

How revolution and gender dynamics affected Nepali 3rd-person honorification from the 1990s to the 2010s: A corpus analysis

Subhekshya Shrestha*

Abstract. One way to express honorification, by which speakers index relations of respect, is through varying pronouns and verb inflection. This phenomena has been studied in European languages with binary 2nd-person pronouns (the T-V model). Past research finds that these systems tend to level with social change. This study explores Nepali, which also uses pronouns and verb-inflections. Nepali has 2nd and 3rd-person honorification, with 4 and 3 different levels respectively. The study employs a corpus analysis to investigate Nepali 3rd-person honorification with regards to the fall of monarchy (1996-2010). The paper finds that this social change correlates to 3rd-person honor leveling.

Keywords. honorification; South Asian linguistics; gender; corpus analysis

1. Introduction. Languages have diverse ways of denoting relationships of respect and deference. For example, English has titles like *lord* or *lady* and Japanese uses suffixes like *-san* or *-sama*. Other ways to employ honorifics surface through morpho-syntactic inflections or lexical changes. One such type is called the Avoidance Register, where certain forms are avoided in the presence of specific groups of people. For instance *mother-in-law language* in Dyirbal is predicated on avoidance of certain words to indicate politeness in the presence of the speaker's mother-in-law. Instead, speakers replace the forms they are avoiding with separate lexemes (Dixon 1990). So, a *mother-in-law language* form for "oak" would be *gurrugun* while the non-avoidance form would be *gurray* (Dixon 1990). Japanese *keigo*, on the other hand, is a complex system of modifying one's speech according to their politeness context. This may take the form of different suppletive verb forms depending on the level of politeness one is using (Wetzel 2004). For instance, the plain form for the verb-phrase "I/you will find out" is *shiru* while polite-honorific form is *gonzōji* (Wetzel 2004). Many other languages have differing pronouns and associated verb inflections dependent on the context of the relationship between addresser and addressee. This paper will focus mainly on this last type of honorification as it pertains to Nepali, an understudied South Asian language.

1.1. THE T-V MODEL. *The pronouns of power and solidarity* (Brown & Gilman 1960) was one of the first works to tackle honorification as expressed through pronouns. It established the T-V paradigm to classify the two types of 2nd-person pronouns present in many European languages, such as *tu* and *usted* in Spanish or *ty* and *vy* in Russian. The T-V model was named after *tu* (T) and *vos* (V) from Latin, with the former being the lower register and the latter being the higher register. The French example in (1) demonstrates that register is expressed through changing pronouns, as well as the verb morphology. In languages with the T-V distinction the relationship between the addresser and addressee surfaces in the grammatical inflection of the verb.

(1) T-V distinction in French

- a. Tú as 18 ans.
 2SG-T have-PRES-T 18 years.

* Author: Subhekshya Shrestha, the University of Chicago (subhekshya@uchicago.edu)

- b. Vous avez 18 ans.
 2SG-V have-PRES-V 18 years.
 You are 18 years old. (lit. ‘You have 18 years.’)

Brown & Gilman explored how these 2nd-person pronouns in European languages differ in usage according to societal contexts. T or V usage was determined by whether the speaker held some power over the addressee or vice versa. The theory states that the person in the higher power position would receive V and would use T for the person in the lower power position. An employer might use T with their employee and receive V. Furthermore, the V pronouns in these languages historically started out as the plural form of the T-form, positing some kind of plurality to the person receiving it based on the power they held. Thus, Brown & Gilman framed power as an inherently asymmetric relation as demonstrated by the non-reciprocal usage of V. Meanwhile, people on equal footing would reciprocally use T intimately and V if there was some perceived distance between speakers. Close friends would use T with each other, while acquaintances might use V. So, varying registers is framed as not only contingent on deference but also familiarity between interlocutors. The authors further argue that societies that have become more egalitarian over their history would leveled their T-V usage to one or other throughout in order to reflect such a change. Such is proclaimed to be the case for the English 2nd-person pronoun shifting to just *you* from the T-V forms *thou* and *you* respectively.

The T-V model of pronouns has been the framework for most work done in the study of honorification and registers. However, newer literature complicates the specific claims made by Brown & Gilman on how changes in the use of T or V can index a particular society’s view on power relations. In her paper, Wales (1983) argues that the English shift towards only having *you* as the second person pronoun was not the result of some egalitarian shift in English society. Rather, social stratification was extremely high in the time that this leveling took place. She purports that this change instead resulted from the middle-class aspiring to habits of polite society by using the more formal V-form, “you”. Another criticism comes from Friedrich (1972), who cites various uses of T- and V- forms in Russian literature fluctuating depending not only on relationships defined by macro-sociological variables like power, kinship or age but also through context depending on the specific speech event.

A lot of scepticism has been cast as to the nature of the T-V model’s indexicality of the power relations in society as noted in Agha (1994). These arguments come from the study of use in interpersonal contexts. A larger quantitative study of honorific use over time could certainly elucidate these societal shifts of attitude towards relations of power, especially at pivotal moments in history. Moreover, Brown & Gilman’s model only accounts for languages that have a two pronoun system. Despite these criticisms, and its lack of universality, Brown & Gilman’s theory on how larger structures of power can affect language use on an interpersonal level has some merit, especially in the context of institutional texts. Within this framework, this study looks at a language with a different system of honorification in the context of a non-European culture. The case of honorification will be examined henceforth as it pertains to Nepali. More specifically, how the trends of the three-level 3rd-person honorification system is affected by historical events in the late 90s and early 2000s, namely the fall of monarchy and the rise of increased gender equality. These trends will be explored using a corpus analysis, which aims to quantitatively capture language change trends in face of historical factors. The next section will give a brief overview of Nepali honorification as well as relevant features of Nepali syntax.

1.2. NEPALI SYNTAX AND HONORIFICATION. Nepali, an Indo-Aryan language, is the lingua franca of Nepal and is spoken in various parts of South Asia. Nepali sentences typically follow a Subject-Object-Verb (SOV) structure (Acharya 1991). Nepali is a split-ergative language, where subject noun-phrases (NPs) in Nepali can be either bare or carry the ergative marker, *le*, depending on the verb aspect. Nepali has subject-verb agreement. The agreement of the verb depends on features of the subject NPs such as gender, number, person, and most pertinent to this paper, honorification. Nepali honorification surfaces through the verb inflection according to the particular subject NP it agrees with, like in the T-V system. However, this system differs from many T-V languages in two ways: (1) it is not binary and (2) it has both 2nd and 3rd person honorification. This means that honorification in Nepali exists past the addresser-addressee relationship, and is obligatorily marked in every sentence in the 2nd and 3rd-person.

In Nepali, honorification is mainly expressed in the 2nd as well as the 3rd-person. There are four distinct levels of honorification in the 2nd-person, shown in (2), and three levels in the 3rd-person, shown in (3). The various verbal inflections for honorification in Nepali are listed below, and they are ordered from least honorific to most. The glosses for the honorification markings are as follows: LoHon = Low Honorific, MedHon = Medial Honorific, HiHon = High Honorific, UHiHon = Upper High Honorific.

(2) **2nd person honorification**

- a. tā kitaab padhchas.
 2-SG-LOHON book-SG read-2-SG-LOHON-M-PRES
- b. timi kitaab padhchau.
 2-SG-MEDHON book-SG read-2-SG-MEDHON-M-PRES
- c. tapāi kitaab padhnuhuncha.
 2-SG-HIHON book-SG read-2-SG-HIHON-PRES
- d. hajur kitaab padhibakshincha.
 2-SG-UHIHON book-SG read-2-SG-UHIHON-PRES
 You read a book. (lit. ‘You book read.’)

(3) **3rd person honorification**

- a. u kitaab padhcha.
 3-SG-LOHON book-SG read-3-SG-LOHON-M-PRES
- b. uni kitaab padhchan.
 3-SG-MEDHON book-SG read-3-SG-MEDHON-M-PRES
- c. uhā kitaab padhnuhuncha.
 3-SG-HIHON book-SG read-3-SG-HIHON-M-PRES
 He reads a book. (lit. ‘He book reads.’)

Only the LoHon and MedHon levels encode gender and number agreement on the verb while HiHon and UHiHon forms only encode honorification. Examples 3 (a)-(c) have demonstrated the 3rd-person masculine inflections, while examples 4 (a)-(c) below will show the feminine forms.

As we can see, the forms for (3c) and (4c) are identical whereas the rest change when the gender of the subject NP changes. Furthermore, Nepali is a language in which only human feminine NPs have overt inflection, the suffix *ii*. For example, the word for "girl", *ketii*, is constructed by adding the feminine suffix to the word for boy, *keta*. Verbs agreeing with the masculine human subject NPs and non-human subject NPs have the same inflection in LoHon. This is because honorification only applies to human entities, and non-human entities (e.g. 'pen' or 'cat') can only take LoHon.

(4) 3rd person feminine honorification

- a. *ketii kitaab padhche.*
girl-SG book-SG read-3-SG-LOHON-F-PRES
- b. *ketii kitaab padhchin.*
girl-SG book-SG read-3-SG-MEDHON-F-PRES
- c. *ketii kitaab padhnuhuncha.*
girl-SG book-SG read-3-SG-HIHON-PRES
A girl reads a book. (lit. 'Girl book reads.')

Similar to how the V-form in many European languages is the plural form of the T-form, MedHon singular, MedHon plural and LoHon plural are inflected identically. We can see this in examples (5a) and (5b), where the verb is inflected identically but the subject is NP plural in (5a) and singular in (5b). This inflection is different from the LoHon singular present in (3a).

(5) 3rd person masculine singular and plural

- a. *keta-haru kitaab padhchan.*
boy-PL book-SG read-3-PL-LOHON-M-PRES
- b. *keta kitaab padhchan.*
boy-SG book-SG read-3-SG-MEDHON-PRES
A boy/some boys read(s) a book. (lit. 'Boy(s) book read(s).')

On the surface, the Nepali system of honorification works like the T-V model, using pronouns and verb inflections to differentiate interpersonal relationships. However, the fact that the Nepali system is not binary, and that both the 2nd-person and 3rd-person carry honorification means that the T-V model is not sufficient in capturing the intricacies of the Nepali honorification. Still, scholars have tried to apply the paradigm to Nepali 2nd-person pronouns. Schmidt (1976) conducted a survey of 47 native speakers where she asked the participants which Nepali 2nd-person pronouns they would use in various social contexts. In this paper, she tried to fit 2nd-person honorification in Nepali into the previously established T-V model, grouping LoHon and MedHon in the category of T and HiHon and UHiHon in the category of V. The responses gathered show people mostly tended toward using LoHon, MedHon and HiHon, with the latter two being used most of the time. Schmidt took this tendency towards using the middle two levels of honorification as evidence of a T-V distinction in Nepali. However, precisely because the data did not fit into the two categories neatly indicates that a binary model such as T-V is not adequate for Nepali honorification, and rather needs to be understood as multi-faceted. Moreover, participants'

self-reported uses of such constructions might not be the most reliable reflection on the natural uses of Nepali honorification.

Van Driem (2019) has detailed the various factors that govern 2nd-person honorification in Nepali, specifically in the context of familial relationship, using Ethnographic data. Since kinship hierarchies in Nepali society are constructed upon aspects such as age, gender, relation, affinity, all of which affect respect and deference, these relationships are a rich source to explore honorification use. Van Driem defines pronominal usage being modulated by multiple axes of societal factors, all of which interact with each other producing very complex rules that govern the social discourse of the household. For example, a daughter-in-law may receive *timi* (2nd-person MedHon) and use *tapi* (2nd-person HiHon) when talking to her mother-in-law, but a son-in-law would both receive and use *tapi* (2nd-person HiHon) under equivalent circumstances. On the other hand, *t* (2nd-person LoHon) can be used affectionately with a child or a close friend of the same age but is derogatory if used without a context of established intimacy. However, honorification in Nepali society goes beyond these close ties of familial relations or friendships coloring all interactions of daily life.

So far, most previous work on the topic has explored 2nd-person honorification, with the focus being on the dynamics between the addresser and addressee (Agha 1994). This stems partially from the fact that honorification is particularly cogent between interlocutors interacting with each other. Another reason for this particular attention on 2nd-person honorification is because the prevalent T-V model was initially based on European languages that only carry the distinction of registers in 2nd-person and not in the 3rd-person. Furthermore, the focus on 3rd-person honorification in particular is important to explore how these relationships of deference persist beyond addresser-addressee interactions. The T-V model we have talked about thus has not so far been extended to pronouns in the 3rd-person, as it mostly deals with the interpersonal relationship between the addresser and addressee. This paper is not drawing directly from the T-V model of pronouns which might not apply to 3rd-person pronouns per se. Rather, it is analysing Nepali honorification through the underlying argument in Brown & Gilman's work as to how societal factors can cause a shift in the overall usage of one pronoun over another. Hence, this study will focus on the rich 3rd-person honorification in Nepali, and societal factors that can affect their usage.

1.3. NEPALI HISTORICAL CONTEXT. Since its inception in 1768 CE, Nepal had been a Hindu monarchy under the Shah dynasty for over 200 years (Thapa & Sharma 2009). In 1990, the country transitioned from an absolute monarchy to a multi-party democracy following the People's Movement. The early nineties saw the formation of the Communist Party of Nepal, Maoists (CPN-M), which propelled the country into a civil war in 1996. The war lasted 10 years and was fought for the purpose of abolishing the monarchy. During the time-frame of the war, King Gyanendra ascended the throne in 2001 following the massacre of his brother, then King Birendra, and his lineage. King Gyanendra's authoritarian policies and anti-democracy sentiments united the political parties across the spectrum who formed a Seven-Party-Alliance against the monarch. In April of 2006, the 240 year long regime of the Shah dynasty collapsed and monarchy was abolished in favor of a federal democracy. In a matter of a couple of decades, Nepal moved from a highly socially-stratified autocratic society to a democratic one. Over the course of 2 decades, the nation went through a momentous change in governance as well as changing attitudes to-

wards monarchy. With a more egalitarian ideology proliferating the zeitgeist as well as a change in the power structures, this paper explores how Nepali speakers might modulate the way they use honorification, especially when they reflect the very relationships of stratification that the country tried to shed in the revolution.

Alongside the anti-monarchy movement, Nepal also experienced a shift in women's rights movements in the same time-frame. The Maoist militants fighting against the Shah regime consisted of 40% women and touted egalitarian gender policies that went against the traditional patriarchal hegemony (Lohani-Chase 2008). Women were also severely disenfranchised and under-represented in Parliament, which only shifted following the interim constitution of 2007 (Nepal 2007), where women made up about 30% of the Constituent Assembly as opposed to the earlier 3 – 6% (Tamang 2009). Furthermore NGOs and INGOs played a huge role in gender equality movements in the country through various programs pertaining to equal access to education, employment and resources. While these movements gave rise to more egalitarian gender dynamics, they still did not and have not deconstructed the underlying structures of patriarchy, caste, class, and other factors still disenfranchising the women in the nation. Even so, the state of gender equality is certainly better than that of before the abolition of monarchy, with women being awarded almost equal rights under the constitution (Secretariat & Durbar 2015).

From the early 90s to the present day, Nepal went through such momentous changes that fundamentally changed the power structures both in the realm of governance and gender. There was a general, institutional move towards more egalitarian ideals that underlined all of these movements and brought Nepal to the point it is at today. With these changes in mind, this paper aims to explore whether the use of 3rd-person honorification has changed over this time-span. If there is a change in the usage over this time, in what direction would the use of honorification change? In a two pronoun system like those that Brown & Gilman's model there is a shift towards the use of one or the other in face of a more egalitarian movements. However, in a three pronoun system like that in Nepali how would such a pattern manifest, and would such a pattern appear in the absence of an addresser-addressee interface in the 3rd-person? This paper aims to explore not only the patterns of Nepali 3rd-person honorification with the broad social changes in mind but also the changing landscapes of gender dynamics.

1.4. PREDICTIONS FOR THE STUDY. When Brown & Gilman first proposed the T-V model, they framed it in terms of how macro-sociological changes and power relations are exhibited in the use of personal pronouns. One of the strongest examples of this shift is through the example of English, which leveled 2nd person pronoun use completely. As mentioned in section 1.1, Wales argues against such a strong relation between such a link. However, it is true that certain social changes precipitated into the leveling of the English pronominal system of address. Schupbach et al. (2007) conducted a cross-linguistic study on French, German, Italian and Swedish that demonstrated the impact of different social movements on T-V usage. This was most salient in the case of Swedish with the T-form being the primary pronoun of address after a late 1960s reform where companies started encouraging the use of only the T-form in various formal and informal contexts between employees. However in the cross-linguistic and cross-cultural contexts, the differing views on the use of T-V pronouns caused somewhat of a resurgence of the V-form. Schupbach et al. note that these pronouns are modulated with social change and often take on a cyclical pattern of leveling to one end or the other through time.

Historical context from other languages show that these pronominal systems often have a

tendency to level in the face of broad societal changes, especially if there is a shift from a stratified environment to a more egalitarian one. In this paper, the usage of 3rd-person honorification in three time periods will be measured: before, during, and after the Nepali Civil war. It should be noted that the patterns seen in the European languages appear far earlier than the time-frame that this paper is focused on. Furthermore, this very time-frame overlaps with the boom of the internet as well as massive globalization, which might affect the results. However, with previous context in mind, due to the higher level of social stratification in the pre-war era, it is more likely that this stratification would be reflected in the varied use of honorification during this time period. This pattern is likely to continue throughout the time period of the war, since the hierarchical structures are still institutionally maintained in spite of rising conflict, especially as the ruling class is trying to hold on to power. However, the post-war period is likely to yield a dramatic shift, as the system of governance has completely changed and ideologies of equality modern democracy has been instated which might influence preference of using one form of honorific over using multiple ones. As the society around them changes to shun previously salient hierarchies, speakers might not prefer indexing these very hierarchies in their language use. This leveling of honor might be reflected in both genders in the post war period.

However, in the time before and during the war, the gender differences may be more salient. van Driem (2019) has demonstrated that gender is a pertinent indicator of how honorification is used in Nepali familial relationships along with age and relative hierarchy. However, even beyond the realm of kinship relations, gender still plays an important role in informing societal discourse, including honorification. Lower honorific forms are more masculine-coded, with men using LoHon forms in both 2nd- and 3rd- person forms to refer to their peers or those socially lower than them (van Driem 2019). Women are under greater social pressure to be more deferential and even with their peers or those lower in the social hierarchy, they tend to use MedHon form as their lowest honorific form. Lakoff (1973) observes that women's speech often reflects their subservience in society, where their usage strong forms of expression is discouraged. Furthermore, women often occupy less powerful positions in society than men do, which means that they are under a much larger threat of disenfranchisement if they do not keep up with the societal norms of politeness (Holmes 2013).

Because of the societal expectation of having to use more polite forms than men do, women might also receive higher forms of honorification through reciprocation under the guise of politeness, despite being lower in the social hierarchy. Moreover women might both use and receive more honorific forms than men, coalescing around the MedHon level, while men will have higher levels of LoHon use. This avoidance of the lower honorific register for women due to societal expectations has also been seen in Basque (Bereziartua & Muguruza 2021), where the informal register referring to women is close to disappearing. The use of higher honorific forms for women might also indicate the link between honorification and social distance, where avoidance of the LoHon level with women signifies their alienation in society. So, before the monarchy was abolished, we might see women receiving more MedHon and HiHon forms as a reciprocation of the expectations to be more deferential than men. Whereas, men might have more instances of receiving LoHon in the same time-period due to the absence of such expectations.

2. Methodology. Past research on Nepali honorification has utilized mainly qualitative measures such as anthropological observation, surveys and interview analysis. This study aims to quantitatively analyze how patterns of honorification modulate in Nepali before, during, and after the

Civil war. For measuring how honorification in Nepali functions in a naturalistic usage context, a corpus analysis of Nepali text data was employed. An advantage of using corpus analysis is its high statistical power, as trends of usage over millions of words is measured. The benefit of this approach is that it would provide insight to the trends in the everyday realities of honorification. Moreover, a corpus study would allow for a temporal analysis, by elucidating trends of how the use of honorific verbal inflection has changed over time.

2.1. MATERIALS USED. One of the corpora used for this project is the Nepali National Corpus (NNC)(Yadava et al. 2008), which consists of 14 million words from papers, books, and websites spanning a wide variety of topics. The original data consists of works between 1990 through 1992, referred to hereafter as NNC-O (for NNC-Original). The extended works span 2000-2006, referred to hereafter as NNC-E (for NNC-Extended). This data is tagged with the NELRALEC tag-set, released by Yadava et al. (2008) with the NNC, on the basis of Parts of Speech (POS) specific to aspects of Nepali grammar. Since Nepali is a language where postpositions, possessives and case-markings are all suffixed to a noun-phrase and verbs reflect gender, honorification, person and number in agreement, it is not enough to tag each word as simply noun or verb. The NELRALEC tag-set instead identifies each of the syntactically and semantically prominent portions of a lexical item and classifies them separately. This means that each individual linguistic token is disambiguated on the basis of syntactic category, gender, person, honorification, punctuation as well as other factors. A single word such as *uni-haru-ko* (theirs', lit. "they-PI-Possessive") would thus have three tags under this tag-set. There are a total of 112 tags that are specific to each linguistic token. The tagging system is particularly useful for the project as it has already parsed the variables this paper is concerned with analyzing, namely honorification on the verb-form, and gender features on the nouns.

In addition to the Nepali National Corpus, a more recent corpus compiled by Rabindra Lamsal (henceforth referred to as the Lamsal corpus) containing 30 million words (6.5 million sentences) with works ranging up to 2020 (Lamsal 2020) was also incorporated in the data pool. This corpus pulls from the archives of various Nepali websites online. Unlike the NNC, however, the Lamsal corpus is untagged in any way except for the date of publication. In order to concatenate the two corpora, an automated tagger was used to tag the Lamsal corpus. NepBERTa (Timilsina et al. 2022), a BERT-based Devlin et al. (2018) model pre-trained on Nepali texts was trained on 90% of the NNC dataset and the remaining 10% was used for validation. The model had an error rate of 0.90% and was used to tag the Lamsal corpus.

2.2. DATA CLEANING AND PROCESSING PIPELINE. Only the verbs inflected in the third-person were extracted as this study focus on third-person honorification. There is a lack of a robust dependency-parser in Nepali to determine subject-verb pairs that could elucidate honorific relationships, due to Nepali NLP research being in its infancy. Instead, the following pipeline was used to identify the subject-NPs and its corresponding verb. The subject-NP in Nepali can either be marked with the ergative marker, *-le*, or be bare. All sentences with multiple *-le* markers were removed from the analysis to avoid ambiguous sentences with nested clauses. From the remaining sentences, the subject-NPs that were marked with *-le* were extracted. Next, a process of elimination was used to find sentences with bare subject NP, since subject-NPs in Nepali cannot have genitive, accusative, or any other post-position markings. If a sentence had multiple NPs that fulfilled such criteria, it was removed to decrease ambiguity. Next, since MedHon masculine verb inflection was identical to the plural inflection, all sentences with plural subject-NP candidates

were removed.

The NNC dataset does not tag the HiHon inflection as the form is a compound verb. The pattern for this form is regular: verb-stem + infinitive marker (*nu*) + conjugations of the verb “to be” so, specific strings including these features were used to filter HiHon from the verbs in 3rd-person. As gender information is only coded on the verb in the LoHon and MedHon levels, gender features of the subject in the HiHon condition needed to be extracted for the analysis. Since Nepali is a language where the feminine is the only gender marked differently from the masculine and neuter, feminine inflections were used to find feminine subject-NPs. First the data was checked to see if there were adjectives, determiners, genitive markers, reflexives right before the subject NP as these grammatical features agree in gender with the subsequent NP. After that pass, subject NPs themselves were inspected to ascertain whether they ended with the feminine marker -*ii*, barring the word *mantrii*, meaning minister, which is gender-neutral and a commonly used loanword from Sanskrit. Since HiHon can only apply to human entities, the rest of the non-feminine subjects had to be masculine.

Finally, the problem of animacy remained, as human LoHon masculine and non-human verb inflections are the same. To mitigate this issue, of possible human NPs was created by collecting the NPs that followed the Nepali classifier *-jana*, which can only refer to human entities. If the subject NP candidates in the sentences with LoHon masculine verbs were not part of the list of entities, the sentence was removed from analysis. While this pipeline sacrifices many ambiguous sentences resulting in fewer data, it is a reasonable workaround in absence of a true dependency parser.

3. Results.

3.1. DISTRIBUTION OF SELECTED SENTENCES. After going through the data cleaning pipeline, the percentage of sentences collected from each corpus was measured to ensure it was a fair sample. The percentages of total and selected sentences can be seen in Table 1, where the data of interest chosen all comprise similar ratios of the entire corpus they came from, despite the difference of size of each corpus. A chi-square test of independence was conducted to ensure that the ratio of not selected sentences to selected sentences from each corpus represented roughly the same distribution. The test yielded a chi-square value of 0.15 (p-value = 0.93), showing that these variables are independent of one another.

Corpus	Total Sentences	Selected Sentences	Selected Sentences %
NNC-O	59,668	10,838	18.16%
NNC-E	741,006	130,533	17.62%
Lamsal	4,124,157	665,624	16.14%

Table 1. Number and percentage of selected sentences from each corpus after filtering for sentences in the 3rd-person singular referring to human entities

The data was divided into sentences with masculine subjects and feminine subjects. Table 2 shows the number of sentences from the selected data based on gender. Overall, there are more sentences that have a masculine subject than a feminine one for each corpus. Another chi-square test of independence was performed which resulted in a chi-square value of 0.87 (p-value = 0.64), showing that the distributions are independent of one another.

Corpus	Masculine Subjects	Feminine Subjects
NNC-O	9,590	1,248
NNC-E	119,550	10,983
Lamsal	612,442	53,182

Table 2. Number of sentences selected with masculine subjects and feminine subjects

3.2. DISTRIBUTION OF HONORIFICATION. The dependent variable in this study are the three levels of 3rd-person honorification: LoHon, MedHon, and HiHon. Meanwhile, gender and time-span are the independent variables. Masculine and Feminine are the categories for gender. There are three categories for time-span, with each corpora analysed in this study corresponds to a time-frame of interest. NNC-O covers the Pre-War time-frame, NNC-E covers the War-Period time-frame and the Lamsal corpus covers the Post-War time-frame. Since the number of masculine subjects and feminine subjects is so different, the percentages will be taken from the total number of observations in the specified gender. Table 3 shows the proportions of honorification and gender across all three corpora. Overall, MedHon is the most commonly used honorification form, while HiHon is the least common. Looking specifically at HiHon in the Pre-War period, it is used proportionally twice as much with feminine subjects than is with masculine subjects. There is also a lower percentage of verbs in LoHon for feminine subjects as compared to masculine subjects, with a higher MedHon for sentences with feminine subjects.

The data in War-Period in Table 3 shows that there is a shift in pattern for honorification for sentences with feminine subjects. HiHon increases two-fold for both masculine and feminine subjects. However, the proportions of honorification in masculine subjects still remains fairly similar. On the other hand, MedHon for feminine subjects increases by almost 10%, with LoHon verbs decreasing by half. We see the most dramatic change in the Post-War period. For both masculine and feminine subjects, there is a tendency to level toward the MedHon, with over 90% of the data representing this honorific level for both genders.

	Pre-War		War-Period		Post-War	
Honor	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine
LoHon	45.69%	36.69%	40.99%	19.81%	2.71%	2.11%
MedHon	50.82%	57.05%	52.89%	67.92%	94.35%	92.95%
HiHon	3.49%	6.26%	6.11%	12.27%	2.93%	4.93%

Table 3. Percentage of honorification divided by gender before, during, and after the Civil War

Figures 1 and 2 demonstrate visually how the ratios of honorification change over time for feminine subjects and masculine subjects respectively. As can be seen from these figures both masculine and feminine subjects end up leveling to MedHon by the Post-War period. However, the shift in honorification patterns starts in the civil war period for the feminine subjects, while the ratios stay pretty similar for the masculine subjects in the same time period. LoHon tends to decrease quite a bit during the War-Period relative to Pre-War for feminine subjects, with an increase in both MedHon and HiHon.

A multinomial logistic regression was performed as the data in this paper has multi-class

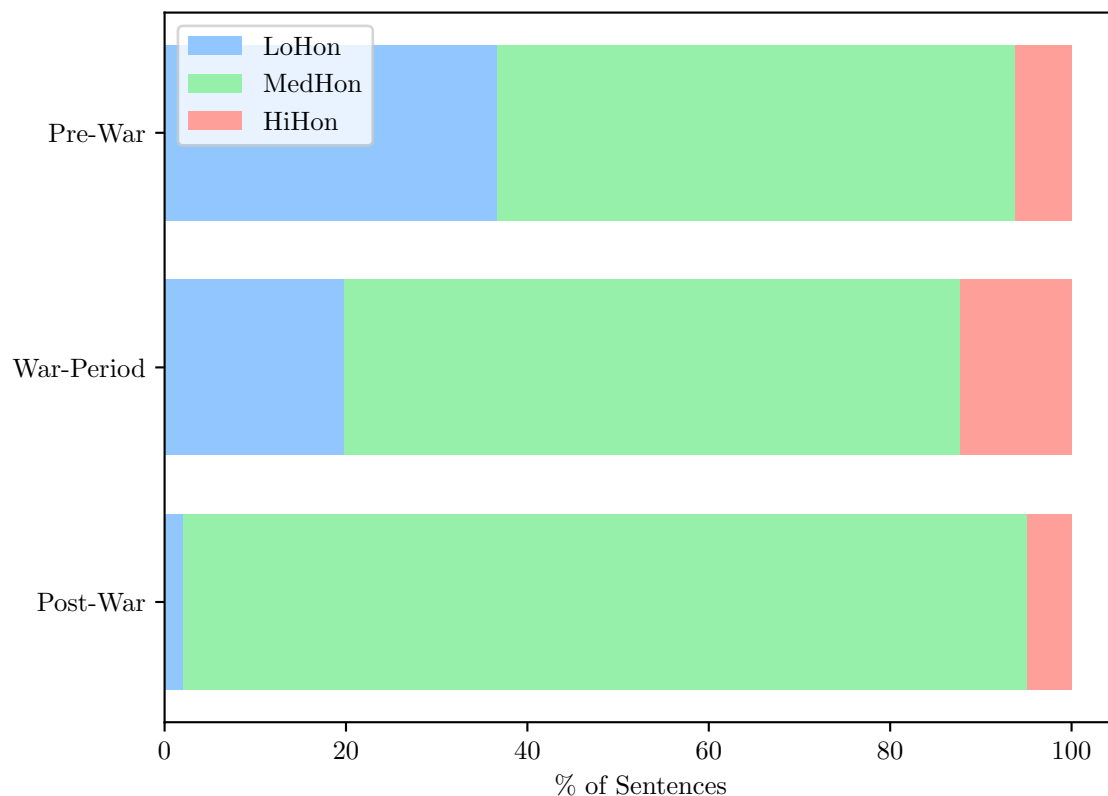


Figure 1. Proportion of different honorification in sentences with feminine subjects in the three time periods

variables, each with several discreet ordered categories. A model was created to study the relationship between the different levels of honorification as the dependent variable and gender and time-span as the independent variables. Since a multinomial regression is a series of logistic regressions between two variables in data with multi-class variables, a referent category needs to be fixed for each variable. MedHon was taken as the referent category for the honorification level. For gender, Feminine was taken as the referent category and for time-span, Pre-War was taken as the referent category. Table 4 shows the coefficients of the regression. Each coefficient shows the relationship between the category of the variable in the table and the referent category. A negative value indicates a decrease in the category with regard to the referent category while a positive values demonstrates an increase. The larger the value of the coefficient, the higher the magnitude of the change. The standard errors for the coefficients in LoHon row were around 0.06 and those for HiHon were around 0.12. Table 4 also shows the z-scores of the multinomial regression. Z-scores are the ratios of the coefficients to the associated standard error, with a higher absolute value signifying a more statistically significant coefficient. Also indicating that these results are statistically significant, are the p-values for the variables, which were less than 0.05.

The multinomial regression model used also included the interactions between the independent variables, with the same referent categories as before. Table 5 shows the coefficients, stan-

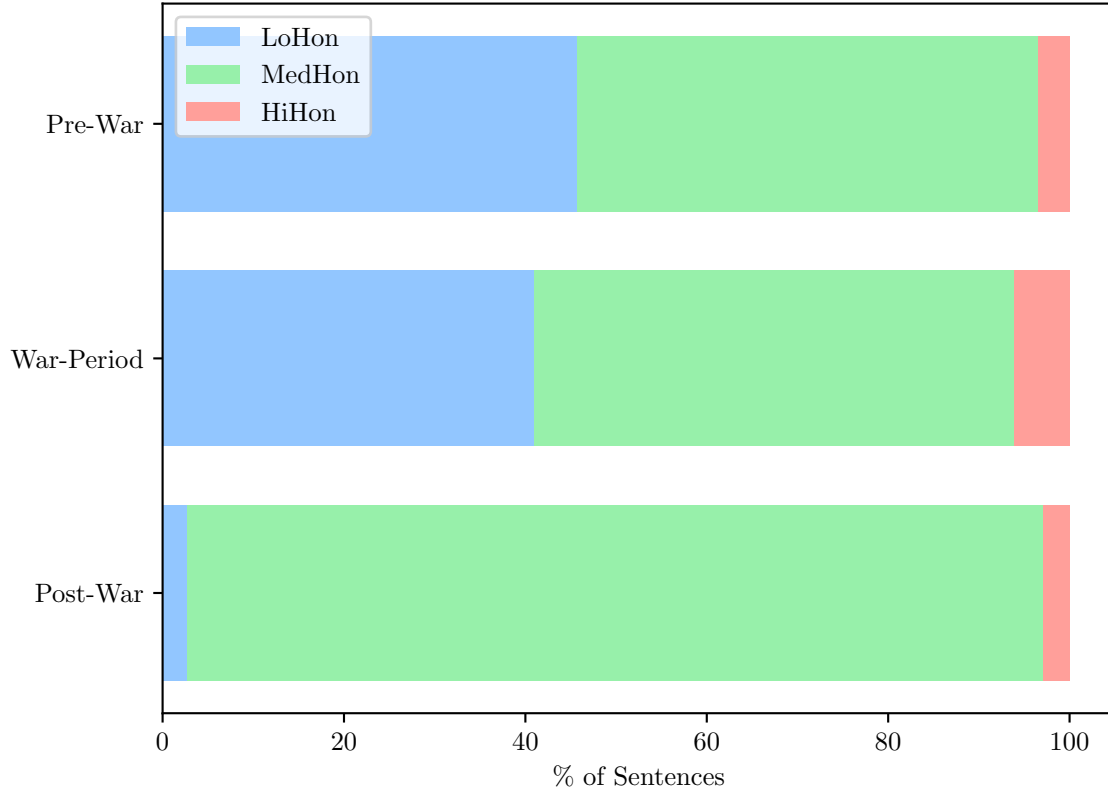


Figure 2. Proportion of different honorification in sentences with masculine subjects in the three time periods

standard errors and z-scores for the interactions between the variables. The standard errors were also 0.06 for LoHon coefficients and 0.12 for HiHon coefficients. Of all the different combinations, the only time we see a significant interaction effect of gender and time-span is LoHon relative to MedHon and Pre-War relative to War-period subset of the data. This particular interaction was the only statistically significant one out of all of the interactions, with a p-value of less than 0.05.

Because it was the only statistically significant interaction from the ones shown in Table 5, the interaction between gender and the War-Period was examined in more detail using a logistic regression. In this analysis, there were $2 \times 2 \times 2$ factors; LoHon and MedHon for Honorification level, Pre-War and War-Period for the Time-span, and Masculine and Feminine for Gender. The standard errors for the coefficients in this regression were around 0.02. All p-values for this regression suggested that the results were statistically significant, as they were lower than 0.05. The results of this regression suggest that Gender had a significant effect on the amount of MedHon and LoHon used between the Pre-War and the War-Period.

4. Discussion.

4.1. ANALYSIS. From the results we can see that speakers were using varied levels of honorification in time before the monarchy was abolished. This varied usage in the Pre-War and War-

Honor	Intercept		Masculine Subject		War-Period		Post-War	
	Coef.	Z-score	Coef.	Z-score	Coef.	Z-score	Coef.	Z-score
LoHon	-0.44	-7.36	0.33	5.28	-0.79	-12.23	-3.34	-49.83
HiHon	-2.21	-18.54	-0.47	-3.55	0.49	4.06	-0.72	-6.01

Table 4. Coefficients and Z-scores of multinomial regression. Baseline variables: MedHon for honorification, Pre-War for time-span, feminine for gender. p less than 0.05

Honor	Masculine Subject $\times$ War-Period		Masculine Subject $\times$ Post-War	
	Z-Score	Coef.	Z-Score	Coef.
LoHon	9.42	0.64	-1.40	-0.09
HiHon	0.16	0.02	-0.49	-0.07

Table 5. Coefficients and z-scores of interactions of time-span and gender. Baseline variables: MedHon for honorification, Pre-War for time-span, feminine for gender

Period times as compared to the Post-War period suggests that this specific aspect of language use might reflect the social stratification present in Nepali society at the time. Even so, most of the sentences used LoHon or MedHon with HiHon being used the least amount, accounting for at most 12% of the sentences. Since the sentences analyzed in this paper came from formal sources such as journals and articles, it could be that the lack of HiHon overall was not reflective of interpersonal relations between the writer and the subject. Rather, these forms were solely reserved to refer to a small percentage of the population: those in power.

There is a clear divide between the Post-War period and the time-frames before it, with the Post-War period demonstrating a clear change. While the Pre-War and War-Period saw significant use of both LoHon and MedHon for both genders, MedHon accounts for 90% of the data in the Post-War time-frame. This particular finding supports the hypothesis put forth that with the change in the system of governance towards a more egalitarian one, there is a shift toward preference of using one honorific form over using multiple forms that denote differing respect levels. The Post-War honor-leveling seems to be present despite the gender of the subject, indicating that this trend is a broader societal one. Moreover, gender did not have a statistically significant effect on the leveling to MedHon. This supports the hypothesis that despite any potential effect of gender before the Post-War period, it all levels to MedHon, due to the shift in wider society. The leveling in honor mirrors the patterns that have been seen in cases like English (Wales 1983), or Swedish (Schupbach et al. 2007). Thus the argument that Brown & Gilman put forth pronominal systems being modulated by societal and historical factors holds even in Nepali for 3rd-person pronouns.

While both masculine and feminine sentences demonstrated a pattern of leveling in the Post-War period, the data from the other time-periods show that there is a difference in how gender can affect how honorification is received. The hypothesis about women receiving fewer LoHon forms than men was also supported. In the Pre-War and War-Period time-spans, women had significantly fewer instances in proportion of being referred to in the LoHon form than men did. Women received 36% LoHon forms in the Pre-War era and 19% in the War-Period while men

	Coef.	Z-score
Intercept	-0.95	-32.75
Gender	0.89	38.69
Time-Frame	-0.21	-10.16

Table 6. Coefficients and z-scores of logistic regression between gender, Pre-War and War-Period for LoHon and MedHon

received 45% and 40% LoHon forms respectively. These results were found to be statistically significant as well. Women’s received honorification mostly coalesced around the MedHon level. Interestingly, they also received more HiHon forms than men did, across all three time spans even though HiHon forms were the least common across gender and time span.

The lower levels receiving LoHon forms for women lines up with Bereziartua & Muguruza’s findings in Basque, although it is not quite equivalent. The Basque patterns shows that the familiar form of address for women is disappearing, which does not seem to be the case for Nepali feminine LoHon forms. This pattern in fewer LoHon forms could be due to the aforementioned masculine indexicality of LoHon forms and the reciprocation of deference that is socially expected from women in Nepali society. We also see an interaction between gender and time-span between the LoHon and MedHon during Pre-War and War-Period. The number of LoHon forms for women decreases in the War-Period while there is an increased use of MedHon. This change in honorification suggests that feminine sentences were shifting towards honorification leveling prior to masculine ones. The earlier shift in honorification patterns in sentences with feminine subjects could be the result of rising gender-equality movements at the time described by Lohani-Chase and Tamang. As these conversations were gaining traction, the tendencies of those writing about women could have been starting to level as seen in the pattern Post-War.

4.2. LIMITATIONS. It should be noted that the data filtering pipeline described in the methodology was built to counter the lack of dependency-parsers in Nepali that could reliably extract exact subject-NP and verb pairs. As NLP research in Nepali is still in its infancy, multiple heuristics were used in lieu of a dependency-parser to approximate these subject-verb pairs of interest. In this process, quite a lot of the data had to be removed from analysis altogether to reduce ambiguity. These excluded data consisted of sentences with nested clauses and sentences with multiple candidates for subject NPs. Furthermore, because the list of human entities was built from NPs that followed the human classifier, there might be some human NPs that was not able to be captured by this pipeline. In the same vein, this round of filtering for human entities discards any data that used proper names, as the classifiers only appeared before common nouns. Pronouns also had to be excluded as potential subject NPs as the 3rd-person pronouns were under the same tag as unmarked determiners under the NELRALEC tag-set. Small random samples of the analysis data were manually verified to check the viability of the data. While there were no glaring errors, the results presented in the paper might not be completely representative of the true patterns in the language of the corpora.

Since most of the data collected in the corpora came from newspapers, journals, and books, the trends reflected in this paper are based not on honorific use in the entirety of the Nepali language but on the specific register of formal writing. The prevalence of the MedHon form in the post-war might be caused by the ideas of journalistic neutrality, of negotiating the differences

between the different honorific levels and landing in the middle. It is fitting then, that we see patterns reflecting more egalitarian ideals in precisely these formal texts. As these texts are written in the formal register, they represent the very institutions and systems that are the undercurrent for the changes in the society. Hence they reflect the egalitarian ideals that these institutions tout, which might not be representative outside of the realm of this specific register. In fact, it could be argued that precisely the persisting imbalance in the number of sentences with feminine subjects and masculine subjects demonstrates the unequal footing of gender in Nepali society that has yet to be eradicated.

Moreover, as with many other societies, there are many factors other than gender that color the interpersonal relationships of deference in Nepali society. These factors include but are not limited to socio-economic class, caste, ethnicity, and age. Nepal is a diverse society with a multitude of different ethnic groups and the politics that underlie the relations inter-and intra-group (Toffin 2009). While this paper has shown the wider societal trends of honorification in the formal register, it does so at the expense of other uses of the language. Nepali honorification is an intricate system that underpins many aspects of social life. For example, Upadhyay's paper on requestive acts demonstrates how the Nepali system on honorification is crucial in indexing politeness, rather than the Western notion of indirectness (Searle 1975). Thus, more research is needed on the field of honorification, wherein the of the system can be teased apart on multiple levels. A future direction could certainly be analyzing a spoken-word corpus where registers other than formal writing are explored. Furthermore, an elicitation task with native speakers might also elucidate the behavioural patterns of honorification use.

5. Conclusion. Relationships of respect and deference manifest themselves in different languages and cultures in various ways, of which the T-V pronominal model is one of the most studied. However, in a language such as Nepali which not only has more than two levels of honorification but also a completely different social context, the previously established model is insufficient to explain it. This paper analysed Nepali 3rd-person honorification in the light of changing gender landscape and the fall of monarchy. Despite all of the differences between the T-V model and the Nepali system, changes in societal institutions and attitudes seem to have an effect on the use of these honorification forms, at least when it comes to formal writing. The results in this paper demonstrate that changes in honorification in Nepali is modulated by complex factors such as gender dynamics, historical events and shifts in systems of governance. While the results in the paper does not capture the full scope of honorification in terms of interpersonal relationships, where it is perhaps the most salient, it does shed light on large macro-sociological trends in how societal changes can affect language use.

This paper focused on two types of relationships of power affecting honorification in Nepali. The aforementioned T-V model helps us explain why Nepali honorification leveled after the fall of monarchy. However, honorification as a linguistic phenomena is far more complex and depends on factors other than what Brown & Gilman purported: power and solidarity. The interaction with language and gender show that the expectation and reciprocation of politeness can also affect how honorification is used. Honorification, especially in low-resource languages like Nepali, has not been studied sufficiently. The many axes and intricacies in honorification elucidate important points about the interaction between society, language and the interpersonal bonds we form.

References

- Acharya, Jayaraj. 1991. *A descriptive grammar of Nepali and an analyzed corpus*. Washington, DC: Georgetown University Press.
- Agha, Asif. 1994. Honorification. *Annual Review of Anthropology* 23(1). 277–302.
- Bereziartua, Garbiñe & Beñat Muguruza. 2021. Basque informal talk increasingly restricted to men: The role of gender in the form of address ‘hika’. *Journal of Language and Linguistic Studies* 17(1). 360–376.
- Brown, Roger & Albert Gilman. 1960. The pronouns of power and solidarity. In Tomas A. Sebeok (ed.), *Style in language*, 252–281. Cambridge, MA: The MIT Press.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *CoRR* abs/1810.04805. <https://doi.org/10.48550/arXiv.1810.04805>.
- Dixon, Robert M. W. 1990. The origin of “Mother-in-Law vocabulary” in two Australian languages. *Anthropological Linguistics* 32. 1–56.
- van Driem, George. 2019. The dynamics of Nepali pronominal distinctions in familiar, casual and formal relationships. In Paul Bouissac (ed.), *The social dynamics of pronominal systems: A comparative approach*, 151–203. Amsterdam: John Benjamins.
- Friedrich, Paul. 1972. Social context and semantic feature: The russian pronominal usage. In John J. Gumperz & Dell H. Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*, 287–295. New York: Holt, Rinehart & Winston.
- Holmes, Janet. 2013. *Women, men and politeness*. New York: Routledge.
- Lakoff, Robin. 1973. Language and woman’s place. *Language in Society* 2(1). 45–79. <https://doi.org/10.1017/S0047404500000051>.
- Lamsal, Rabindra. 2020. A large scale Nepali text corpus. Dataset. <https://doi.org/md8w>.
- Lohani-Chase, Rama S. 2008. *Women and gender in the Maoist people’s war in Nepal: Militarization and dislocation*. New Brunswick: Rutgers University dissertation.
- Nepal, People Of. 2007. *The Interim Constitution of Nepal 2063 (2007)*. Government of Nepal: Ministry of Law, Justice, and Parliamentary Affairs.
- Schmidt, Ruth L. 1976. The Nepali system of honorific registers. *Kailash* 4(3). 213–225.
- Schupbach, Doris, John Hajek, Jane Warren, Michael Clyne, Heinz-L Kretzenbacher & Catrin Norrby. 2007. A cross-linguistic comparison of address pronoun use in four european languages: Intralingual and interlingual dimensions. Presentation. Annual Meeting of the Australian Linguistic Society in Brisbane.
- Searle, John R. 1975. Indirect speech acts. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts*, 59–82. Leiden: Brill.
- Secretariat, Constituent Assembly & Singha Durbar. 2015. *Constitution of Nepal 2015*. Kathmandu: Constituent Assembly Secretariat.
- Tamang, Seira. 2009. The politics of conflict and difference or the difference of conflict in politics: The women’s movement in Nepal. *Feminist Review* 91(1). 61–80. <https://doi.org/10.1057/fr.2008.50>.
- Thapa, Ganga B. & Jan Sharma. 2009. From insurgency to democracy: The challenges of peace and democracy-building in Nepal. *International Political Science Review* 30(2). 205–219. <https://doi.org/10.1177/0192512109102437>.
- Timilsina, Sulav, Milan Gautam & Binod Bhattarai. 2022. NepBERTa: Nepali language model

- trained in a large corpus. In Yulan He, Heng Ji, Sujian Li, Yang Liu & Chua-Hui Chang (eds.), *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, 273–284. Online only: Association for Computational Linguistics. <https://aclanthology.org/2022.aacl-short.34>.
- Toffin, Gérard. 2009. The *janajati/adivasi* movement in Nepal: Myths and realities of indigeneity. *Sociological Bulletin* 58(1). 25–42. <https://www.jstor.org/stable/23620834>.
- Upadhyay, Shiv R. 2003. Nepali requestive acts: Linguistic indirectness and politeness reconsidered. *Journal of Pragmatics* 35(10-11). 1651–1677. <https://doi.org/fsvrt3>.
- Wales, Kathleen M. 1983. *Thou and you* in Early Modern English: Brown and Gilman re-appraised. *Studia Linguistica* 37(2). 107–125. <https://doi.org/d2r2kv>.
- Wetzel, Patricia J. 2004. *Keigo in modern japan: Polite language from meiji to the present*. Honolulu: University of Hawaii Press.
- Yadava, Yogendra, Andrew Hardie, Ram Lohani, Bhim Regmi, Srishtee Gurung, Amar Gurung, Tony Mcenery, Jens Allwood & Hall Pat. 2008. Construction and annotation of a corpus of contemporary nepali. *Corpora* 3. <https://doi.org/10.3366/E1749503208000166>.