# Statistical Study of Significance of *Anusvāra* (ं, /ẽ/, ṃ) in Konkani

**Anonymous AACL 2025 submission**

## Abstract

A multilingual country like India which has a vast diversity of culture and languages. Having this, there is a huge scope to understand the nature, the structure and the properties of the Indian languages and their associated dialects. Konkani, one of the scheduled languages of India, has been rarely studied, statistically. This paper illustrates the significance of one of the key property of Konkani language, *i.e.,* an *Anusvāra* (ं, /ẽ/, ṃ). *Anusvāra* plays a significant role in the syntax and semantics of Konkani language. Paper discusses the occurrences, analysis, detection and importance of *anusvāra* in Konkani. Further, it statistically analysis on how *anusvāra* plays a pivotal role in differentiating Konkani from other sister languages.

## 1 Introduction

India is a multilingual country having several languages, which in turn has multiple dialects. In India, there are 22 scheduled languages and around 720 dialects[1]. Many of these dialects are on its way of extinction and some of them cease to exist. Very less study has been done on most of these dialects. There are several reasons for this and is out of the scope of this paper. On the same note to this, very less study has been done on some of the less popular scheduled languages of India. In this paper, the focus is on one of the rarely studied scheduled language of India, the Konkani language.

### The Konkani Language

Konkani is an Indo-Aryan language and is one of the official languages in India[2]. Konkani is written in five scripts *viz.,* Devanagari, Roman, Kannada, Malayalam and Perso-Arabic. Devanagari is an official script for Konkani. It was chosen as the script for Konkani by modern revivalists as well (Sardesai, 2004). As per the 2011 census of India, there were 22,56,502 Konkani speakers in India[3]. It is widely spoken in Western India's state of Goa and parts of Karnataka, Maharashtra and Kerala (Sardesai, 2006).

### Genesis of Konkani

Konkani belonged to the Indo-Aryan family which is a sister language to other Indian languages such as Marathi, Hindi, *etc*. The commonality among these languages is that all of them are written in the Devanagari script. Phonetically as well, Devanagari script is giving a better representation to the contrasts significant to Konkani such as the dental v/s retroflex, the unaspirated v/s aspirated (Dalgado, 1893; Sardesai, 2004).

---

[1] https://www.justlanded.com/english/India/India-Guide/Language/Languages-in-India

[2] https://wikivividly.com/wiki/Konkani_alphabets

[3] http://www.censusindia.gov.in/2011Census/Language_MTs.html

| | FRONT | | CENTRAL | BLACK | |
|---|---|---|---|---|---|
| | unrounded | | | rounded | |
| | long | short | short | long | short |
| CLOSE | ई $/$ iː $/$ | इ $/$ ɪ $/$ | | ऊ $/$ uː $/$ | उ $/$ ʊ $/$ |
| CLOSE MID | ए $/$ eː $/$ | ऎ $/$ e $/$ | अ $/$ ə $/$ | ओ $/$ oː $/$ | ऒ $/$ o $/$ |
| OPEN MID | ऐ $/$ ɛː $/$ | | | | औ $/$ ɔː $/$ |
| OPEN | ऍ $/$ æ $/$ | | | | आ $/$ aː $/$ |

**Table 1:** Vowels in Konkani

Konkani is predominantly spoken in the state of Goa and some communities in neighboring states. Across this geographical area, there is a dialectical change in Konkani. Konkani language has shown a lot of tendency towards standardization by producing school textbooks, novels, stories, newspaper, teaching in Konkani in Devanagari script at the degree and the university levels. All this opened up opportunities for standardization of Konkani (Sardesai, 2004). Konkani language is rich with respect to its grammatical properties as well.

Table 1 gives vowels in Konkani and Table 2 gives the consonants in Konkani. Konkani has set of 12 (+1 *i.e.,* ओ) vowels and 34 consonants[4]. While, च (*ca*) /tʃ/, ज (*ja*) /dʒ/ and फ (*pha*) /f/ are letters from loanwords[5] of Arabic and Persian origin *e.g.,* काफ़ी, हज़ार, (*kāphaī, hajaāra,*) *etc.* Also क्ष (*kṣa*) and ज्ञ (*jña*) are combined consonants of guttural, non-aspirated, non-voiced, stop consonant क (*ka*) and retroflex, aspirated, non-voiced, fricative ष (*ṣa*) and palatal, non-aspirated, voiced, stop consonant ज (*ja*) /ɟ/ and palatal, non-aspirated, voiced nasal ञ (*ña*) /ɲ/ respectively. The vowels [ə], [ɪ], [ʊ] are always short in length, while the vowels [ɑː], [iː], [uː], [eː], [oː], [ɛː], [ɔː] are always considered long. In these vowels, short, central a is known as schwa in which the unstressed central vowel represented by the symbol (ə) in the International Phonetic Alphabet(IPA)[6]. In Konkani, schwa implicit is mainly in the written scripts obligatorily deleted for correct pronunciation. For instance, करता is *kartā* [kər'tɑː], and not *karatā*. But ऋ (*r̄*) and ऌ (*l*) are rarely used in Konkani except to render *tatsam* words *e.g.* कॢप्ती (*kḷptī*)/kl̩pt̪iː/, ऋष्यशृंग (*ṛṣyaśṛmga*) /ɹ̩sjəɕm̩ŋgə/ and hence, these vowels are not included in the vowel table.

Texts in Konkani are usually written in a standard script. However, there are texts which are written in other scripts as well. Also, some texts have a dialectical influence in both, the text data and the speech data. Most of the Konkani text is available in both Devanagari and Roman (*romi*) scripts. Please note that the *romi* script is not included in the scope of this paper due to time and space constraint.

There are various dialects of Konkani, spoken across different regions, *viz.*, *pednekari*, *bardeskari*, *saashti*, *malvani*, *antruz*, *chitpavni*, *karwari*, *manglorean*, *etc.* Among these, the dialects from Pernem ( *i.e.*, *pendnekari*), Canacona (*i.e., kankonkari*), *Saashti*, Bardez (*i.e., Bardeskari*) or Antruz (*i.e., antruz*) are easily identifiable. Apart from this, a Goan variant of Konkani (as well as a Roman variant of Konkani) is easily distinguishable from a non-Goan variant such as *Karwari*, *Mangalorean* or a Cochin variant. Thus on the basis of speech, a Konkani speaker could be situated on the regional map of Konkani without much difficulty (Sardesai, 2004, 2006).

Further, the Konkani speaker usually utters a word in a nasalized form. This differentiates a Konkani speaker from a non-Konkani speaker. The nasalized form is due to presence of *anuswara* as a property in Konkani. *Anuswara* plays a vital role in the Konkani language, in both speech and text processing. This has been discussed in details in next sections. This paper presents the significance of *anusvāra* in Konkani.

Rest of the paper has following parts: Sec-

---

[4]https://archive.org/details/akonkanigrammar00maffgoog

[5]http://www.ruf.rice.edu/~kemmer/Words/loanwords.html

[6]https://wikivividly.com/wiki/Konkani_alphabets

| | ST | | | | NS | AP | FR | | AF | |
|---|---|---|---|---|---|---|---|---|---|---|
| | -vo | | +vo | | | | -vo | +vo | -vo | +vo |
| | -asp | +asp | -asp | +asp | -asp | | +asp | +asp | -asp | -asp |
| GU | क /k/ | ख /kʰ/ | ग /g/ | घ /gʱ/ | ङ /ŋ/ | | | ह /ɦ/ | | |
| PL | च /c/ | छ /cʰ/ | ज /ɟ/ | झ /ɟʱ/ | ञ /ɲ/ | य /j/ | श /ɕ, ʃ/ | | | |
| RX | ट /ʈ/ | ठ /ʈʰ/ | ड /ɖ/ | ढ /ɖʱ/ | ण /ɳ/ | र /ɾ/ | ष /ʂ/ | | | |
| DN | त /t/ | थ /tʰ/ | द /d/ | ध /dʱ/ | न /n/ | ल /l/ | स /s/ | | | |
| LB | प /p/ | फ /pʰ/ | ब /b/ | भ /bʱ/ | म /m/ | व /ʋ/ | | | | |
| AL | | | | | | | | | च /tʃ/ | ज /dʒ/ |
| LD | | | | | | | | फ /f/ | | |
| RL | | | | | | ळ /ə/ | | | | |

**Table 2:** Consonants in Konkani

*(GU. Guttural; PL. Palatal; RX. Retroflex; DN. Dental; LB. Labial; AL. Alveolar; LD. Labiodental RL. Retroflex lateral flap)*

*(ST. Stop; NS. Nasal; AP. Approximant; FR. Fricative; AF. Affricative)*

*(-vo: -voiced; -asp: -aspirated; +vo: +voiced; +asp: +aspirated)*

tion 2 discusses *Anusvāra* in Konkani; Section 3 discusses on experiments and statistics conducted to understand why *anusvāra* is significant to Konkani; Section 5 concludes the paper with scope for future work.

## 2 *Anusvāra* in Konkani

*Anusvāra* is one of the important symbol for writing system in Indo-Aryan languages. It provides the nasalized sound and is represented as ◌ं *(/õ̃ /, ṃ)* in Devanagari. The languages which predominantly uses *anusvāra* are Sanskrit, Hindi, Marathi, Konkani, Nepali, *etc*. In Konkani, the *anusvāra* is traditionally defined for representing a nasal stop homorganic to a following plosive, (*anunāsika*) and also vowel nasalization[7]. It is an important property of formation of Konkani language syntax and semantics.

In this paper we investigate the usage and importance of *anusvāra* in Konkani language. The corpus is taken from Konkani Wordnet[8] (Walawalikar et al., 2010). While Marathi

WordNet[9], Hindi WordNet[10] and Sanskrit WordNet[11] (Kulkarni et al., 2010) has been used as a comparable corpus. All these wordnets are developed under the umbrella of IndoWordNet (Bhattacharyya, 2017). A WordNet is an online lexical resource which has a set of synonyms (synsets), concept definition (gloss), example sentences (word usage) and is categorized based on the four parts-of-speech categories *viz.*, Noun, Verb, Adjective, Adverb (Miller, 1998).

**Examples of *anusvāra* in Konkani**

Some of the examples of words in Konkani having *anusvāra* are - अर्दें (*ardem*, an half), अळमें (*aḷamem*, a mushroom), केळें (*keḷem*, a banana), खरें (*kharem*, a truth), धवें (*dhavem*, white), फकांडां (*phakāṃdāṃ*, fun), बरें (*barem*, good), आतांचें (*ātāṃcem*, of now), उकतें (*ukatem*, open).

Some examples of usage of *anusvāra* words in Konkani are:

1. बोकड्यो शेतांत रोंप्यांचीं पानां खातात (*bokaḍyo śetāṃta roṃpyāṃcīṃ pānāṃ khātāta*, Goats are eating plant leaves in a field).

2. स्वतंत्र भारतांत आतां अठ्ठावीस राज्यां जाल्यांत (*svataṃtra bhāratāṃta ātāṃ aṭhṭhāvīsa rājyāṃ jālyāṃta*, Independent India has now twenty-eight states).

These examples have most of the words having an *anusvāra* in them. We have identified two types of *anusvāra*:

- An *anusvāra* which is borrowed or present in the *tatsam* words[12] and *desi* words[13]. For our convenience, let us call this type of *anusvāra* as *tadanusvāra*.

---

[7] http://www.wikiwand.com/en/Konkani_alphabets

[8] http://konkaniwordnet.unigoa.ac.in/

[9] http://www.cfilt.iitb.ac.in/wordnet/webmwn/index.php

[10] http://www.cfilt.iitb.ac.in/wordnet/webhwn/index.php

[11] http://www.cfilt.iitb.ac.in/wordnet/webswn/index.php

[12] http://www.hindikunj.com/2017/10/tatsam-tadbhav.html

[13] anusvāra

- The Konkani specific *anusvāra* which we can find in words of Konkani language only. Let us call this type of *anusvāra* as *konkanusvāra*.

In the sentences 1 and 2 mentioned above, a word स्वतंत्र (*svataṃtra*) has *tadanusvāra* and words शेतांत, पानां, भारतांत, आतां, राज्यां, जाल्यांत ( *śetāṃta, pānāṃ, bhāratāṃta, ātāṃ, rājyāṃ, jālyāṃta*) has *konkanusvāra*. However, word रोंप्यांची (*roṃpyāṃcīṃ*) has both the types *i.e.*, 1ˢᵗ and 3ʳᵈ *anusvāra* are *konkanusvāra* and 2ⁿᵈ *anusvāra* is a *tadanusvāra*. This suggests that the *konkanusvāra* appears more often than the *tadanusvāra* in the Konkani text.

### *Anusvāra* in Types of Words in Konkani

There are basically four types of words in Konkani *viz.*, Original Konkani, *Tatsam*, *Desi* and Loanwords. Table 3 shows the type of words in Konkani with few examples, all having *anusvāra*. Here, *Tatsam* words are words directly taken as it is from Sanskrit language. *Desi* words are words from sister languages like Hindi, Marathi, *etc.*. Loanwords[14] are words which are taken from foreign languages. The original Konkani words are those words which are specific to Konkani language and are frequently used in this language itself.

### Phonological Properties of *Anusvāra*

The following are the rules for assigning *Anusvāra* when vowel becomes nasalized (Maffei, 1882; GoaKonkaniAkademi, 2007):

- Between a long vowel and a voiced stop - [/aː/ + / b/, /d/, *etc.*] = [/aː̃/ /b/, /d/, *etc.*]. Examples, तांबो (*tāṃbo*, [taːmbo], copper), चांदी (*cāṃdī*, [caːndiː], silver).

- Between a long vowel and a voiceless stop = [/aː/ + /ṭ/, /c/, *etc.*] = [/aː̃/ /ṭ/, /c/, *etc.*]. Examples, दांत (*dāṃta*, [daːnt], tooth), कांच (*kāṃca*, [kaːɲc], glass).

---

<sup>14</sup>https://blogs.transparent.com/hindi/loan-words-in-hindi/

| | |
|---|---|
| **Tatsam words** | चिंतन, वंशज, मंद, संज्ञा, संदेश, संस्कार *(ciṃtana, vaṃśaja, maṃda, saṃjñā, saṃdeśa, saṃskāra)* |
| ***Desi* words** | लोखंड, अंजीर, अपंग, भंडार, मंजुरी *(lokhaṃḍa, aṃjīra, apaṃga, bhaṃḍāra, maṃjurī)* |
| **Loanwords** | जहांपना, कांगारू, अंदाज, डायटींग, डिसेंबर *(jahāṃpanā, kāṃgārū, aṃdāja, ḍāyaṭīṃga, ḍiseṃbara)* |
| **Original Konkani words** | हरवें, फुडें, भोंवडी, सोंपे, हांसरें, संवसार, तांतीं, गुदावं, मांव, खंय *(haraveṃ, phuḍeṃ, bhoṃvaḍī, soṃpe, hāṃsareṃ, saṃvasāra, tāṃtīṃ, gudāvaṃ, māṃva, khaṃya)* |

**Table 3:** Types of Konkani words

- Between a short vowel and an obstruent. Example, आंकडो (*āṃkaḍo*, [aᵊ̃kədo], to support).

- Preceding voiced, non aspirated, approximate, repeated dental /l/ final long e vowel = /ll/ + /eː/ = /ll/ + /eː̃/. Example, कल्लें (*kalleṃ*, a type of utensil), केल्लें (*kelleṃ*, done).

- Preceding voiced, non aspirated, nasal, retroflex /ṇ/ final a long e vowel = /ṇ/ + /eː/ = /ṇ/ + /eː̃/. Example, उणें (*uṇeṃ*, less), उताणें (*utāṇeṃ*, up side down)

- Preceding non voiced, non aspirated, guttural & palatal stop consonant with final a long e vowel = /k/,/c/ + /eː/ = /k/,/c/ + /eː̃/ = /keː̃/, /ceː̃/. Examples, दावकें, दुदकें, धुकें and अंकाचें, अखेरचें, अग्रीचें (*dāvakeṃ, dudakeṃ, dhukeṃ* and *aṃkāceṃ, akheraceṃ, agnīceṃ*).

- Preceding non voiced, aspirated, fricative, palatal /ʃ/ consonant with final a long e vowel= /ʃ/ + /eː/ = /ʃeː̃/. Example, उशें, कशें, कांशें (*uśeṃ, kaśeṃ, kāṃśeṃ*).

- Neuter nominative singular Nouns, adjective = final long e vowel = /eː/ - /eː̃ /. Examples of Nouns, केळें (*keleṃ*,

[kelẽ], banana), खोरें (*khorem*). Examples of Adjectives, कुसकें (*kusakem*), उताणें (*utāṇem*).

- Feminine nouns in ending long i vowel = /iː/ -> /ĩː/. E.g., बोकडी (*bokaḍī*, Goat).

- In plural form of neuter noun in ending long, mid, front, nasal e vowel changes into long, close, front, nasal i vowel = /eː̃ / -> /ĩː/. Examples, केळें (*keḻem*, a banana)- केळीं (*keḻīm*, bananas), तवशें (*tavaśem*, a cucumber) - तवशीं (*tavaśīm*, cucumbers).

## 3 Discussions: Experiments and Statistics

To understand the occurrence of *anusvāra* in Konkani, a statistical experiment was conducted using Konkani WordNet data and Marathi WordNet data. Example sentences from both these wordnets were taken and sorted based on number of words per sentence (grams). There were 24785 Konkani sentences and 30336 Marathi sentences in their respective wordnet corpus. These sentences were categorized into 2-grams, 3-grams, 4-grams, 5-grams, 10-grams, 15-grams and 20-grams.

Further, the number of *anusvāras* per sentence was calculated from each of these categories. The figures can be seen in table 4. The occurrence of *anusvāra*, on an average, in Konkani (on LHS) and in Marathi (on RHS) sentences with respect to the figures in table 4 can be analyzed and presented as:

- 2-gram: 0 to 1 >= 0 to 1

- 3-gram: 0 to 1 >= 0 to 1

- 4-gram: 1 to 2 > 0 to 1

- 5-gram: 1 to 2 > 0 to 1

- 10-gram: 3 to 4 > 1 to 2

| Length of the sentence | Average occurrence of *anusvāra* in Konkani sentences | Average occurrence of *anusvāra* in Marathi sentences |
|---|---|---|
| 2-gram | 0.19 | 0.11 |
| 3-gram | 0.66 | 0.35 |
| 4-gram | 1.03 | 0.58 |
| 5-gram | 1.47 | 0.62 |
| 10-gram | 3.00 | 1.18 |
| 15-gram | 3.54 | 1.66 |
| 20-gram | 4.50 | 1.75 |

**Table 4:** Occurrence of *anusvāra* in n-gram sentences in Konkani and Marathi

| Part-of-speech category | Konkani WordNet words | Words with *anusvāra* | Occurrence of *anusvāra* in words (%) |
|---|---|---|---|
| Noun | 34805 | 8271 | 23.76 |
| Verb | 7003 | 1738 | 24.82 |
| Adjective | 9794 | 4456 | 45.50 |
| Adverb | 1273 | 450 | 35.35 |
| Total | 52875 | 14915 | 28.21 |

**Table 5:** Percentage of occurrence of *anusvāra* in words from **Konkani** Wordnet

- 15-gram: 3 to 4 > 1 to 2

- 20-gram: 4 to 5 > 1 to 2

In all the above cases, it is observed that occurrence of *anusvāra* is more prominent in Konkani sentences than in Marathi sentences. Hence, we can say that *anusvāra* is more significant to Konkani language than to the Marathi language.

In a second experiment, a set of all the words from Konkani WordNet were taken in one pool and a set of all words from Marathi WordNet were taken in another pool. In this experiment, all the Konkani words with *anusvāra* (having both *tadanusvāra* and

| Part-of-speech category | Marathi WordNet words | Words with *anusvāra* | Occurrence of *anusvāra* in words (%) |
|---|---|---|---|
| Noun | 39350 | 7236 | 18.39 |
| Verb | 5975 | 670 | 11.21 |
| Adjective | 10092 | 1500 | 14.86 |
| Adverb | 1273 | 132 | 10.37 |
| Total | 56690 | 9538 | 16.82 |

**Table 6:** Percentage of occurrence of *anusvāra* in words from **Marathi** Wordnet

5

| Types of words | Percentage of Occurrence of *anusvāra* in Konkani WordNet words (%) |
|---|---|
| *Tatsam* words | 2.20 |
| *Desi* words | 14.32 |
| Loanwords | 1.23 |
| Original Konkani words | 82.25 |

**Table 7:** Percentage of Konkani with *anusvāra* w.r.t. the types of words

*konkanusvāra*) were calculated. Similarly, all the Marathi words having *anusvāra* were calculated. Refer table 5 for Konkani statistics and table 6 for Marathi statistics. It is observed that out of 52875 Konkani words, 14915 words have *anusvāra*. However, for Marathi data, out of 56690 Marathi words, 9538 words have *anusvāra*. That is, Konkani data has about 28.21% *anusvāra* words while Marathi data has about 16.82% data, which is less as compared with Konkani and considering the total words in Marathi as well. Here also, we can say that *anusvāra* is an important property of Konkani language.

In the third experiment, a set of all *anusvāra* words from Konkani were taken and checked whether they belong to any of the four types of Konkani words (*viz.*, *tatsam* words, *desi* words, loanwords, original Konkani words). *tatsam* words were calculated by comparing this Konkani set with Sanskrit WordNet words, *desi* words were calculated by comparing this set with Marathi and Hindi WordNet data and loanwords were calculated manually. The remaining words were considered to be Konkani specific and validated manually as well. In this experiment, it is observed that out of these 14915 *anusvāra* words in Konkani, there are very few loanwords (about 1.23%), few *tatsam* words (around 2.20%), about 14.32% *desi* words, while about 82.25% words are the original Konkani words. This infers that words with *anusvāra* is an essential ingredient of Konkani literature. Also, *konkanusvāra* is more frequently used than the *tadanusvāra* in these words.

Considering all these experiments, we can say that *anusvāra* definitely plays a significant role in Konkani language.

## 4 Limitations

The paper presented the role of *anusvāra* in the Indian Language, Konkani. However, the authors have considered a limited data which is available for this low-resourced lanuage, Konkani. The authors could not present all the statistical analysis which shows the significance of *anusvāra* in Konkani, however it has covered sufficient statistical data which can surely help in colcluding the results. Authors agree that Results and Discussions section could have been elaborated in more details.

## 5 Conclusion and Future Work

The paper presents the role of *anusvāra* in the Indian Language, Konkani. The tables presented in this paper illustrates the significance of *anusvāra* in Konkani. It is observed that Konkani language can stand by itself and different than other Indian sister languages because of the presence of nasalization, which is made prominent by the usage of *anusvāra* in the Konkani language. Hence, we conclude that an *anusvāra* plays a significant role in the Konkani language.

In future, we shall work on *romi* Konkani script and along with other dialects of Konkani. Also, we shall conduct experiments to check if *anusvāra* remains intact across the dialectical geographical area or it goes away and whether it semantically and/or syntactically influences the underlying dialect. Speech processing is one of the major area which is ongoing and is being experimented.

## References

Pushpak Bhattacharyya. 2017. Indowordnet. In *The WordNet in Indian Languages*, pages 1–18. Springer.

Sebastiao Rudolpho Dalgado. 1893. Dic-cionario komkani – portuguez.

GoaKonkaniAkademi. 2007. Konkani shuddhalekhanache nem.

Malhar Kulkarni, Chaitali Dangarikar, Irawati Kulkarni, Abhishek Nanda, and Pushpak Bhattacharyya. 2010. Introducing sanskrit wordnet. In *Proceedings on the 5th Global Wordnet Conference (GWC 2010), Narosa, Mumbai*, pages 287–294.

Angelus Francis Xavier Maffei. 1882. Konkani grammar.

George Miller. 1998. *WordNet: An electronic lexical database*. MIT press.

Madhavi Sardesai. 2004. Mother tongue blues. In *India Seminar*, pages 44–51. Seminar - Web Edition, The monthly symposium, New Delhi.

Madhavi Sardesai. 2006. *A comparative linguistic and cultural study of lexical influences on konkani*. Ph.D. thesis, Goa University.

Shantaram Walawalikar, Shilpa Desai, Ramdas Karmali, Sushant Naik, Damodar Ghanekar, Chandralekha D'Souza, and Jyoti Pawar. 2010. Experiences in building the konkani wordnet using the expansion approach.