

FORMATION OF GENERALIZATION WORDS ("MOHMIL") IN URDU

HUDA SARFRAZ

ABSTRACT

"Mohmil" words are meaningless words in the Urdu Language used mostly for generalization purposes. The formation of these words was investigated by studying documented generalization words and by conducting a survey, in which the subjects were to form generalizations for a list of words.

1. INTRODUCTION

The rules of Urdu Grammar divide Urdu words into two categories, meaningless ("mohmil") and meaningful ("mauzoh") (Rafiq).

The word "mohmil" means meaningless or useless (Feroz al Lughat) and "mohmil" words are defined as those "meaningless words that are spoken along with meaningful words without any reason" (Siddiqi).

The meaningful word usually precedes the meaningless word, but there are some cases in which the opposite is true, e.g. [əla bəla] (evil or bad thing) (Feroz al Lughat), where [bəla] can independently mean something bad but [əla] alone is meaningless.

These words are mostly used for generalization purposes, e.g. [əla bəla], and sometimes to denote the excessive presence of a property, e.g. [gol mol] (round) or enthusiasm, e.g. [tʰik tʰak] (well or good) (Feroz al Lughat).

Since "mohmil" words are most commonly used for generalization purposes they will be referred to as generalization words.

Some generalization words are those that are fixed and can be found in dictionaries along with an actual meaningful word e.g. [gol mol] (Feroz al Lughat), where [gol] is an actual word meaning round, but [mol] is not and only makes sense if used with [gol].

The two words are used together to show excessive roundness (Feroz al Lughat).

Some generalization words are apparently formed randomly during speech e.g. the word [kʰana] may be generalized as [kʰana ʃana] or as [kʰana vana]. These are not found in dictionaries but may be found as examples of "mohmil" words in books. This paper investigates the formation of such words.

2. LITERATURE REVIEW AND PROBLEM STATEMENT

Hardly any formal analysis of generalization (mohmil) words exists. They are most extensively covered in school textbooks where they are defined and explained with examples. In Urdu Grammar books "mohmil" words are simply defined to be meaningless words and the explanation given for not discussing them any further is that a compilation of rules should only include meaningful words (Rafiq), i.e. there are no rules defining the formation of meaningless words.

The problem is to find out how generalization words are formed and if they follow a pattern of formation. For this study, generalization words have been categorized into two parts.

The first category includes those generalization words quoted in books as examples of "mohmil" words or in dictionaries and those that have been observed repeatedly in native Urdu speakers. These may include generalizations made by replacing the initial phoneme by a [v] or a [ʃ] or inserting a [v] or a [ʃ] in the beginning, if they are found quoted in books or dictionaries. Generalization words frequently heard in Urdu speech will also be included. This category will be referred to as documented generalization words.

The second category includes those that may be generalized using a [v] or a [ʃ] or possibly any other way, which is not documented. This category will be referred to as undocumented generalization words.

Since two types of generalization words are being considered, the problem statement has two parts.

First, what types of documented generalization words exist?

Second, what types of undocumented generalization words exist?

Are [ʃ] and [v] the only phonemes used for generalization?

Is it fixed for each word whether a [ʃ] or a [v], or any other phoneme will be used to generalize it?

If it is fixed, then is it predictable?

What is the influence of Punjabi and Urdu backgrounds on the formation of generalization words in Urdu?

What types of words are normally generalized? Can words with four or more syllables be generalized?

3. METHODOLOGY

To investigate the formation of documented generalization words, a list of generalization words from different books and dictionaries was compiled and categorized.

For the undocumented words, a survey was taken of word generalization using 43 Urdu words. The subjects were 34 native Urdu speakers.

Most of the words used in the survey were commonly used words, which are commonly generalized. Some of the words included were of more than three syllables and some were uncommon and not usually generalized, (See Appendix A for a complete list of the words used in the survey) to see what the reaction of the test subjects would be.

To check the influence of an Urdu background 5 specifically Urdu speaking subjects were chosen, and for Punjabi 5 were chosen who normally spoke Punjabi but could speak Urdu.

The subjects were asked to fill in the generalization they would normally use, and to skip any words that they wouldn't normally use or generalize. Some examples were also given along with the list of words. The test subjects were also told to give alternatives for the examples if they thought that they existed.

After the survey, a word was considered to be generalized by a [ʃ] or a [v] if 65% or more of the subjects considered it to be a

[ʃ] or a [v] word respectively. A percentage of 65 instead of 50 was used because it was noticed that while choosing between [ʃ] and [v] for generalization the subjects would sometimes switch between the two often before deciding on one; 65% was chosen to reduce any error in the results due to random decisions.

The percentage of [ʃ]s and [v]s used by each person was also calculated to check whether people had a preference for one or the other, in particular for Punjabi speaking and Urdu speaking people.

Also, a list of generalizations other than the expected [v]/[ʃ] generalizations obtained from the survey was compiled to see if the subjects used any other types of generalizations.

4. RESULTS

The following categories of generalization word formation were found (see Appendix B for complete list of words and generalizations compiled) from documented words and the survey:

1. Replacement/Insertion of initial phoneme, e.g. [ʃəlt səlt], [dʒadu vadu], [ulṭa pulṭa] etc. If the initial phoneme is a consonant it is replaced, e.g. [gum sʊm]. If it is a vowel a phoneme is inserted in the beginning, e.g. [əlek səlek]. This included the word [gol mətəl] (and other similar ones), in which the initial phoneme was replaced by a complete syllable. The replacement or insertion may be by the following phonemes: [v], [s], [k], [p], [d^h], [m] and [ʃ], [m] and [ʃ] being the most common. This was the most common generalization category.
2. Replacement of the middle phoneme. This category included only single syllable words and the middle phoneme that was replaced was always a vowel, e.g. [tʃʊp tʃap], [d^hum d^ham].
3. Replacement of most of the word (mostly only single phoneme retained), e.g. [g^hər bər].
4. Increase in number of syllables. In this category the generalization word had more syllables than the actual word and was significantly changed. In some cases the syllabic phoneme(s) was retained and in some the consonantal phoneme(s) was retained. There was not enough data to

predict what would be retained. Examples include [tʃori tʃəkari] and [dant dəpət]. This was the second most common type of generalization used.

5. No relation to preceding word, e.g. [d^həka pel].

The results of the survey are shown in Tables 1 and 2. Table 1 shows the percentage of [ʃ] and [v] generalizations for each word (see word list from Appendix A).

TABLE 1 Percentages of [ʃ]s and [v]s used for each word in survey (all other types of generalizations excluded)

word#	[v]%	[ʃ]%
1	45	55
2	74	26
3	77	23
4	70	30
5	47	53
6	68	32
7	54	46
8	59	41
9	55	45
10	67	53
11	59	41
12	79	21
13	46	54
14	45	55
15	76	24
16	60	40
17	68	32
18	73	27
19	41	59
20	61	39
21	32	68
22	73	27
23	50	50
24	40	60
25	67	33
26	71	29
27	66	34
28	86	14
29	12	88
30	62.5	37.5
31	48	52
32	71	29
33	94	6
34	73	27
35	23	77
36	52	48
37	40	60
38	54	46
39	52	48

40	5	95
41	20	80
42	93	7
43	35	65

From Table 1, the words with 65% or greater consensus (and therefore considered as fixed generalizations) from the survey as a [ʃ] or a [v] generalization are:

khana vana
 kentʃi ventʃi
 kursia vursia
 pərhai vərhai
 bətʃe vətʃe
 tʃor vor
 məsla vəsla
 tʃot vot
 tʃiʃiã vɪʃiã
 dʒadu vadu
 pani vani
 roti voti
 hat^hi vat^hi
 məzahira vəzahira
 deɦʃətɡərɪ vɛɦʃətɡərɪ
 səhafi vəhafi
 bariʃ variʃ
 əlmari ʃəlmari
 bəhavəlpur ʃəhavəlpur
 əfyanistan ʃəfyanistan
 məhavərat ʃəhavərat
 fəvara ʃəvara
 səvab ʃəvab

Table 2 shows the percentage of [ʃ], [v] and other types of generalizations for each individual subject.

TABLE 2 Percentages of [ʃ]s and [v]s used by each test subject in survey (all other types of generalizations excluded)

(P) – Punjabi Speaking
 (U) – Urdu Speaking

subject #	[v]%	[ʃ]%	others%
1	21	79	0
2	88	0	12
3	86	12	2
4	95	0	5
5	0	95	5
6	39	61	0

7	18	82	0
8	7	63	30
9	40	58	2
10	42	42	16
11(U)	97.5	0	2.5
12(U)	95	5	0
13(U)	67	26	8
14(U)	68	23	10
15(P)	46.5	46.8	4.7
16(P)	22	71	7
17(P)	26	74	0
18(P)	7	73	20
19	50	33	17
20	46	50	4
21	20	20	60
22	16	41	43
23	76	24	0
24	97.5	2.5	0
25	76	14	10
26	81	19	0
27	58	42	0
28	72	28	0
29	84	14	2
30	40	60	0
31	27.5	67.5	5
32	56	37	7
33(P)	49	51	0
34(U)	69	0	31

Other generalizations used by test subjects in the survey are listed below word by word, the number of times they occurred is also stated in the following format:

Word number generalization (number of occurrences) – test subject numbers

1. ləte (2) – 23, 30
mərɽe (1) - 31
2. k^hane (1) - 5
mana (1) - 21
pina (1) - 34
3. pəntʃi (2) – 8, 32
4. mʊrsiā (1) - 4
5. mʊɽe (1) - 21
6. lik^hai (2) – 14, 19
7. pərɽai (1) - 10
dʒəɽa (1) - 14
kətai (1) - 21
mərɽai (1) - 22

- b^hərɽai (1) – 30
8. dəpət (11) – 2, 13, 14, 18, 19, 20, 21, 25,
30, 31, 34
pit (1) – 22
9. kuta (2) – 21, 22
gilli (1) – 8
10. bʊtʃtʃe (1) - 22
sətʃtʃe (2) – 8, 31
bale (1) - 30
kətʃtʃe (2) – 10, 11
11. none
12. tʃar (1) - 30
mor (3) – 10, 21, 22
lor (1) – 8
13. muha (3) – 2, 13, 21
14. tʃətəɽe (1) - 16
15. mʊsle (1) - 5
mʊsla (1) – 18
16. none
17. mot (6) – 2, 10, 13, 15, 21, 22
18. kəbʊɽər (1) - 22
pɪɽiā (1) – 8
19. t iur (1) - 21
20. mərəɽt (1) - 34
21. vəɽərə (1) - 22
ɽari (1) – 31
təlmari (1) – 34
22. none
23. kətʃalu (5) – 2, 8, 22, 25, 29
malu (2) – 21, 34
kalu (1) – 10
ulu (1) - 14
24. vur (1) - 15
ləɽgur (1) – 19
məɽgur (1) – 21
25. tuna (3) – 21, 22, 30
26. d^hani (1) - 15
puni (1) – 22
27. boti (1) - 21
tʊkər (1) – 22
28. sat^hi (2) – 8, 19
hut^hi (1) – 18
mat^hi (1) – 21
29. pəhavəlpur (1) - 3
kəhavəlpur (1) – 8

- smd (1) – 27
 ravəlpur (1) – 29
 30. maludgi (2) – 8, 22
 tʃʰaludgi (1) – 4
 faludgi (1) – 10
 ʃudgi (1) – 31
 31. none
 32. ʃahira (1) -15
 məzuhəra (1) – 18
 zahira (1) – 21
 vaira (1) – 22
 33. vərđi (2) – 11, 27
 ʃərđi (1) – 15
 34. məhafi (2) – 8, 10
 səhufi (1) – 18
 35. tʃʰafnistan (1) - 4
 taleban (1) – 8
 vanəstan (1) – 15
 ʃəntan (1) – 17
 əfyunistan (1) – 18
 pʰəfyānistān (1) – 34
 36. tʰam (1) - 34
 kadʒ (1) – 36
 37. rəmbil (1) – 34
 38. pʰəndʒir (1) – 34
 məndʒir (1) – 22
 39. tʃəmbəl (1) – 32
 kumbəl (1) – 22
 40. pʰəhavəraʃ (1) – 34
 41. pʰəvara (1) – 34
 fəvura (1) – 18
 42. burəʃ (1) – 18
 tarɪʃ (1) – 34
 43. təfyunistan (1) – 18
 məvab (1) – 32
 dʒəvab (1) – 10
 əzab (1) – 8

5. DISCUSSION

From the survey it was observed that other than [v] and [ʃ], the phoneme commonly used (but not that commonly observed) for generalization by inserting or replacing in the beginning of a word is [m], the most

notable example being [tʃuha muha]. Another common way of generalization was to change a vowel, e.g. [hatʰi hutʰi], [pani puni], [tʃor tʃar]. These words can fit into the categories defined for documented words, so basically the ways of formation of documented and undocumented generalization words are the same.

From the fixed (23 out of the 43 words) [v]/[ʃ] generalizations obtained from the survey, the [ʃ] replacement/insertion only occurred when there already was a [v] in the word except in [əlmari ʃəlmari]. There wasn't enough data to conclude whether this was a special case or there were other words that used [ʃ] as well. Also wherever such examples were quoted in books, a [v] was always used. There were no examples with [ʃ]. Therefore though it can be seen that the use of [v] is more common overall, it is inconclusive whether [v] and [ʃ] placement is predictable.

One possibility is that generally when a word starts with a vowel it is generalized by a [ʃ], but most of the words in the survey started with a consonant, so this is inconclusive. The words starting with vowels were [əlmari] (68% ʃ, [alu] (50% ʃ), [əŋgur] (60% ʃ), [aludgi] (37.5% ʃ), and [əfyānistān] ʃ.

If the choice of a [v] or a [ʃ] is predictable, it will probably be dependent on more than just the initial syllable of the word, as can be seen from the two words [aludgi] and [əlmari].

17 out the 43 words had a division between 40 and 60% for the [v]/[ʃ] replacement/insertion, the word [alu] having an exact 50% for each. This indicates that the selection of a [v] or a [ʃ] may be random, at least for some words.

Out of the 34 subjects 8 used [v] for insertion/replacement on more than 80% of the words, 2 of these subjects were Urdu speaking. 2 subjects used [ʃ] on more than 80% of the words, neither of these subjects were of any particular background.

4 of the 34 subjects chose only one of [ʃ] or [v] for all their generalizations, but they also used other types of generalizations for some words.

Therefore it was concluded that there wasn't a general tendency for individuals to use either a [v] or a [ʃ].

21 subjects used other types of generalizations for some of the words.

Overall the five Urdu speaking test subjects tended to use [v] more, and also used [p^h], [t] and [t^h]. The Punjabi speakers tended to use [ʃ] more, and also changed vowels more often, as in [hat^hi hut^hi].

The survey showed that the most commonly generalized words are everyday words with one or two syllables. The three syllable words in the survey were mostly not generalized by the test subjects, and in some cases the syllables of the words were reduced in order to generalize them, e.g. [aludgi ʃudgi].

There is the possibility that with a larger number of test subjects, the results quoted above about words generalized by [v] and [ʃ] may change. Initially the survey was conducted using only 20 test subjects, but the results varied too erratically for almost all the words, so the number of test subjects was increased to get a more dependable result. A second list of words may also be compiled based on these results to reach a better conclusion.

6. REFERENCES

Feroz al Lughat, Urdu Jadeed. Ferozsons Lahore, Rawalpindi, Karachi.

Rafiq, M. *Urdu Qawaid o Insha Pardazi, Hissa Awal.* Ferozsons Lahore, Rawalpindi, Karachi.

Rafiq, M. *Urdu Qawaid o Insha Pardazi, Hissa Dom.* Ferozsons Lahore, Rawalpindi, Karachi.

Siddiqi, A.K. *Aina Urdu Lazmi, Pilot Urdu Grammar and Composition, Urdu Lazmi Parcha (Alif-Bay) Mukamal, baray jamat naham, daham.* Urdu Bazaar Lahore.

Siddiqi, A.K; Siddiqi, A.T. *Aina Urdu Lazmi, Qawaid o Insha Pardazi, baray jamat hashtam, parcha bay.* Urdu Bazaar Lahore.

7. APPENDIX A

List of words used in survey for undocumented generalization words:

1. kəpɾe	clothes
2. k ^h ana	food
3. kentʃi	scissors
4. kursiā	chairs
5. dʒute	shoes
6. pəɾhai	studies
7. ləɾai	fight
8. dant	scold
9. billi	cat
10. bətʃtʃe	children
11. seb	apple
12. tʃor	thief
13. tʃuha	mouse
14. pətəɾe	fire crackers
15. məsla	problem
16. tʃɪpkəli	lizard
17. tʃot	injury
18. tʃɪɾiā	birds
19. trar	ready
20. dəɾəɾt	tree
21. əlmari	cupboard
22. tʃ ^h ətri	umbrella
23. alu	potato
24. əŋgur	grapes
25. dʒadu	magic
26. pani	water
27. roti	bread
28. hat ^h i	elephant
29. bəhavəlpur	bahawalpur
30. aludgi	pollution
31. gədagəri	beggary
32. məzahira	demonstration
33. dəhʃətgərđi	terrorism
34. səhafı	journalist
35. əfyanıstan	afghanistan
36. kam	work
37. zəmbil	small bag
38. zəndʒir	chain
39. kəmbəl	blanket
40. məhavərat	
41. fəvara	fountain
42. barıʃ	rain

43. səvab

8. APPENDIX B

Generalization words listed by type (words from the survey were included if they were used by more than 3 people or if they were also commonly observed):

1. Initial phoneme replacement/insertion:

ibarət vıbarət (Rafiq)
 kəlam vəlam (Rafiq)
 ɣəlt səlt (Rafiq)
 mel kətʃel (Rafiq)
 ulti pulta (Rafiq)
 pani vani (Rafiq)
 pəkət d^həkət (Rafiq)
 dʒadu vadu (Rafiq)
 nəher vəher (Rafiq)
 hat^hi vat^hi (Rafiq)
 bətʃa kətʃa (Siddiqi)
 əlek səlek (Siddiqi)
 gəm sum (Siddiqi, Siddiqi)
 gol mol (Feroz al Lughat)
 əla bəla (Feroz al Lughat)
 nəmaz ʃəmaz (observed)
 k^hana vana (observed)
 gənd mənd (observed)
 mıtı ʃıti (observed)
 pəse vəse (observed)
 rona d^hona (observed)
 roz bəroz (observed)
 dım bədı (observed)
 tʃur mur (observed)
 alu kətʃalu (survey)
 alu kətʃalu (survey)
 tʃot mot (survey)
 pani d^hani (survey)
 tʃuha muħa (survey)
 khana vana (survey)
 kentʃi ventʃi (survey)
 kursıā vursıā (survey)
 pətʃhai vətʃhai (survey)
 bətʃe vətʃe (survey)
 tʃot vot (survey)
 tʃırtıā vırtıā (survey)
 dʒadu vadu (survey)

pani vani (survey)
 roti voti (survey)
 hat^hi vat^hi (survey)
 məzahıra vəzahıra (survey)
 deħʃətɡərđi vəħʃətɡərđi (survey)
 səħafi vəħafi (survey)
 bariʃ variʃ (survey)
 əlmari ʃəlmari (survey)
 məħavərat ʃəħavərat (survey)
 fəvara ʃəvara (survey)
 səvab ʃəvab (survey)

2. Middle phoneme replacement:

t^hik t^hak (Rafiq)
 b^hıt b^hat (Rafiq)
 tʃup tʃap (Rafiq)
 d^hum d^ham (Rafiq)

3. Replacement of major part of word:

bat tʃit (Rafiq)
 səda səlf (Rafiq)
 sətʃ mutʃ (Rafiq)
 b^hut pəret (Siddiqi)
 ıka duka (Siddiqi)
 g^hər bar (observed)
 g^has p^hus (observed)
 kərtə ləte (survey)

4. Increase in syllables:

tʃori tʃəkari (Rafiq)
 d^hol d^həməka (Rafiq)
 gol mətol (Rafiq)
 ləmba tərəŋga (Rafiq)
 dana dənka (Siddiqi)
 kala kələta (Siddiqi)
 tal mətol (Siddiqi)
 dım bədı (observed)
 roz bəroz (observed)
 dant dəpət (survey)
 alu kətʃalu (survey)

5. No Relation:

d^həka pel (Rafiq, *Hissa Dom*)
 kona k^hudra (Rafiq, *Hissa Dom*)
 dʒadu tuna (survey)
 roti tukər (survey)