Nepali *Le* as a Marker of Categorical Subjecthood

Luke Lindemann, *Yale University*

Abstract

The purpose of this paper is to investigate an account of ergative-nominative alternations in non-perfective clauses in Nepali. As observed in many Indo-Aryan languages, the ergative case marker is obligatory in perfective tenses. The variable presence of ergative marking in the non-perfective domain of Nepali has been articulated as an expression of emphasis, subject animacy, or individual-level predication. I argue that that =*le* marks the subject of a **categorical proposition** in the sense of Kuroda (1972). I explore the felicity of =*le* in particular discourse contexts depending upon whether the response is presented thetically or categorically. I also note that =*le* may only be found in quantifier phrases with strong construal, which provides support for the notion that =*le* marks categorical subjects.

1   Introduction

Many modern Indo-Aryan languages have a split in ergative case-marking conditioned by aspect. In general, the verb agrees with the nominative subject of a transitive clause outside of the perfective domain, and in the perfective domain there is ergative case-marking and the verb agrees with the unmarked object. Masica describes the "classic NIA split" as "quasi-ergative case-marking and agreement in the Perfective only, vs nominative-accusative patterns in non-Perfective tenses"(Masica 1993, 342). We can see this pattern as it occurs in Hindi (Deo and Sharma 2006, 376):

(1) a. ram=ne   **cidiya**   **dekh-i**
         Ram.M=Erg sparrow.F.Sg.Nom see-Perf.F.Sg
         'Ram saw a sparrow.'

     b. **sita**   ram=ko   **dekh-t-i h-ai**
         Sita.F.Nom Ram.M=Acc see-Impf-F.Sg be-Pres.3.Sg
         'Sita sees Ram.'

In (1a), the verb is marked in the perfective and agrees with the object. There is ergative case-marking (-*ne*) on the subject of the sentence, *Ram*. For perfective tenses such as in (1a), the verb agrees with the object if there is no accusative case marking, and agreement is default otherwise. In (1b), the verb is marked in a non-perfective tense and agrees with the subject. The object is marked with the accusative case marker -*ko*. This is because human-denoting specific objects in Hindi obligatorily take accusative marking. Thus we have an ergative pattern for perfective tenses and a nominative-accusative pattern in non-perfective tenses.

There is a substantial amount of variation in this pattern among the Indo-Aryan languages (cf. Deo and Sharma 2006). Some languages like Bangla have completely lost all ergative case marking. Other languages, like Assamese, have extended the ergative case marking to all transitive clauses in the language (Masica 1993, 344). Nepali presents with a particularly unique pattern. Ergative marking is obligatory in the perfective domain, and it varies with the nominative elsewhere in the language. In the non-perfective domain, the same sentence may be expressed with or without ergative case marking. Verbal agreement is always with the subject whether or not the subject carries the

ergative case marker =*le*:[1]

(2) a. **ram=le**   cidi   **dekh-yo**
     Ram.M=Erg sparrow.Nom see-Perf.3.Sg
     'Ram saw a sparrow.'

   b. **sita**   / **sita=le**   ram=lai   **dekh-chin**
     Sita.F.Nom / Sita.F=Erg Ram=ACC see-Impf.F.3.Sg
     'Sita sees Ram.'

As in Hindi, the case marker (=*le*) is obligatory in the perfective. Unlike Hindi, the verb agrees uniformly with the subject. In both (2a) and (2b), the perfective verb agrees with the subject, even if there is an ergative case marker on the subject. This is true whether or not the object has accusative case-marking. The most interesting case is (2b), in which we find optionality. Here we see that the subject may take an ergative marker or may be unmarked. This optionality is apparently present for all verbs in non-perfective tenses:

(3) a. **sunita** / **sunita=le**   khana **pɑk-aun-dɑi-chɑ**
     Sunita.F / Sunita.F=Erg food   cook-Caus-Prog-Impf.M.3.Sg
     'Sunita is cooking food.'

   b. **ram** /   **ram=le** hɑrek din euṭa   ãp   **khan-chɑ**
     Ram.M / Ram.M=Erg every day single mango eat-Impf.M.3.Sg
     'Ram eats a mango every day.'

   c. **mɑ** / **mai=le**   **pucc-ãula**
     I   / I.Obl=Erg clean-Ind.Fut.1.Sg
     'I will clean it.'

   d. **cor** /   **cor=le**   scarf **cor-thyo**
     thief / thief=Erg scarf steal-Hab.M.3.Sg
     'Thieves would steal the scarf.'

No other language among the contemporary Indo-Aryan languages exhibits such a pattern of ergative marking in the non-perfective domain. This pattern in Nepali has been observed at least since Grierson (1904), who takes ergative marking as an expression of emphasis. It has been similarly characterized as an expression of emphasis by Masica (1993) and Clark (1963). Abadie (1974) argues that =*le* disambiguates the subject argument when it would otherwise be ambiguous. Finally, Butt and Poudel (2007) notes the correspondence between the appearance of =*le* and individual-level predication.

In this paper, I review the various proposals that have been made regarding =*le* as it appears in the imperfective domain. I will discuss my own proposal, which is that =*le* optionally marks the subject of a categorical proposition as formulated by Kuroda (1972) . Thus, whenever =*le* appears on the subject of an imperfective transitive clause, that clause is a categorical proposition. Next, I will discuss the predictions that this theory makes for the use of =*le* in particular discourse contexts. I will also examine evidence from the appearance of =*le* on quantificational determiners.

2 Previous research on =*le*

This unusual extension of =*le* into the non-perfective domain in Nepali has been noted since at least Grierson (1904), who describes this usage of =*le* as "idiomatic" and "emphatic", and attributes its usage to influence from Tibeto-Burman languages. The term "emphasis" is invoked in Clark (1963) and Masica (1993), although it is never explicitly defined. Bickel (2011) takes this as evidence that =*le* marks focus, but Abadie (1974) and Verbeke (2011) both disagree with the notion that the

---

[1]Tense and aspect markers are glossed as PRES: Simple Present (Imperfective), PERF: Perfective (Simple Past), HAB: Past Habitual, PROG: Progressive, FUT: (Nondefinite) Future, DEF.FUT: Definite Future, CAUSE: Causative, PASS Passive. Case markers are ERG: Ergative, ACC: Accusative, OBL: Oblique, INSTR: Instrumental, ABL: Ablative, GEN: Genitive, LOC: Locative.

element marked by =*le* must be focused. The term "emphasis" captures the intuition that the argument marked by =*le* is given special attention in the discourse, but this is still a rather vague notion. Abadie (1974) suggests that =*le* may be used to avoid ambiguity when there would otherwise be a meaning difference, as in the following (Abadie 1974, 15):

(4) a. yo gai=le khan-chɑ
         this cow=Erg eat-Pres.M.3.Sg
         'This cow eats.'

     b. yo gai khan-chɑ
         this cow eat-Pres.M.3.Sg
         'This cow eats / This (one, person) eats cow.'

When =*le* is present as in the first example of (4a), "cow" must be the subject. In the second example of (4b), there are two possible interpretations: "This cow eats," and "This (one, person) eats cow." The determiner *yo* may be marking an elided subject. While it is true that (4b) is ambiguous in a way that (4a) is not, it is easy to find situations in which =*le* may be preferred even though it cannot serve to disambiguate the subject argument:

(5) gai=le   yo   khan-chɑ
     cow=Erg this eat-Pres.M.3.Sg
     'A cow eats this.'

The above example is a minimal adjustment to Abadie's example. In (5), =*le* cannot be serving to disambiguate the subject argument, because the position of the determiner after 'cow' indicates that it must be marking an object (in canonical SOV order). And yet here speakers still use =*le*, and may even find it odd for it to be absent. While =*le* does have the effect of disambiguating the subject of a sentence when it would otherwise be ambiguous, more needs to be said about the function of =*le* in the many situations in which =*le* does not disambiguate the arguments.

A promising proposal comes from Butt and Poudel (2007), who suggest that =*le* marks individual-level predication. This distinction is illustrated with the following examples (Butt and Poudel 2007, 7):

(6) a. calɑk=le   gaḍi cɑlaun-chɑ
         driver.M=Erg car   drive-Pres.M.3.Sg
         'The driver drives the vehicle.'

     b. guru   gaḍi cɑlaun-chɑ
         teacher.M.Nom car   drive-Pres.M.3.Sg
         'The teacher is driving/will drive the vehicle.'

The context in the first example of (6) is that a school's bus driver drives the children everyday: that is his occupation. In the second, the teacher just happens to be driving the bus today because the bus driver is out. Note that the simple present tense of the verb *calāuncha* 'to drive' has three possible interpretations: it may have a habitual reading, an immediate present reading, or a future-oriented reading. In this sentence, =*le* marks the reading with individual-level predication, that is, the habitual reading.

When presented with sentences like those above, many speakers tend to agree with the idea that =*le* marks the difference between a person who drives taxis as an occupation, and a person who simply may be doing these things at the moment. However, as we will see below, =*le* may also be found in clauses for which the predicate is neither habitual nor individual-level. We have already seen an example of =*le* when the sentence is in the present progressive(3b). Verbeke provides more examples in which individual-level predication does not seem to be present (Verbeke 2011, 165):

(7) raja=le   sodh-e   hɑn yɑs=le   pheri ke   bhan-dai-che
     king=Erg ask-Perf.M.3.Sg Q   she=Erg again what say-Prog-Pres.F.3.Sg
     'The king asked: "What is she saying?"'

(8) aphu=le na-bok-era   kɑs=le   bok-i-din-chɑ
    self=Erg Neg-carry-Conj who=Erg carry-Lnk-give-Pres.M.3.Sg
    'If I don't carry it myself, who will carry it?'

In summary, the marking of =*le* on the subject of transitive perfective clauses has been described as giving the subject greater emphasis, as disambiguating the subject argument, or as distinguishing individual from stage-level predication. All of these notions express intuitions about the way in which a sentence with a marked subject has a different meaning. In the next section, I will argue that a distinction between thetic and categorical propositions can tie many of these notions together to represent the contribution of =*le* in the clause.

3   Proposal

Kuroda (1972), in analyzing topic marking in Japanese, proposes a distinction between two classes of propositions - thetic and categorical. This distinction comes from the theories of judgment propounded by the 19th century philosophers Brentano and Marty (Kuroda 1972, 1990, Ladusaw 2000, McNally 1998). A **thetic proposition** is a description of an entity or an eventuality in which no element of the sentence is given particular discourse prominence. Ladusaw (2000) describes a thetic judgment as an existential commitment to a description. A **categorical proposition** relates the occurrence of the described event to a particular entity. It is a "double judgment": an entity is first presented, and then a property is predicated of this entity. Ladusaw notes that the subject of a categorical judgment must be presupposed in the discourse, and must be strongly construed.

   The argument advanced here is that =*le* marks the subject of a categorical proposition in Nepali. If we take =*le* to be a marker of categorical subjecthood, this provides us with an explanation for intuitions about minimal pairs such as (9) below:

(9) a. mɑ curoṭ   khan-chu
       I   cigarette eat-Pres.M.3.Sg
       'I smoke cigarettes.' (I have a habit)

   b. mɑi=le curoṭ   khan-chu
      I=Erg   cigarette eat-Pres.M.3.Sg
      'I smoke cigarettes.' (I am a smoker)

The first statement may be a simple statement of habit, but in the second statement =*le* emphasizes that the person is a *curoṭ khane mance* 'a cigarette-smoking person.' This is what it means to make a categorical proposition: we first bring attention to an entity, and then we predicate a property of that entity. Similarly, the unmarked form in (10a) below does not emphasize, as (10b) does, that the speaker is building his own house:

(10) a. mɑ aphno ghɑr banau-dɑi-chu
        I   own   house build-Prog-Pres.M.1.Sg
        'I am building (my) own house.'

    b. mɑi=le   aphno ghɑr banau-dɑi-chu
       I.Obl=Erg own   house build-Prog-Pres.M.1.Sg
       'As for me, I am building (my) own house.'

Here both predicates have a stage-level interpretation, but -*le* has the effect of directing attention first to the speaker and then to the activity. My consultant Timila Dhakwa expressed it the following way:

   "It seems that when it is a general statement you can leave out the =le and it still makes sense. But having the =le just makes it clear as to who is doing the action. You are making the extra statement that it is [the subject] doing the action."

By "marker of a categorical subject," I mean that =*le* marks the sentence argument that is logically the subject of a categorical proposition. This distinction corresponds somewhat to topic, particularly

in that the categorical subject is distinct from the syntactic or grammatical subject of the sentence. Kuroda advances the thetic/categorical distinction to explain the Japanese marker *wa*, which is frequently characterized as a prototypical topic marker. The Japanese marker *wa* may mark the object or even certain non-NPs that can be construed as the sentence topic.

The situation in Nepali is different because *=le*, as the ergative marker, may only mark the grammatical subject of the sentence. However, the entity marked by *=le*, when it is present in non-perfective clauses, clearly has a particular discourse prominence. Portner and Yabushita (1998) discuss Japanese as it relates to a topics-as-entities formalization of topicality in which topics denote entities that the sentence is "about." The topic information is part of the common ground. The sentence element identified as a topic must be definite and its existence must be presupposed in the mind of its speakers. I believe that this is what is intended by "emphasis" - the element marked by *=le* refers to an entity that is what the sentence is about.

This analysis shares with Butt and Poudel (2007) the intuition that *=le* brings attention to a particular argument in order to express a property of that object.[2] It is compatible with a view like Abadie's in which a statement may be expressed with *=le* in order to clearly delineate arguments, but it is more generally an expression of information structure. In many discourse contexts a speaker may felicitously express the same statement either thetically or categorically, and this is the source of the optionality of *=le*.

4   Discourse Context

The felicitous utterance of a thetic or categorical proposition is related to discourse structure. As Caro notes for Spanish, "often the same proposition may be expressed thetically or categorically by the speaker, by choosing to initiate his/her message with a topic and then proceed with the focus information or by presenting the state of affairs as a compact event involving no obvious parts" (Caro 2009, 18). If a sentence marked with *-le* must be a categorical proposition, then such a sentence:

1. may contain an individual-level or stage-level predicate, but
2. cannot be a thetic proposition.

Furthermore, the NP marked by *-le*:

1. must refer to a presupposed entity, i.e.,
2. must have a strong construal.

If *=le* marks the subject of a categorical judgment, then it cannot appear in a thetic clause. So a clause in which *=le* appears is a categorical proposition. If the subject is unmarked, then it is a thetic proposition which may not be a felicitous response to a question about a particular entity. The example below illustrates this interaction. In this situation, a man hears a loud noise outside and notices his friend looking out the window:

(11) Q: bahirɑ ke   hun-dɑi-chɑ?
        outside what happen-Prog-Pres.M.3.Sg
        'What is happening outside?'

    A: shikari / shikari=le   mriga sɑmat-dɑi-chɑ
        hunter / hunter=Erg deer   catch-Prog-Pres.M.3.Sg
        'A/The hunter is hunting a deer.'

In (11), the question is a request for information about a general state of affairs. A felicitous response may contain a description (a thetic proposition) or relevant information about a particular argument (a categorical proposition). If the existence of a hunter is not presupposed in the discourse, then the subject *shikāri* will not be marked with *=le*. This corresponds to the English "A/some hunter is

---

[2]Butt and Poudel briefly equate stage-level predicates with thetic propositions and individual-level predicates with categorical propositions, but I believe these notions are logically separate. Thus *=le* may be found with either individual-level or stage-level predicates.

hunting deer." If the existence of the hunter *is* presupposed (for example, if there is one particular hunter who is always lurking around the property such that the hunter's existence is known to both speakers), then *=le* may be present or absent, depending upon whether the speaker wishes to express the observation thetically or categorically. The corresponding English sentence for both situations would be "The hunter is hunting a deer."

If, conversely, the question is the one expressed by (12), then only a categorical proposition will do as a response. A request for information about a particular entity probably cannot be answered with a thetic proposition. Here we are looking for information about a particular hunter. Thus, *=le* must be present, and the corresponding English sentence would again be "The hunter is hunting a deer."

(12) Q: shikari=le ke gar-ɑi-cha?
       hunter=Erg   what do-Prog-Pres.M.3.Sg
       'What is the hunter doing?'

    A: #shikari / shikari=le   mriga samat-ɑi-chɑ
       #hunter / hunter=Erg deer   catch-Prog-Pres.M.3.Sg
       'The hunter is hunting a deer.'

In many contexts, a Nepali speaker may have the choice of expressing the same statement either thetically or categorically. It is because of this that the use of *=le* appears to be optional in many contexts. But if the question concerns a particular entity like the hunter, then *=le* is obligatory.

5   Construal of Quantifiers

Following the notion that the element marked by *=le* must have a strong construal, we can examine the distribution of *=le* for particular quantifiers which can be construed as either weak or strong. Following the discussion of ambiguous quantifiers like 'some' and 'many' in Partee 1983, we can similarly construe Nepali quantifiers like *dherai* 'many' and *kohi* 'some' as ambiguous between a strong construal, which presupposes the existence of a set, and an indefinite weak construal. Other quantifiers, like *dheraijaso* 'most', seem to be unambiguous. *Dheraijaso* has only a strong reading. The usage of particular quantifiers thus interacts with the placement of *=le*, because the subject of a thetic proposition cannot have strong construal:

(13) dherai biddyarthi din-ko   dui-tin ghɑnṭa sik-chɑn
     many student   day-Gen two-three hour learn-Pres.M.3.Pl
     'Many students / Many of the students study 2-3 hours a day.'

There is both a weak and a strong interpretation available for unmarked *dherai* as in (13) above. A strong construal may be appropriate in a discourse about a particular group of students and their study habits, while a weak reading may be a more general statement about students. However, when *=le* marks the quantifier the weak reading is no longer available:

(14) dherai=le biddyarthi din-ko   dui-tin ghɑnṭa sik-chɑn
     many=erg student   day-Gen two-three hour learn-Pres.M.3.Pl
     'Many of the students study 2-3 hours a day.'

The quantifier *kohi* 'some', a similarly ambiguous quantifier, shows a similarly distinction between marked and unmarked forms, in which the strong reading is the only one available with *kohi-le*:

(15) a. kohi.kohi biddyarthi din-ko   dui-tin ghɑnṭa sik-chɑn
       some.Red student   day-Gen two-three hour learn-Pres.M.3.Pl
       'Some students / Some of the students study 2-3 hours a day.'

    b. kohi.kohi=le   biddyarthi din-ko   dui-tin ghɑnṭa sik-chɑn
       some.Red=Erg student   day-Gen two-three hour learn-Pres.M.3.Pl
       'Some of the students study 2-3 hours a day.'

An unambiguously strong quantifier like *dheraijaso* 'most', whether marked or unmarked by *=le*, has only one reading:

(16) dheraijaso / dheraijaso=le biddyarthi din-ko   dui-tin ghaṇṭa sik-chan
    most / most=Erg   student   day-Gen two-three hour learn-Pres.M.3.Pl
    'Most of the students study 2-3 hours a day.'

This is summarized in Figure 1.

	Weak	Strong
dherai N	Yes	Yes
dherai N-*le*	**No**	Yes
kohi N	Yes	Yes
kohi N-*le*	**No**	Yes
dheraijaso N	**No**	Yes
dheraijaso N-*le*	**No**	Yes

FIGURE 1 Marked and bare forms of quantifier phrases

We can also look at quantifier phrases with elided subjects. If it can be inferred from the context, the subject in a quantifier phrase may be deleted. In this case, *=le* may directly mark the quantifier directly. If the subject is elided, it must be strongly construed, so there will only be one interpretation of the quantifier phrase. Ambiguous quantifiers may be marked or unmarked, depending upon whether the sentence is interpreted thetically or categorically:

(17) a. kohi.kohi / kohi.kohi=le   biddyarthi din-ko   dui-tin ghaṇṭa sik-chan
       some.Red /some.Red=Erg student   day-Gen two-three hour learn-Pres.M.3.Pl
       'Some of the students study 2-3 hours a day.'

     b. dherai / dherai=le   biddyarthi din-ko   dui-tin ghaṇṭa sik-chan
       many.Red many.Red=Erg student   day-Gen two-three hour learn-Pres.M.3.Pl
       'Many of the students study 2-3 hours a day.'

The unambiguously strong quantifier *dheraijaso*, as in (18), seems to be generally dispreferred with elided subjects. While this does not follow directly from the analysis presented here, it certainly bears further investigation. It is one of the few contexts in which the absence of *=le* seems to be dispreferred under any reading:

(18) dheraijaso / dheraijaso=le biddyarthi din-ko   dui-tin ghaṇṭa sik-chan
    most / most=Erg   student   day-Gen two-three hour learn-Pres.M.3.Pl
    'Most of the students study 2-3 hours a day.'

6 Conclusions

I hope to have shown that describing *=le* as a marker of categorical subjecthood unifies many of the varied proposals and intuitions about the contribution of *=le* to the clause. In particular, Butt and Poudel's notion that *=le* marks individual-level predication captures the same basic intuition that *=le* brings attention to a particular argument in order to express an inherent property of that object, but this account allows for the presence of *=le* with stage-level predicates.

This makes the prediction that we should not find *=le* in a thetic clause, and that the referent should be presupposed. We can see how this interacts with discourse, in which a speaker often has the opportunity to choose between a thetic or a categorical response. This also aligns nicely with quantifiers in Nepali, in which (strong) proportional readings must be correlated with *=le* marking.

Moving forward, I believe that we should explore cases like those expressed in the previous section to see how commonly *=le* may be found on indefinite referents, because this represents a puzzle. It

would also be useful to explore the relationship between $=le$ and the prosodic structure of the clause, because this can give us insight into focus and topic as it is expressed more generally in Nepali.

Acknowledgments

References

Abadie, Peggy. 1974. Nepali as an ergative language. *Linguistics of the Tibeto-Burman Area* 1(1):156–177.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3):435–483.

Bickel, Balthasar. 2011. Grammatical relations typology. In J. J. Song, ed., *The Oxford handbook of language typology*, pages 399—445. Oxford University Press.

Butt, Miriam and Tikaram Poudel. 2007. Distribution of the ergative in nepali. *Manuscript, University of Konstanz* .

Caro, Elena Matínez. 2009. Pragmatic frames, the thetic-categorical distinction and spanish constituent order. *ALFA: Revista de Lingüística* 51(2).

Clark, Thomas Welbourne. 1963. *Introduction to Nepali: a first-year language course*. School of Oriental and African studies, University of London.

Cohen, Ariel. 2001. Relative readings of many, often, and generics. *Natural Language Semantics* 9(1):41–67.

Deo, Ashwini and Devyani Sharma. 2006. Typological variation in the ergative morphology of indo-aryan languages.

Grierson, George A. 1904. Specimens of pahari languages and gujuri. *Linguistic Survey of India* 9(4).

Kuroda, S-Y. 1972. The categorical and the thetic judgment: Evidence from japanese syntax. *Foundations of language* pages 153–185.

Kuroda, S-Y. 1990. The categorical and the thetic judgment reconsidered. In *Mind, meaning and metaphysics*, pages 77–88. Springer.

Ladusaw, William. 2000. Thetic and categorical, stage and individual, weak and strong. *Negation and Polarity. Syntactic and Semantic Perspectives, Oxford University Press, Oxford* pages 232–242.

Li, Chao. 2007. Split ergativity and split intransitivity in nepali. *Lingua* 117(8):1462–1482.

Masica, Colin P. 1993. *The Indo-Aryan Languages*. Cambridge University Press.

McNally, Louise. 1998. On recent formal analyses of topic. In *The Tbilisi symposium on language, logic, and computation: Selected papers*, vol. 14, pages 147–160. Stanford, CA: CSLI Publications.

Portner, Paul and Katsuhiko Yabushita. 1998. The semantics and pragmatics of topic phrases. *Linguistics and Philosophy* 21(2):117–157.

Schwarz, Anne. 2010. Discourse principles in grammar: the thetic/categorical dichotomy. *Ietropic: electronic journal of studies in the tropics* .

Verbeke, Saartje. 2011. *Ergativity and alignment in Indo-Aryan*. Ph.D. thesis, Ghent University.