9) reveal that when the former is just a statement that he will go home, the latter emphasizes that he will certainly go home. But when /tɔ/ is used after some constituents in a sentence, it not only adds emphasis to those, but also expects another clause to make the sentence complete. The following example will drive home the point.

10. ramɔ *tɔ eţhiki* asibɔ, syamɔ bi asibɔ
Ram tɔ here will come, Shyam also will-come (en)
Not only Ram, but also Shyam will come **here.** (en)
Nie tylko Ram, ale i Shyam **tu** przybędzie. (pl)
Non seulement Ram, mais aussi Shyam viendront **ici**. (fr)
11. ramɔ *eţhiki tɔ* asibɔ, seţhiki bi jibɔ
Ram here will-come, **there** also will-go (en)
Ram nie tylko przyjedzie tutaj, ale **i tam** pojedzie. (pl)
Ram will not only come here, but he will also go **there**. (en)
Ram ne viendra pas seulement ici, mai il ira aussi **là-bas**. (fr)

Sometimes one may come across only the first part of (10) and (11), i.e. /ramɔ tɔ eţhiki asibɔ/ and /ramɔ eţhiki tɔ asibɔ/ respectively as full sentences. But in such cases they would clearly indicate ellipsis of the next parts and be interpreted solely on the basis of the contexts in which they appear.

2.3. /pɔra/

The particle /pɔra/ can be used in both declarative and interrogative sentences in Odia. Its main function in declarative sentences is to convey the meaning of recall and emphasis. In other words, it is used when the speaker recalls something and also wants to emphasize something in a sentence.

Consider the following examples:

12. tu pɔra bhatɔ khaibu
Oh! You (emphasis on *you*) will eat rice. (en)
(*To ty*) będziesz jadł ryż. (pl)
C'est toi qui vas manger du ris. (fr)
13. tu bhatɔ pɔra khaibu
O! You are going to eat (*rice)* (emphasis on *rice*). (en)
O! (Ty) będziesz jadł (*ryż*! (pl)
O! Tu vas manger du ris! (fr)
14. tu bhatɔ khaibu pɔra
O! You are going to *eat (rice*)! (strong emphasis on *to eat rice*). (en)
O! Czyżby to ty (teraz) będziesz *jadł*! (pl)
O! C'est bien toi qui *vas manger (maintenant*)!

The examples below (15 and 16) show that incorrect placement of the particle /pɔra/ may not only change the meaning of the sentence but also make it ungrammatical.

The particle */pɔra/* can also emphasize a clause as in the following example:

15. tu asibu boli **pɔra** se ɔpekhya kɔrici
you will-come COMP pɔra he waiting has-done
He has been waiting for your arrival (emphasis).

If /pɔra/ is advanced to the pre-COMP position in (15), the output will be unacceptable:

16. *tu asibu **pɔra** boli se ɔpekhya kɔrici
You will come pɔra COMP he waiting has-done

The cause of ungrammaticality of (16) is that /tu asibu boli/ is a constituent, not /tu asibu/. So /para/ has to follow /tu asibu boli /; it cannot be inserted into it.

2.4. /sina/

Use of the particle/sina/ in a clause demands that another clause must follow it to make a complete sentence. In other words, it is used in the subordinate or embedded clause in a sentence. Not only that, it emphasizes the element that precedes it and the corresponding element in the matrix clause must contrast with the former. For example :

17. se sina bhatɔ khalla, ta bhai ruţi khaila
he sina rice ate his brother chapatti ate
He (emphasis) ate rice, but his brother ate chapatti.
18. se bhatɔ sina khaila, tarɔ ruţi khalbaku iccha thila
he rice sina ate, his chapatti to-eat wish was
He ate (emphasis) rice, but he wanted to eat chapatti.
19. se bhatɔ khaila sina, au ţike pare banti kɔribɔ
he rice ate sina, another little after vomiting will-do
He ate (emphasis) rice, but he will vomit after some time.

Notice that in (17), /sina/ is used after the subject of the matrix clause /se/ 'he' and the subject of the embedded clause /ta bhai/ 'his brother' is contrasted with it. Similarly,

in (18) and (19) it is used after the object /bhatɔ/ ‘rice’ and the finite verb /khaila/ ‘ate’ respectively in the embedded clauses; and these are being contrasted with the corresponding object /ruti/ ‘chapatti’, and /banti karibɔ/ ‘will-vomit’ of the matrix clauses of the concerned sentences. A characteristic feature of this particle is that it cannot be used in a clause if its verb is in the imperative mood. The following example is illustrative:

20. se asu sina, eṭhi rəhiparibɔni
    he let come sina, here cannot stay
    Let him come, but he cannot stay here.

3. Particles: a challenge for Machine Translation technologies

In this work, we focus on indeclinable particles in Odia in the context of the development of machine translation technology. We identified major problems in obtaining high quality machine translation between languages similar to Odia and most European languages.

For this work we selected three Indo-European languages representing medium-sized speaker populations: Odia, Polish, and French. Despite the structural similarities inherited from Sanskrit and Latin, there are significant differences between them at the pragmatic level. They make it difficult, or even impossible, to obtain a good quality translation without analysis that allows for the identification f semantic differences. This is because, in order to guide the reader in the correct reading of the text, authors often resort to the use of indeclinable particles specific to particular languages.

A translation problem may arise when the indefinable particles used in the source text do not have direct equivalents in the target language. In such cases, the correct interpretation of the particle requires analysis of the context of use. In the sections 3 to 8 of this work we have given examples of typical contexts corresponding to individual particles. Correctly interpreting particles requires analyzing the context of use. Applying modern technologies (e.g., to leverage the potential of GenAI) will require training language models based on large tagged corpora of representative usage examples. /Nb. in traditional dictionaries addressed to humans, the method of assigning examples of use to the dictionary entry was commonly used./

Currently used machine translation systems have so far relied primarily on descriptive, traditional, rule-based grammars. This was satisfactory for medium quality translation ignoring (partially) pragmatic level. Today MT modes refers to traditional grammatical (descriptive) categories, like verb, noun, etc. This is not the case of particles, as this category is language dependent.

Analyzing the examples provided in the sections 3-8 above, it seems that to achieve a higher level of quality it will be necessary to supplement the current IT workshop with tools such as (human-driven) generative artificial intelligence (GenAI).

4. Concluding remarks

To conclude, we discussed the distribution and pragmatics of seven discourse particles of Odia which have been neglected by both the traditional grammarians as well as the linguists. The aim of this discussion was to help primarily the MT scholars working on Indian languages to formalize the pragmatics of these particles for MT purposes and also look for their nearest equivalents in other Indian languages for a closer and more effective translation from Odia to other languages.

References

De Beaugrande. R. 1978. *Factors in a theory of poetic translating*. Assen van Goccum.

Grimes, J.E. 1975. *The threads of discourse*. Berlin: Mouton.

King, M. 1987. *A tutorial on machine translation*. Geneva: ISSCO.

Lefevere, A. (ed.). 1992. *Translatoan/history/culture: A source book*. London: Routledge.

MOB (*madhyɔmikɔ oḍia byakɔraṇɔ*). 1973. Cuttack: Board of Secondary Education Orissa.

Tucker, A.B. 1987. Current strategies in machine translation research and development, In *Machine translation: theoretical and methodological issues*, ed. by S. Nirenburg. Cambridge: Cambridge University Press.

Widdowson, H.G. 1979. *Explorations in applied linguistics*. Oxford: Oxford University Press.