

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/309202308>

Recognizing Urdu Compound Words

Conference Paper · October 2016

CITATIONS

0

READS

4,915

2 authors:



Abdul Jabbar

Virtual University of Pakistan

66 PUBLICATIONS 1,776 CITATIONS

[SEE PROFILE](#)



Sajid Iqbal

Bahauddin Zakariya University

7 PUBLICATIONS 55 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



ResiliNets [View project](#)



Diagnosis and grading of Human Brain Tumor in MRI [View project](#)

Urdu Compound Words Manufacturing: A State of Art

Abdul Jabbar

*Department of Computer Science
Institute of Southern Punjab, Multan, Pakistan
a.jabbar73@hotmail.com*

Sajid Iqbal

*Department of Computer Science
Bahauddin Zakariya University, Multan, Pakistan
Sajid.iqbal@bzu.edu.pk*

Abstract

Compounding is a device to invent new words in Urdu language. In this paper, we discuss two types of compound words, first that is created by affix word such as *جیل خانہ جات* and *نمہ داری* and second that is created by Mohmil words such as *کپڑے وپڑے* and *جنتر منتر*. Such compound words are constantly invented in Urdu language; it is impossible to list them in Urdu lexicon and also difficult to identify the boundary of such compound words. In this article, we generate the first Urdu Affix Words List (UAWL) and Urdu Mohmil word patterns both are used to coin compound words. The criteria for the selection of affix word are the frequency of the affix word and head word must belong to Urdu lexicon. The corpus, we study consists of five Urdu grammar books, 149466 Urdu words and 100 news articles from BBC Urdu and Dawn newspaper. The UAWL and Urdu Mohmil words pattern may provide a good level of confidence to develop Urdu Natural Language Processing (UNLP) applications such as word segmentation, phrase chunking, parts of speech tagging and machine translation.

Keywords

Urdu, Affix words, Mohmil words, compound words, Urdu corpus

1. Introduction

Urdu is spoken by more than 100 million people around the World¹. In Pakistan, approximately 60% of Pakistanis learn Urdu as a second language Dar et al (2015). In spite of large number of Urdu speakers in the World, Urdu is still known as a resource-poor language (Syed et al. 2011).

¹ <http://www.bbc.co.uk/languages/other/urdu/guide/facts.shtml>

In Urdu language, compounding is a productive device that generates new words Rahman, (2012). A comprehensive study of Urdu compound can be found in Rahman, (2012). Compounding is a common phenomenon for morphologically rich language like Urdu and it is a challenge for the natural language processing (NLP) systems Kuncham et al. (2015) such as machine translation, parts of speech tagging etc. Since compound words are not often found in the dictionaries or lexicon. However, their recognition and splitting could be of benefit for various NLP tasks (machine translation, information retrieval, terminology extraction, etc.). In this paper, we intended to contribute small, publicly available Urdu affix words list and pattern of Mohmil words for the Urdu Natural Language Processing (UNLP) community.

1.1 Objective of study

Almost all the natural language processing applications based on the word and words are derived by tokenization process. Urdu text tokenization based on whitespace produces unwanted results, such as *داری ذمہ* wrong split into two words, which is, in reality, one compound word (Ijaz et al. 2007). The aim of this study is to collect the affixes that are used to coin compound words and secondly, we develop the pattern of Mohmil words that leads to morphological modification of headword. In this change, Mohmil word is used as affix. Mohmil words belong to open class words and a user of language can coin new Mohmil words easily. Further, a great variety is observed in such word It is, therefore, impossible to list all Mohmil words in dictionary or lexicon.

Words structure in Urdu

Urdu has complex words structure and derivation is a sole process to coined new words in Urdu language, for example from an Urdu *خوش* (happy), by adding the affix, a new word *خوش مزاج* (pleasant) is coined (Syed et al 2014). Islam, (2012) describe two types of derivation, first derivation by affixation such as *داری ذمہ* in which *داری* is a non-word suffix (Hardie, A. 2004) and second by derivation compounding in which two independent word are concatenated to coined a compound words *خوش اخلاق*, in such compound words mostly one constituent come from the Persian or Arabic language (Islam 2012), compound can be hybrid *کریانہ سٹور* (Qureshi et al 2012)

2. Review of literature

Urdu Zabta Takhti (UZT) considered a standard code for Urdu language approved by Government of Pakistan (Hussain and Afzal, 2001). A first Urdu corpus of 1.6 million words was established by EMILLE (Enabling Minority Languages Engineering) Project at the Universities of Lancaster and Sheffield (McEnery et al. 2000). Ijaz et al. (2007) collect 18 million Urdu words with domain specific. Hussain, (2008) corpus contains 100,000 manually tagged Urdu words in which about 40,000 words annotated with Named Entity tags. Another corpus is produced by Muaz et al (2009) consisting of 230K tagged Urdu words.

3. Method

3.1.Data collection

Three main sources, we use to collect data, first is online 149466 Urdu word list², some affix words extract from the five books (Schmidt, 1999; Board, 2010; Bloch, 2012; Haq, 1996; UEP, 2014) and we also study the online news article 100 news articles from three disciplines including science and technology, sports and politics taken from BBC Urdu³ and Dawn news⁴, and two online Urdu dictionaries available at^{5, 6}.

3.2.Data processing

Since the standardization of derived data, digits, special characters, English words and stop word is eliminated for lexical analysis. From the Urdu word list, if a line contains the single word then remove it. Two principal frequency of affix words and after extracting the affix, headword must belong to Urdu lexicon are guided to select the affix words.

The computer software would read all derivations of a word

The data processing mainly comprised the tokenization and standardization. The Urdu text got from the online sources convert into the UTF-16 standard character encoding scheme. From the continuous Urdu text, content words are derived by assuming Punctuation marks, special symbols (\$, %, #, @, +), English alphabets (A-Z and a-z), numbers (0-9 and ۰-۹), Urdu stop words, Carriage return, tab, Line feed and Form feed as a boundary of content words. Stop words

² http://www.cle.org.pk/software/ling_resources/wordlist.htm

³ <http://www.bbc.com/urdu>

⁴ <http://www.dawnnews.tv/>

⁵ <http://www.urduencyclopedia.org/urdudictionary/index.php>

⁶ <http://182.180.102.251:8081/oud/default.aspx>

are common words and have no linguistic meaning (Atwan et al 2013) in Urdu language, postpositions, pronouns, determiners, and conjunctions have considered stop words (Gupta et al. 2015). White space is not used for tokenization, since any compound words such as خوف ناک will be wrongly tokenized into two words.

3.3.UAWL development

We develop a criterion on which affix words are collected first affix word must come with more than two words and second after removing the affixes word, headword must belong to Urdu lexicon. For example پانے جات، پانی، دھکم دھکا are compound words, in which پانے جات and پانی are affixes and after removing the affixes headwords are جات and پانی that present in Urdu lexicon. The complete list of such affix is presented in appendix A

3.4.Pattern of affix word development

Another type of compound words is those in which Mohmil word is used as affix word Such as پانی، دھکم دھکا In which پانی and دھکم are Mohmil words. Haq (1996) and Sarfraz () are also describe some Mohmil words and their pattern. The complete list of Urdu Mohmil word patterns is given in appendix B.

4. Conclusion

The resources we develop consist of 768 affix words and 20 patterns of Urdu Mohmil words. The UAWL constructs in this paper can serve as base to identify the Urdu compound words either bi-gram or tri-gram words. The pattern of Mohmil words will also provide a strong ground to identify the Mohmil words in Urdu text and also identify the compound words created by Mohmil words. We are hopeful that these resources will provide a good level of confidence for many Urdu Natural Language Processing (UNLP) applications. Since this research work is a preliminary study on Urdu affix words that is used to coin compound words therefore it contains some limitation: first corpus size is limited, and second, no Urdu standard Urdu affix list and Mohmil patterns are available to validate our constructed UAWL and Mohmil pattern. In future, we intend to grow our data size and enhance the list of affix words and explore more Urdu Mohmil words to extract the patterns. Our aim is to develop an automated system to segment the compound Urdu words.

Acknowledgements

I would like to express my heartfelt thanks to Masood Jhandir Research Library, Sardar pur Jhandir Mailsi, District Vehari, Punjab, Pakistan (jhandir.com) for that they have made some important resources available.

References

- Atwan, J., Mohd, M., & Kanaan, G. (2013). Enhanced arabic information retrieval: Light stemming and stop words. In *Soft Computing Applications and Intelligent Systems* (pp. 219-228). Springer Berlin Heidelberg.
- Bloch, Dr. Sohail Ahmad (2012), “بنیادی اردو قواعد”, Muqtadrah Qumi Zuban Pakistan, Islamabad.
- Board, P. T. (2010). “اردو قواعد وانشاء” for Class-10th. Lahore: Punjab Textbook Board.
- DAR, M., ANWAAR, H., VIHMAN, M., & KEREN-PORTNOY, T. A. M. A. R. (2015). DEVELOPING AN URDU CDI FOR EARLY LANGUAGE ACQUISITION. Editors, 1.
- Gupta, V., Joshi, N., & Mathur, I. (2015, February). Design & development of rule based inflectional and derivational Urdu stemmer ‘Usal’. In *Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, 2015 International Conference on (pp. 7-12). IEEE.
- Haq, Molvi Abdul (1996), “قواعد اردو”, Anjuman Tariqi e Urdu, New Dehli (Hind)
- Hardie, A. (2004). The computational analysis of morphosyntactic categories in Urdu (Doctoral dissertation, Lancaster).
- Hussain, S. (2008, January). Resources for Urdu Language Processing. In *IJCNLP* (pp. 99-100).
- Hussain, S., & Afzal, M. (2001). Urdu computing standards: Urdu zabta takhti (uzt) 1.01. In *Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century. Proceedings. IEEE International* (pp. 223-228). IEEE.
- Ijaz, M., & Hussain, S. (2007, August). Corpus based Urdu lexicon development. In the *Proceedings of Conference on Language Technology (CLT07)*, University of Peshawar, Pakistan (Vol. 73).
- Islam, R. A. (2012). The morphology of loanwords in Urdu: the Persian, Arabic and English strands.
- Kuncham, P., Nelakuditi, K., Nallani, S., and Mamidi, R. (2015). Statistical sandhi splitter for agglutinative languages. In *Computational Linguistics and Intelligent Text Processing*, pages 164–172. Springer.
- McEnery, A. M., Baker, J. P., Gaizauskas, R., & Cunningham, H. (2000). EMILLE: towards a corpus of South Asian languages. *Br. Comput. Soc. Machine Trans. Specialist Group*, 11, 1-9.

- Muaz, A., Ali, A., & Hussain, S. (2009, August). Analysis and development of Urdu POS tagged corpus. In Proceedings of the 7th Workshop on Asian Language Resources (pp. 24-29). Association for Computational Linguistics.
- Qureshi, Anwar & Awan, (2012, September) "Morphology of the Urdu Language", International Journal of Research in Linguistics and Lexicography, INTJR-Volume 1-Issue 3.
- Rahman, M. (2012). A Linguistic Approach to Comprehensive Study of Compound Words-With Special Reference to Urdu Language. Language in India, 12(2).
- Sarfraz, huda, "FORMATION OF GENERALIZATION WORDS ("MOHMIL") IN URDU", Center for Research in Urdu Language Processing
- Schmidt, Ruth Lail (1999). URDU: AN ESSENTIAL GRAMMER.
- Syed, A. Z., Aslam, M., & Martinez-Enriquez, A. M. (2011, November). Sentiment analysis of urdu language: handling phrase-level negation. In Mexican International Conference on Artificial Intelligence (pp. 382-393). Springer Berlin Heidelberg.
- Syed, A. Z., Aslam, M., & Martinez-Enriquez, A. M. (2014). Associating targets with SentiUnits: a step forward in sentiment analysis of Urdu text. Artificial Intelligence Review, 41(4), 535-561.
- UEP (2014), "تخلیق اردو گرائمر", for class 8th, Unique Education Publisher, Urdu bazar Lahore.

Appendix A (Affix word list) $128 * 6 = 768$

Affix words List					
انگیز	والا	سر	آمیز	ہزار	دانی
پس	ہارا	سہ	آباد	ہشت	شدہ
پر	ترین	شاہ	دان	ہفت	گزار
پنچ	تر	شہ	خانہ	ہم	غیر
پیش	جگر	شش	دہندہ	ہمہ	خلاف
تہ	ناک	صاحب	نو	یک	فی
خر	سٹور	صدر	سٹی	ٹیکس	ثانوی
خوش	اصطلاحات	غیر	سالانہ	مند	جرائم

Affix words List					
گان	دستاویزی	رنگ	زنجیری	سر	ثلاثی
زیر	مولویانہ	نیم	رومانی	مہا	ٹیڑھی
زیر	لیفٹیننٹ	نو	روحانی	شاہانہ	تقصیر
زود	متصوفانہ	ہر	روایتی	زہریلے	ٹھنڈا
آخراں	کوہستانی	تادم	روحانی	ژولیدہ	تیزیہ
اخڈال	جمالیاتی	پنجہ	روایتی	سائنسی	تیزاب
نورال	معصومانہ	پلنگ	ڈسچارج	معروضی	تکمیل
ہومیو	قرارداد	پستہ	خوشامد	مجازی	تکلیف
وقرال	اظفارال	پتھر	تیزابی	متوسط	تیرگی
نوائے	کیمیائی	پائے	تکمیلی	گرمیہ	تقویم
نظامت	کیمیائی	پاؤں	تکبیری	گردشی	تقطیر
متبرک	طلسماتی	تخمی	پائیہ	صحیفہ	جلالت
ماتمی	خاندانی	آبلہ	تقابلی	زنانہ	تعویذ
لیڈیز	ترازوئے	سوڈو	تعدیلی	جوابی	تعدیل
کھوٹا	ابتدائے	روئے	تصدیقی	ٹھنڈی	تصفیہ
کھوٹی	ابتدائی	ایکس	تحویلی	تکبیر	تصفیہ
مستقل	سائنٹفک	انڈر	پیمانہ	تشنہ	پارٹی
مرکزی	شیرینیہ	آہنی	ابنائے	تشبیہ	پارچہ
مذہبی	طلسمانی	باسی	اتفاقی	تسخیر	پائین
مخملی	گورنمنٹ	بازی	بازاری	تحریف	ترکیب
مخلوط	مجرمانہ	بادہ	اٹینہ	تحریر	تذکرہ
محکمہ	محافظتی	بندہ	پاکیزہ	تائید	تحلیل
محکمہ	محکمانہ	برنج	پبلشنگ	تاریخ	تحسین
مجلسی	مدافعتی	برتن	پیمانہ	پچھلے	تحریم
متفقہ	مشرکانہ	آثار	تاریخی	ابالی	تحریر
مادری	مدافعتی	اپلا	تاریک	ابجدی	پچھلی
ماتمی	نفسیاتی	ابجد	اولوال	ابدان	پھاڑی
لشکری	برادران	ابدی	تبدیلی	ابروے	شکستہ

Affix words List					
تجنیس	ابناء	نافذال	آبرو	صوبے دار	مابعد
مدبر	ابنای	واحدال	ابلا	احقرال	ثرائی
عددی	ابوال	گلوگیر	ابلق	مانکرو	کشادہ
طبعی	ابدال	نادرال	ابلق	ماہانہ	قانون
صحیح	آتشیں	ناقابل	اپیل	مردانہ	طریقہ
شکست	اتمام	ناجائز	مدار	مجموعہ	طالع
سبزہ	آلودہ	تدریجی	نظام	بہترین	ما فوق
زہرہ	انگشت	ترجیحی	ہوال	متوازی	صحیفہ
زنان	بازار	مزاحیہ	ناقص	متنازع	شیریں
دماغ	باران	مخملیں	نازک	کورانہ	شومیء
حکمت	بادہء	مجموعی	مرکب	طوفانی	شوخی
حرکی	آئینہ	تقاضاے	ماہر	طبقاتی	شوخیء
حاصل	آوارہ	تقطیری	ماتم	سلطانی	شہادت
چوبی	بادیہ	مجموعہ	لیزر	سرکاری	سیاسی
جذبہ	باراں	تکوینی	لیڈی	بان	سلطان
ترکی	باریک	تولیدی	کورا	لوش	زنانی
ترکش	باطنی	گیسوئے	مسلک	گوشہ	زاویہ
تحیت	بدرجہ	ٹکسالی	مردم	گورا	دماغی
تازی	بندہء	کاتبین	مذاق	گندہ	دعائے
نیک	برہان	قانونی	محفل	گلشن	دارال
برگ	برائے	عمرانی	مجلس	گرمی	خلافت
بدھ	ولدال	زاویہء	مثبت	گرما	حیثیت
باٹ	میثاق	طریقہء	مادر	گردش	حاضری
باب	برادریوں	زہریلی	لباس	کورٹ	جوہری
اگن	مخالفانہ	صوبائی	گوشہ	کثیر	بدرجہ
اگر	العلمانہ	شیرینی	حاضر	کامل	پتلی
اسم	سپرداری	بیر	بقاے	کاتب	پتخہ
ابن	پاشیدگی	بحث	بگڑا	قومی	پبلک

Affix words List					
غیر	آہنگیوں	ظفر	بلند	آتشی	ابیض
ہمہ	وندکریم	طبع	چہار	فوجی	موجب
ہفت	نخواستہ	ضعف	چائے	عضوی	پاکٹ
ہشت	ناگہانی	بحر	جوہر	پرست	پارہ
زور	رسانیوں	صلح	جذبہ	عالی	پارہ
خوش	رسانیاں	صدر	ترنج	طلاق	پارس
نیم	آر ایانہ	متحدال	جائے	الہی	جلدی
راہ	گزارانہ	داریوں	ثقلی	ضعیف	جلال
اثر	شامیانہ	خارجیہ	تیغہ	انشا	جلاء
آبا	تازیانہ	دارانہ	تیغہ	بابا	تیرہ
ابو	سوفیانہ	طفلانہ	تیزی	باقی	ادھ
ابی	مجرمانہ	پروانے	تکیہ	شوخی	سرد
ابی	جداگانہ	ورزیاں	تکیہ	بانگ	خار
نور	پروریاں	واٹرز	تکلف	بایو	ہیڈ
خوب	انگریزی	گردانی	تشنہ	بخشی	ماہ
اہل	نوازیوں	برادری	تشبہ	شعلہ	کڑی
لوح	نوازیوں	دلنشین	ترنم	شعلہ	کچے
لخت	بیانیاں	گرافیز	صاحب	سیرت	کچی
گلا	فشانیاں	ٹیگیاں	بلیک	سجدہ	عطر
پری	اندازوں	لیواؤں	بزرگ	سجدہ	عدت
گرم	تراشیاں	گاہنیں	تراش	بدری	ضرب
گرد	پروریوں	نویسوں	تخمی	سیزہ	زیر
کوے	خوانیوں	سکنوں	تختہ	بدعت	زہر
کوہ	پسندانہ	بوسیوں	تختہ	سادہ	روح
کور	حاکمانہ	سازیاں	برزخ	برہم	دشت
کند	مالکانہ	ریزیاں	تبسم	دامن	برقی
کشش	پرستانہ	کاریوں	تباہ	زریں	چوب
آہن	شفابخانہ	رانیوں	تائے	رویت	جوش

Affix words List					
کرہ	ظریفانہ	آفرینی	تانب	روشن	جام
کرہ	گسترانہ	بہاراں	باڈی	روبہ	ٹکٹ
ایک	غائبانہ	ءمراں	تازہ	رونق	تخت
باج	شاعرانہ	اختران	تارک	رائے	تال
باد	پنجگانہ	نگاراں	بادہ	ذہنی	پیر
بار	غاصبانہ	کارانہ	تابہ	بستر	پنچ
عون	خطیبانہ	بازانہ	تابع	بسیط	پسر
باگ	خودداری	جویانہ	پیلی	بعید	پاے
بام	شعارانہ	روزانہ	پہلا	حالت	پان
عضو	کشائیوں	مندانہ	پھٹا	روروں	جلا
عرش	کشائیاں	مستانہ	گراں	خواری	بخت
عدم	نویسیوں	غرضانہ	پنجہ	شعاری	ابر
نامے	وندزادہ	افسانہ	پنبہ	معنوی	آہن
ارتے	آرائیوں	جرمانہ	پنبہ	طرازی	آبی
وائز	آرائیاں	بیگانہ	بالا	بازوں	زود
بندی	رسدخانہ	گواری	پارہ	آفرین	بال
خوری	افزائی	دانوں	پختہ	مندوں	تتگ
پوشی	خداداد	بیانی	طالع	گاہیں	اشک
نامی	بردار	آرائی	گردی	گاہوں	اروں
دانی	خانوں	کنندہ	پراں	خوانی	ئگان
جاتی	لاہیاں	گزران	ازمز	گاروں	ئگان
جات	ناترسی	دوراں	کیسز	گیروں	فہمی
یانا	یفکیشن	آزادی	نواز	داروں	ہٹوں
	دستیوں	شکنوں	آمیز	ثانیہ	یاری
نشین	دستیاں	نگاہی	آموز	خارجہ	یاتی
نیت	بینیاں	یافتہ	کناں	دہانی	ترین
سازی	خواروں	انگیز	نگیں	شناسی	شناس
یناس	بہتہ	ڈہال	ہندی	مندان	تامہ

Affix words List					
کاری	پرداز	پتیاں	گاہی	گساری	ورزی
تانی	ابٹیں	افشاں	خانے	کشیدہ	پرور
آزما	انگیزی	ستانی	وانے	رسیدہ	آوری
گوار	اندوزی	اندیش	واس	انداز	فروش
وانا	بازیوں	دہندہ	باش	ورسٹی	بینی
خانہ	بازیاں	نمائی	داد	نگاری	مقدس
ٹھاک	کاریاں	تھانی	بین	گزاری	کدوں
	ناکیوں	فشانی		شدگان	فشاں
	ناکیاں	فروشی	جات	پرستی	خوان
پنات	کثیر	سرائی	اری	منڈی	گداز
پنیت	بندیوں	رسانی	یچہ	کارہ	ناکی
پرست	بندیاں	پروری	یئن	یالی	نشیں
پذیر	خیزیوں	نشینی	فس	ٹائی	کردہ
پسند	خیزیاں	گرافز	نوش	آیند	جوئی
یچوں	گواریت	اندوز	روی	دستی	خیزی
شکنی	پذیروں	پرداز	توز	صاحب	گزیں
صاحبہ	پذیری	بریوں	بین	آگیں	خواں
دوست	پارینہ	گریوں	کشی	خورہ	داری
خواہ	خوانوں	سراؤں	آرا	بانی	نژاد
یابی	دہندوں	آلودہ	زدا	پسندی	رساں
فراز	داریاں	آمیزی	نگر	زدگی	فکری
آزاد	نویساں	ماراں	راں	نویس	خوار
پناہ	نگاروں	حیراں	کدے	گوئی	طراز
گانہ	دانیوں	نگراں	انے	گیری	گسار
انتا	دانیاں	میراں	ویز	وائی	کوشی
اننا	دیوانہ	باراں	وتی	کراں	سوزی
انتی	گوئیوں	الدین	اتی	گراں	آلود
گزار	گوئیاں	شبانہ	خور	آباد	گیوں

Affix words List					
نفسی	آمے	گین	ترانہ	اندازی	ٹاٹتی
دیدہ	زاتوں	نہا	ہشانہ	گاریاں	اٹڑڈ
نگار	الو	ندی	گرانہ	اسیویں	یتیں
زدوں	دان	کدہ	مٹانہ	سازانہ	یتوں
چاری	ناک	کشائی	کار	پسندوں	زادہ
سنجی	ساز	نامہ	افراز	دوگانہ	سرا
بختی	پوش	آور	افروز	ترکانہ	زدی
رنگی	گیر	یاس	پزیری	علمانہ	شکن
وایا	حور	پنا	روائی	الہانہ	بند
سویں	چیں	گار	آفریں	زمانہ	خیز
ٹانہ	ریز	وان	گاران	دارنی	زنی
نما	سوز	دار	انائی	زدگان	گری
حال	زدہ	تاؤ	ہٹیں	نویسی	وار
وری	کشا	اہٹ	اعظم	باز	مندى
کوش	بوس	یاب	اعلیٰ	دوز	بازی

Appendix B (Patterns of affix word)

Patterns id	Patterns	Examples
1	First character replaces with د	اکیلا دکیلا، منا دنا
2	First character replaces with و	پانی وانی، نام وام
3	First character replaces with م	سچ مچ، ننگے منگے، نگا منگا
4	First character replaces with س	غلط سلط
5	First character replaces with ب	ہزاری ہزاری ، حیص بیص
6	First character replaces with پ	لوٹ پوٹ، اونے پونے، اتا پتا
7	First character replaces with ج	مل جل
8	The first two characters, replace with گ	پوچھ گچھ ، پوچھے گچھے
9	The first two characters, replace with ب	سمجھا بجھا
10	The first two characters, replace with م	جھوٹھ موٹھ
11	Second characters ۱ replace with و	خالی خولی، مار مور
12	Second characters ۱ replace with ۱	لوٹ لاٹ
13		
14		
15		

16		
17	Last character replace with م	دھکم دھکا
18		چوری چکاری، چوڑا چک
19		، جھاڑو بہارو
20	Last character replace with ا	کھوکھا، دھو دھا