

Plural Problems in the Nominal Morphology of Marathi*

Shalmalee Pitale^a, Vaijayanthi Sarma^a

^aDepartment of Humanities and Social Sciences, IIT Bombay,
Powai, Mumbai-400076, India
{shalmaleep, vsarma}@iitb.ac.in

Abstract. In this paper, we describe the two tests developed and designed for Marathi using *non-words*, a) plural formation for *non-words* b) intuition test for gender assignment in which subjects were asked to assign gender to non-words. We look at the distribution of nouns across noun classes and genders and discuss the congruence between the problematic classes as observed in the tests and the actual class distribution and frequency in the language.

Keywords: Marathi, plurals, gender, non-words, morphological rules

1 Introduction

Marathi, an Indo Aryan language, makes a two-way number distinction, singular and plural. The number feature is synthetically marked on the noun itself. Two linguistic factors determine the plural formation of Marathi nouns: (a) grammatical gender, whether a noun is masculine, feminine or neuter and (b) the final segment of the noun. As a part of our ongoing research on the acquisition of Marathi morphology, we developed tests to see whether Marathi speaking children and adults apply tacitly known rules of inflectional morphology to new-coined words.

One of the tests we developed was designed for plural formation with *non-words*. In this test, subjects were required to choose the correct plural suffix for a given non-word. Ample agreement cues for gender were also provided. This test was adapted from Jean Berko's classic study on the acquisition of the inflectional morphology of English (1958) and Bettina Spreng's work on the acquisition of plural morphology in German (2004). We also developed an intuition test for gender assignment in which subjects were asked to assign gender to non-words to measure the effect of the phonological shape of the token on the choice of gender. This test was inspired by A. Mill's (1986) work on acquisition of gender in German.

The results of these two tests indicated that the performance of the subjects, both adults and children, is better for some noun classes compared to others. In other words, phonological cues were reliable for some of the noun classes and facilitated gender choice, but this was not a reliable or unique cue, and led us to study the distribution of nouns into gender classes in Marathi.

In this paper, we describe the two tests and analyze their results in brief. We look at the distribution of nouns across the noun classes and genders and discuss the congruence between the problematic classes as observed in the tests and the actual distribution of the same in the language.

*We would like to thank the three anonymous reviewers for their reviews. We have tried to address their comments and incorporate their suggestions in this version. Any errors are, of course, the authors.

2 Testing Plural morphology

Singular/bare forms of countable nouns in Marathi may end in a vowel or a consonant. The phonological ending of different nouns and their corresponding plural suffixes across all three genders in Marathi are given with examples in Table 1. It may be seen that there is significant overlap of both the phonological form and affixes across classes. Some nominal shapes are restricted to a single gender class. Thus, -ə or -u ending nouns are always masculine and e/ə alternating nouns are always neuter. However such one-to-one map is not seen with other shapes. Consonant final nouns could potentially belong to any of the three gender classes. Further, -a could either mark the bare noun shape (feminine or masculine) or the plural form of certain classes (the C or ī ending feminine nouns). The cells in grey indicate null classes, for example, there are no feminine or neuter nouns that end in -ə or in -u.

Table 1: Noun classes in Marathi¹

Gender	Masculine	Feminine	Neuter
Singular	Plural		
ə	∅ mitrə ‘friend’		
C	∅ hat ‘hand’	C → a ī (i) vaT → vaTa ‘path’ (ii) vihir → vihirī ‘well’	C → e/ə ghər → ghəre/ə ‘house’
ī	∅ mənī ‘bead’	ī → a nədī → nədya ‘river’	
ū	∅ laDū ‘a sweet’	∅ or a (i) vəstū ‘thing’ (ii) pisū → piswa ‘a mosquito’	ū → e/ə limbū → limbe/ə ‘lemon’
a	a → e vaDa → vaDe ‘mansion’	∅ šaLa ‘school’	
e/ə			e/ə → ī təLe/ə → təLi ‘pond’
u	∅ šətru ‘enemy’		

Based on this classification of nouns, we developed the tests for plural formation. Subjects were asked to provide appropriate plurals of given tokens, all *non-words*. Testing with non-words show us how subjects treat words for which they have no prior knowledge and permits inferences about the default assumptions that are made about words in the language and how morphological rules are applied. The picture material and procedures for the study were adapted from Berko’s *wug* test (1958). While gender was (and is) not directly marked on nouns, it is marked morphologically through gender agreement. Agreement cues were given to facilitate the choice of gender and the selection of the corresponding plural form. The test groups consisted of ten Marathi speaking (invariably bilingual) adults (5 male, 5 female) and ten (monolingual or marginally bilingual) Marathi-speaking children (5 male, 5 female).²

Table 2 contains the results of the tests for both groups of subjects per gender class. Some of the tokens used are also provided per class.

¹ The capital letters indicate retroflex consonants and grey cells indicate absence of lexical items in that class.

² While the sample group appears small, it is usually sufficient to determine how rules are used by speakers. We do plan to test with a larger test group to address concerns of sample size quirks, once we have a general understanding of the kinds of difficulties that Marathi morphology poses for its speakers.

Table 2: Results for test of *non-words*

Gender/ Noun Class	Number		Token	Results (%)	
	Sg	Pl		Children sg to pl	Adults sg to pl
M1	ə	no change	pūmbə	100	100
M2	C		tʃag	100	100
M3	ī		pəNī	100	100
M4	ū		maTū	100	100
M5	a	e	tuDa	80	80
F1	a	no change	gaTa	100	100
F2	ū		bhakū	100	100
F3	ī	a	sakvī	60	100
F4	C	a	raT	10	30
F5	C	ī	Dhak	0	20
N1	C	ə/e	khaD	20	80
N2	ū	ə/e	jambū	0	20
N3	ə/e	ī	bhəLə	0	30

Both children and adults appear to have no difficulty with the non-changing classes (M1-4, F1&2). The results are also encouraging for F3 and M5 classes. However, consonant final feminine classes (F4& F5) and the three neuter classes prove less tractable. Spreng's (2004) conclusions that some plural types are easier than others and those that are difficult remain so for both adults and children also holds true for Marathi. Children's performance is worse than the adults' performance for the difficult classes and, is particularly poor for the tokens simulating the neuter classes. These classes appear difficult for both groups. When in doubt, all subjects typically leave the nouns unmarked, substituting a zero suffix, though this is less frequent with adults than with children who opt for this strategy in providing answers. Other errors are restricted to the *potential* suffixes available for that particular gender class. For example, choosing ə/e where ī is required. The error is restricted to a specific gender class; in other words, gender drives affix selection but the target affix selected is incorrect though appropriate to the gender class. Though gender cues were provided, children as well as adults sometimes chose an affix that may apply to multiple genders though still relevant to the gender under consideration.

We can conclude that since the match between the phonological shape and the gender of the words is not unique, this makes the task of choosing the affixes for non-words complicated. The task reveals the high degree of ambiguity that is built into the system and is suggestive of the degree of difficulty in acquiring this system. Additionally, neuter nouns appear more problematic for both children and adults.

3 Intuiting Gender

Gender³ in Marathi is morphologically unmarked and is reflected primarily through agreement of the noun with phrasal elements such as adjectives, demonstratives and verbs. Grammatical gender is congruent with natural sex distinctions (male versus female) for animate nouns but is otherwise arbitrarily selected and not associated with readily identifiable phonological or semantic features.

³ Marathi along with Gujarati and Konkani are the only Indo-Aryan languages to have retained the three-gender system of Sanskrit where gender is arbitrarily assigned. Most languages have either shifted to the two-gender system or map grammatical gender with natural gender straightforwardly. Even loanwords in Marathi are reassigned arbitrarily to gender classes. In this Marathi resembles German more than its sister languages.

In this test, the subjects were asked to ‘guess’ the gender of non-words to measure the effect of the phonological shape of the noun on gender assignment. This test was inspired by A. Mill’s (1986) work on the acquisition of gender in German where she developed a test using non-words for German speaking adults in order to study the psychological status of morphological and phonetic rules in German.

The test group consisted of ten Marathi speaking adults (5 M and 5 F) from Mumbai. The pictures used for the *wug* test (Berko 1958) were modified for our purposes. These simple, coloured pictures of some unfamiliar objects were associated with non-words which adhered to the phonotactic rules of Marathi. We created multiple tokens to rule out possible form-specific problems. The results are given in Table 3.

Table 3: Gender assignment and its frequency

Singular Ending	Potential Gender Preference (in %)	
a	M	75
	F	25
ī	M	5
	F	90
	*N	5
ū	M	35
	F	0
	N	65
ə	M	20
	N	75
	*F	5
Consonant	M	50
	F	30
	N	20

The first column provides possible singular endings for nouns in Marathi. The second column describes the *potential* gender/s possible for each ending of the singular form. For example, consonant ending tokens can be masculine, feminine or neuter. The third column shows the preference of the subjects (in %) while assigning gender to these non-words.

The cells in grey show the preferred gender for the choices possible per class – the winning gender as it were. Most subjects assigned masculine gender to ‘-a’ ending tokens, feminine gender to ‘-ī’ ending tokens and neuter gender to ‘-ə’ ending tokens. No one assigned feminine gender to ‘-ū’ ending tokens. For the -ī and the -ə classes, one subject also chose a gender that should have been ruled out by its shape; this is indicated in the table above with an asterisk (*).

In the absence of any other information (semantic or real-world), phonological shape biases the choice of gender and subjects attempt to distinguish classes given that cue. So we see that (a) subjects preferentially associate particular phonological shape(s) with particular genders and (b) that their responses allow other for other/alternate possibilities that we expect.

4 Frequency and Class Size

From the results above, we conclude that the match between the phonological shape and the gender of the words is not unique in Marathi and that this makes the task of choosing the affixes complicated. Further, of the three gender classes, neuter nouns appear more problematic in all tests for both children and adults. It thus becomes important to consider the frequency of these nominal classes in Marathi and consider whether the difficulties associated with both plurals in general and neuter nouns in particular shows class frequency effects. Since such frequency lists are not readily available for Marathi, we chose a contemporary dictionary compiled for students, the Navneet Marathi-English Dictionary, from which to cull the tokens for our analysis. This

dictionary contains approximately 40,000 words. It represents a standard vocabulary of Marathi speakers. We collected all the nouns from this dictionary, approximately 18,000 tokens. The data needed some cleaning in order to address issues such as gender mismatch, compound words removal, manual inspection/verification of gender assignment by native Marathi speakers etc. At a final count, we had 17,840 nouns marked for gender. We further classify these nouns into 5 classes based on the phonological endings viz, a, ī, ə, ū and C (consonant). Table 4 shows the basic distribution of nouns across the three genders.

Table 4: Distribution of nouns across genders

Gender	Number of nouns	%
Masculine	7594	42.6
Feminine	6365	35.7
Neuter	3878	21.7
Total	17837	100.0

Nouns in Marathi are not distributed equally among the three genders. Masculine nouns occur most frequently (42.47%) in the data. Neuter nouns are least frequent and the difference between the frequency of neuter nouns and masculine or feminine nouns is significant.

In Table 5, we show the distribution of nouns across each nominal class and for all three genders. The first column describes the phonological ending of the nouns. Columns 2, 3 and 4 describe masculine, feminine and neuter nouns respectively. These columns are further divided to show the number of nouns allotted to each sub-class (phonological shape matched to one of the three genders) and their relative frequency (in %).

Table 5: Distribution of nouns within classes

Singular	Masculine nouns		Feminine nouns		Neuter Nouns	
	No.	%	No.	%	No.	%
C	3983	52.4	1887	29.6	3434	88.5
a	2029	26.7	1254	19.7	0	0
ī	623	8.2	3168	49.8	0	0
ū	226	3.0	56	0.9	63	1.6
ə/e	0	0	0	0	381	9.9
ə	733	9.7	0	0	0	0
Total	7594	100	6365	100	3878	100

There are six classes of masculine nouns. The consonant final one is the most frequent, followed by the 'a' class. Classes ending in 'ī', 'ū' and 'ə' are marginal classes of the masculine gender and it is observed that most of the nouns of these classes are borrowed from Sanskrit. The most frequent feminine class is the 'ī' class, followed by consonant and 'a' final nouns. The feminine class ending in 'ū' is marginal, constituting less than 1% of the total number of nouns. Of the three neuter gender classes, the consonant final class is the most frequent (88% of all neuter nouns). The other two classes are much less frequent.

Table 6 (read horizontally) is derived from the data shown in Tables 4 and 5. It represents the categorization of nouns into different classes based on their phonological endings. The first column shows the phonological ending of nouns. Columns 2, 3 and 4 describe masculine, feminine and neuter nouns respectively. These columns are further divided to show the numbers of nouns allotted to each class and their relative frequency. For example, in row 2, for the 'a' ending noun class 61.8% of the words are masculine, 38.2% of the nouns are feminine and none belong to the neuter gender. Further, as seen in the last column, this class constitutes 18.4 % of the total number of nouns.

Table 6: Distribution of nouns based on phonological ending

Singular	Masculine		Feminine		Neuter		Total Nouns	% Total Nouns
	No.	%	No.	%	No.	%		
C	3983	42.8	1887	20.3	3434	36.9	9304	52.0
a	2029	61.8	1254	38.2			3283	18.4
ī	623	16.4	3168	83.6			3791	21.4
ū	226	65.5	56	16.2	63	18.3	345	2.0
ə	733	65.8			381	34.2	1114	6.2
	Total						17837	100

We see that the consonant final nouns are the most frequent (52% of total nouns) and the nouns ending 'a' and 'ī' are much less so. Consonant final nouns are distributed among all three genders, but this distribution is not equal, masculine nouns being more frequent (42.8%) than either the feminine (20.28%) or the neuter (36.92%). Nouns ending in 'a' are distributed among masculine and feminine classes in a ratio of almost 2:1. The low frequency 'ə' and 'ū' classes contribute a fair number of neuter nouns. Thus neuter nouns are both less frequent in absolute numbers and also not uniquely identifiable. The insights we gain from a study of the lexical distribution of the classes now explains some of the difficulties faced by subjects in the previous tasks. We discuss this in the next section.

4.1 Comparison of tests

A comparison of the results of the previous tests and the class distribution that we just saw yields several generalizations. In Table 7, we present a comparison of class distribution with the intuition test, and in Table 8, a comparison of class distribution with the test for plural marking.

Table 7: Comparison of results of intuition test with distribution of nouns

Singular	Potential gender and its frequency - results of intuition test (%)		frequency of gender given a noun class (%)	frequency of noun class in all nouns (%)
a	M	75	61.8	18.4
	F	25	38.2	
ī	M	5	16.4	21.4
	F	90	83.6	
ū	M	35	65.5	2.0
	F	0	16.2	
	N	65	18.3	
ə	M	20	65.8	6.2
	N	75	34.2	
C	M	50	42.5	52.0
	F	30	20.3	
	N	20	36.9	

In Table 7, column 1 gives the potential phonological endings of the nouns in Marathi, column 2, the results of the intuition test (the potential gender given a particular phonological ending, column 3, the frequency of nouns allotted to a particular gender based on the phonological ending and column 4, the frequency of a noun class based on phonological ending given the total number of nouns.

It is noteworthy that the results for the '-a' ending tokens in the intuition test correlates with the distribution of the nouns ending in 'a' (cells marked in grey). Masculine gender is preferred by the subjects (potential genders being M&F) and 'a' ending masculine nouns are indeed more frequent (61.80%) than 'a' ending feminine nouns (38.2%).

A similar correlation is observed for the nouns ending in 'ī' (cells marked in grey). Feminine nouns ending in 'ī' are more frequent than masculine nouns and subjects preferred the feminine to masculine gender when intuiting gender for such tokens. Thus, nouns ending in 'ī' are intuited to be feminine and 'ī' ending nouns are predominantly feminine. For consonant ending nouns, a similar pattern is observed to some extent. All three classes occur equally frequently in the overall distribution of nouns.

The 'ə' and 'ū' ending nouns are judged to be neuter though their distribution indicates that they are more frequently masculine rather than neuter. It is observed that these two classes are marginal in the overall distribution of nouns and therefore preclude or limit more successful correlation between phonology and gender.⁴

In Table 8 which is very similar to Table 2 but contains an additional column which describes the frequency of nouns for a particular class, given the total number of nouns in Marathi we compare the results of the plural marking test with class distribution.

Table 8: Comparison of results of plural marking test with distribution of nouns

Gender	Sg	Pl	Non- words			% frequency of noun class
			Token	Results (%)		
				Children	Adults	
			sg to pl	sg to pl		
M1	ə	no	pūmbə	100	100	4.1
M2	C	change	tʃag	100	100	22.3
M3	ī		pəNī	100	100	3.5
M4	ū		maTū	100	100	1.3
M5	a	e	tuDa	80	80	11.4
F1	a	no	gaTa	100	100	7.0
F2	ū	change	bhakū	100	100	0.3
F3	ī	a	sakvī	60	100	17.8
F4	C	a	raT	10	30	10.6
F5	C	ī	Dhak	0	20	
N1	C	ə/e	khaD	20	80	19.2
N2	ū	ə/e	jambū	0	20	3.5
N3	ə/e	ī	bhəLə	0	30	2.1

It may be observed that for non-changing classes the results of the test is 100%. The results do not reflect the pattern of distribution of nouns in the lexicon. For changing classes, the results of '-a' ending masculine and '-ī' ending feminine tokens correlate with the distribution of nouns. The higher frequency of nouns leads to better performance for both adults and children. It is interesting to observe that though the incidence of consonant final neuter nouns in the overall distribution of nouns is considerable (19.24 %), children are not able to select the correct plural affix (20%). However, the performance of adults for the same class of nouns is encouraging (80%). This suggests that there is some effect of maturation in the acquisition of the rules of plural morphology. Children may defer inducing or applying the rules till such time as they have ample evidence for the same.

The frequency of the other two subclasses of neuter gender (N2 and N3) is significantly low ('ū' ending nouns 3.53% and 'ə' ending nouns 2.13%) and both children and adult are unable to mark the plural suffix correctly. The less frequent classes of masculine and feminine gender do not inflect overtly for number. Only the less frequent classes corresponding to neuter gender inflect overtly for number. Hence, low frequency probably accounts for the poor performance of

⁴ As remarked by one reviewer, we cannot rule out that something other than class marginality may be going on. We are increasing the number of tokens in all these tests and expect to reevaluate our results per class with this larger sample.

the subjects for these classes. Class F4 and F5 (consonant ending feminine nouns) also show poor performance in the selection of plural affixes. These noun classes are distributionally frequent but the ambiguity of affix selection accounts for the poor performance. The comparison shows not a straightforward correlation between noun distribution/frequency of nouns but a nuanced one where frequency, uniqueness of affix assignment and zero-suffixation all play a role in the performance of children and adults.

5 Conclusion

We have shown through a series of studies that the plural morphology of neuter gender nouns is difficult for subjects, both children and adults. When we relate the performance data to the distribution of nouns we gain several insights into this problem. While speakers do have an internalized system of rules and apply it appropriately, the performance and rule application is limited by other factors. It can be seen that nouns marked for neuter gender are the least frequent in the overall distribution of nouns. Hence, it is possible that the low overall frequency of neuter nouns triggers the poor performance in selecting plural suffixes. It can be concluded that while assigning gender, the gender classes which are distributionally most frequent are preferred over others. This effect is restricted to the classes which contribute a larger proportion of nouns in the overall distribution of nouns in Marathi. When the overall frequency of a class is low, that class also demonstrates problems with plural morphology assignment if, in addition, it also changes in the plural and its shape is class ambiguous. We can also infer that the rules for noun classes with high frequency are easy to acquire than those classes which have fewer nouns in it. Finally, since Marathi is one of the few Indo-Aryan languages with the three gender system, there may be additional pressure on the language to move towards a two-gender or an attenuated gender system as may be seen in other languages of this family.

References

- Berko, J. 1958. The Child's learning of English Morphology. *Word*, 17 (2-3), pp.150-177.
- Corbett, G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Damle, M. K. 1970. *Shastriya Marathi Vyakaran (A scientific grammar of Marathi)*. K. S. Arjunwadkar (ed). Pune, India: Deshmukh & Company.
- Dhongde, R. V. 1983. *Arwatsin Marathi*. Pune, India: Continental Prakashan.
- Kelkar, A. R. 1958. *The Phonology and Morphology of Marathi*. Ph.D Dissertation, Cornell University.
- Lust, B. 2006. *Child Language Acquisition and Growth*. Cambridge: Cambridge University Press.
- Mills, A. 1986. *The Acquisition of Gender*. Berlin and New York: Springer-Verlag.
- Pandharipande, R. 1997. *Marathi*. London and New York: Routledge
- Pinker, S. 1994. On the Acquisition of Grammatical Morphemes. *Journal of Child Language*, 8, pp.477-484.
- Prabhudesai, S. 2011. *Navneet Marathi – English Dictionary*. Naveet Publication: Mumbai
- Spreng, B. (2004). Error patterns in the acquisition of German plural morphology: Evidence for the relevance of grammatical gender as a cue. *Toronto Working Papers in Linguistics* 23(2), pp.147-172.
- Child Language Data Exchange System (CHILDES). 2003. wugs.pdf retrieved from <http://childes.psy.cmu.edu/topics/>