HJRS HEC Journal Recognition System Category Y

Vol 2 Issue 3 (April-June 2025)

ISSN (Online): 3006-4740

ISSN (Print): 3006-4732

Distributional Morphological Approach to Word Productivity and Base Isolation in the Urdu Language

Nimra Saeed

MPhil Applied Linguistics (scholar), Department of Linguistics and Communications, UMT, Lahore
Email ID: f2024084003@umt.edu.pk

Muneeba Jamil

MPhil Applied Linguistics (scholar), Department of Linguistics and Communications, UMT, Lahore
Email ID: f2024084011@umt.edu.pk

Ms. Shumaila Ahmad

PhD Linguistics, Department of Linguistics and Communications, UMT, Lahore
Email ID: shumaila.ahmad@umt.edu.pk

Abstract

This study explores the morphological complexity of the Urdu language through the lens of the Distributional Morphological (DM) framework proposed by Halle and Marantz (1993). Urdu's word formation patterns, deeply influenced by Persian and Arabic grammatical systems, present both theoretical and applied challenges for linguists. Using a qualitative methodology with a descriptive and exploratory design, this research aims to analyze word productivity and the identification of monomorphic bases in standard Urdu. The findings demonstrate that Urdu exhibits a highly systematic yet complex morphological structure, where monomorphic roots can be effectively isolated and analyzed. The implications of this study extend beyond theoretical linguistics to computational applications, including enhanced lexical identification, corpus development, improved translation software, and the reduction of machine translation errors. While the scope is limited to the standard dialect of Urdu, the study opens pathways for future investigations into regional varieties and cross-linguistic

morphological comparisons, particularly with closely related languages such as Hindi and Persian.

Key Words: Distributional Morphology, Word Productivity, Base Isolation, Urdu Morphology, Linguistic Analysis

Introduction:

The Urdu language was called Lashkari (camp language) in the Mughal reign. It was named so because of its versatility and carrying words from multiple other languages. The morphology of this language is as complex as any other language of the world. It expresses various morphological phenomena and proves that it also has richness and adaptability like the other languages. Word productivity and base isolation in the Urdu language mirror its morphological intricacies. This qualitative study focuses on linguistic analysis to look for productivity and base isolation in the Urdu language. It delves into the syntactic process of Urdu language to observe that how new words are created and root words are isolated as single entities. By examining derivation, inflection, compounding, and reduplication, the productivity is measured. The borrowed techniques are taken into consideration as well to check how they lead toward the Urdu lexicon expansion

Words with a similar form that do not fall into morphological makeup are also discussed under the category of potential false analysis. This is an issue that is commonly faced in machine translations. The purpose of this study is to separate the base forms and identify the words that can be morphologically altered. This can help in machine translations, Urdu lexicography, and computational linguistics. In Pakistan Urdu has not gained the status of a national language even though it is the first language of a large number of inhabitants. According to Rehman (2004) and Grimes (2000), this language is the second most spoken language in the world. In contemporary times, Urdu has not been treated well. The vocabulary is not well-established, words are majorly borrowed especially for technical terms and no widely recognized research is being done.

Azim (1975: 259) states that Urdu has been a victim of “genocide” over the years, and the government has not paid attention to it. This research is significant in analyzing Urdu morphology from the perspective mentioned above. It can be helpful in updating translation software and aiding computational linguistics. This study highlights the morphological intricacies of the Urdu language. This research on morphological base isolation and word productivity in Urdu carries theoretical, linguistic, and practical importance. Urdu, a morphologically complex language spoken by millions of speakers globally yet, remains underexplored in terms of its morphological structure and processes. By focusing on base isolation and morpheme distribution, this study explains a key gap in the linguistic

understanding of Urdu, contributing to a profound understanding of its morphological system and its role in linguistic variety.

The analysis of word productivity and the morphemic structure also depicts how new lexical items are produced in Urdu, expressing meaning intricacies. This study also serves the linguistic theory while showcasing the structured system of semantic diversity and word formation. The findings of this study are beneficial for the understanding of Urdu's linguistic process which strengthens the productivity of a language and serves advanced and evolving vocabularies.

This study also shows the significance of the isolated bases identification in the Urdu language. These units serve as the foundational blocks of morphology and their identification is essential for understanding the distinction between monomorphemic and multimorphemic words. These findings serve as a basis for development of accurate morphological models that enrich linguistic theory. There is also a possibility for cross-language analysis such as Urdu with Persian and Urdu with Hindi.

Practically, these findings suggest improvement in computational linguistics and natural language processing (NLP). These precise methods, if developed, can serve for improving technological features such as machine translation, automatic text processing and speech recognition. This evolution is significantly important in fields of educational resources, cross-cultural communication, and forming of Urdu language's digital tools, which are non-existent as of yet. This research is a beneficial step towards linking theoretical linguistics and technological applications, promoting a deeper understanding of Urdu language while handling contemporary linguistic and computational challenges.

This research is a beneficial step towards linking theoretical linguistics and technological applications, promoting a deeper understanding of Urdu language while handling contemporary linguistic and computational challenges.

The present study is limited in its nature due to various reasons. The limitations of time and resources hold this study from exploring all the words that exist in Urdu. Hence, limited words' analysis has been provided. Also, there are multiple dialects in the Urdu language which have their own variations in lexicon. This study is in short supply of the regional dialects, dealing with the standard dialect only. Moreover, there is a possibility for cross language analysis for instance Urdu-Hindi or Urdu-Persian to study this morphological aspect.

Statement of the problem

The word productivity in Urdu language is very intricate due to its own expressive nature as well as the borrowing techniques it has adopted from the foreign languages (as discussed

further). Certain words are used in multiple contexts which results in altered syntactic properties of those words in a given context. Similarly, base isolation is a meticulous phenomenon. These lexical items are supposed to be understood properly, especially in the case of language learners and machine translations. These words are mistranslated when translated into a foreign language through computer assisted translations. In this research, along with the word productivity, these isolated bases are explained as well. The distributed framework is used to explain their formation. This type of data can be put into software for precise translations.

Research questions

This research is majorly concerned with finding the answer of the following questions:

- What is the pattern of morphemic structure and distribution in the Urdu language productivity?
- How does base isolation manifests itself in the Urdu language?
- Does the identification of these characteristics potentially aid in the development of linguistic resources?

Literature review:

Morphology is defined as examining morphemes and their positioning in forming words (Eugene Nida, 1948). It suggests that the main focus of morphology is on word formation. It is a source of expanding the word bank of any given language in a systematic and rule-governed manner. This expansion also aids in producing stylistic variations in the discourse of that language. Fischer (1998), defining the word formation, stated that the word formation is an expression that has lost its standing of infrequent formation but it is still the one. And it is considered novel by most of the representatives of a speech circle. This is a morphological process and new words are introduced in the languages frequently, bringing forth a new concept (Algeo, 1980; Lehrer, 1996). Saleem and Ahmad (2025) explore Urdu compounding through Lieber's Lexical Semantic Framework, demonstrating its applicability to Urdu and revealing novel insights into the language's compound formation and semantic relationships. This is helpful for researchers to inspect linguistic innovations deeply. Base isolation supports the analysis of roots by separating them from affixes and explaining the word formation.

Khan (2013), in his thesis, states that the exhibition of word formation is primarily linked with morpheme-based morphology. In this approach, the words and their integral parts are put under analysis. He labels morphology as a given set of principles that transforms the morphemes into words. Further, he describes the morphemes as the integral part of words, defined as the nominal linguistic units that give any lexical or grammatical meaning. He also proved the generalization about the Urdu suffixes with examples, one being that the suffixes decide the word category in the Urdu language. This study explains how productive suffixes are a source

of expanding the lexical bank of the Urdu language. Hardie (2003: 35), talking about the inflection in the Urdu language, shared that it is mainly deployed on the suffixation where the suffixes are integrated, vigorously composed of a syllable, or a unit of vowel. Suffixes in the Urdu language can also represent various attributes such as gender, number, and cases marked on the nouns. The analysis of these suffixes can play an important role in the identification of word productivity in the Urdu language.

The pattern of change in Urdu nouns is explained by Moizuddin (1989: 20) as there are *declinable and indeclinable* nouns, that may or may not change structurally, multiple constructional substitutions are compatible with the number and the gender agreement, and the rest of the changes are based on the relationship of noun to its case markers. This change of pattern is identifiable and follows a set of rules. However, there are some exceptions as well in the morphological synthesis which are explained in the discussion. These systematic variations can be studied in the word productivity patterns. Moreover, Urdu morphological structures mainly come from Persian, Arabic, and native Urdu (Islam, 2011). He also states that the Urdu language has a range of native and loan derivational affixes. There are only a couple of native affixes which are mainly vowel-based but manifest in a lot of words. Likewise, a lot of Persian suffixes are borrowed, but they link with only a limited number of words. All loan affixes have semantic properties, but few of them do not make a semantic change. This affixal diversity in the Urdu language aids in shaping the Urdu lexicon.

Urdu compounds can have both left as well as right to the root branching. This combination is made up of synonymous or near-synonymous words. The compound can contain the words of two different sources as in *Zulm-o-sitem* where the initial is Arabic and the latter is Persian. This compound also contains an infix -o- which is a technique borrowed from Arabic. This acts as a linking source between constituents. Verbal compounds in Urdu do not take infixes rather have dummy verbs which is a form of modified Persian auxiliary. The word productivity brings both derivational and inflectional changes in the Urdu language where the inflectional transformations are frequently recurrent and follow the native Urdu pattern. (Islam, 2011).

He also highlights that along with the lexical borrowing from the Persian and the Arabic, Urdu borrowed their derivational morphological patterns as well. Thus, the same word can be of various derivational categories. This practice, over the years, has become a part of the Urdu morphology. The Urdu language has also borrowed the English loan affixes to make plurals with -s, and -ies ending. However, this mostly applies to the words with -a and -I ending. For example, *mull(a)h**s**, vader(a)**s**.* Matras (2009) states that languages adopt morphological adaptations from other languages to accommodate complex loan verbs. Through the distributive morphological approach, the phonological as well as morphological adaptations of the foreign affixes' incorporation can be examined.

Urdu is a morphologically rich and complex language, having multi-word phrases, inflectional and derivational morphology, and letters that change shape depending on their positional context (Shafi et al., 2023). Both the free morphemes or the meaningful units able to stand alone, and the bound morphemes, or significant units that only function if attached to other morphemes, are part of Urdu morphology. Examples of bound morphemes include prefixes, suffixes, and linking morphemes, while free morphemes can occur alone or as part of a compound word. One of the morphological features shared widely among Urdu is reduplication. Here, the words partially modify using the structure of an echo or totally repeating; for example, "DN BDN" stands for "day after day" (Shafi et al., 2023).

A constant omission or incorrect positioning of spaces in the Urdu cursive Nastalik script blurs the boundaries between the words. For instance, in a lack of contextual markers, compound words get misconstrued. The language expresses gender, case, number, and verb forms through morphology, which makes computational analysis a bit more difficult. Case markers and postpositions, for example, are essential to Urdu grammar but are often expressed as separate entities, requiring careful tokenization handling. According to their morphological background, some words have several meanings. For example, how the context necessitates a morphological intervention is shown with the word "bashy" which might point to "citizens" (noun) or “stunning” (Shafi, et al. 2023).

Urdu, as an Indo-Iranian language with roots in the Indo-Aryan family, has a various vocabulary influenced by languages such as Arabic, Persian, and Sanskrit, making its diminutive morphology particularly complex (Mangrio, 2016). Urdu diminutives are primarily formed through suffixation, with common suffixes like **"-i"**, **"-a"**, and **"-ri"** applied to nouns and adjectives. For example, "kutta" (dog) transforms into "kutti" (bitch), and "phool" (flower) becomes "pankhri" (petal). These morphological operations are not restricted to nouns but also the adjectives, as in "bara" (big) becoming "bari" (big), demonstrating Urdu's productive diminutive system (Bögel et al., 2008). Distinct to the other languages, Urdu's diminutives repeatedly necessitate gender markers, with the suffix representing the gender associated with the word’s root. This is a feature very important in Urdu’s grammatical structure.

The socio-pragmatic role of Urdu diminutives is also emphasized upon in the study of Batool & Saleem (2023). These forms are very doubtlessly used in informal settings, especially in child-directed speech, to display affection, compliment, or understanding. Diminutives in Urdu language, for example, "muni" (little girl) or "muna" (little boy) are vocative expressions often used within family or close social circle, showcasing the cultural positioning of diminutives (Lockyer, 2014; Batool & Saleem, 2023).

Moreover, diminutives in Urdu convey metaphorical and connotative meanings, adding layers of implicit information. For example, "naala" (drain) diminishes to "naali" (small drain or tube)

but can also metaphorically suggest undesirable traits when used contextually, such as in the phrase “gandi naali ka keeda” (sewage worm) to indicate despicable character traits (Batool & Saleem, 2023).

Methodology:

Research Philosophy:

This research falls under the category of constructivist philosophy. It deals with the study of lexical productivity in the Urdu language. It looks into the formation of words, their structure, and their peculiarities. Language is constructed through the contact of speakers of the same as well as different languages. This proves that a language is not absolute rather it is transformative, generated through various phenomena. This approach can successfully capture and explain the complexity and nuances of the language. Through shared knowledge, language users bring innovations into the already existing lexicon and shape their current knowledge. It helps explain how the users actively construct and negotiate the meaning of words, making this study relevant and insightful to the research on linguistic change.

Research Paradigm:

The study under discussion is descriptive as it describes the pattern of lexical production. It records and organizes this information in a useful pattern. This approach proves to be helpful in cases of language documentation. The paradigm of this research is interpretivism. It deals with the speakers’ perception and interpretation of the meanings. It also analyzes how the speakers attach meaning to the new or derived words. It delves into the speakers’ subjective associations as well to study the nuanced meanings.

Research approach and design:

This research uses an inductive approach as it undertakes the specific-to-general route to explain how the words productivity and base isolation manifest in the Urdu language. Being data-driven research, it uses real-life lexical items from the Urdu language. Having an exploratory nature, it does not attempt to test any hypothesis but rather explains the general lexical productivity. By the analysis of a limited data set, it identifies the basic trends in this area of linguistics. This approach is useful specifically for linguistics-related studies to identify the patterns of production and change.

Data Set:

This study analyzes a word set of 41 items, in which each phenomenon of productivity and base isolation is explained through five examples. The limitation of time and resources as well as the less frequent occurrence of some phenomena in Urdu hold this research to the analysis of this limited number. The data source is Feroze-ul-Lughat, the most authentic dictionary of Urdu to exist in Pakistan. The data set consists of items that best convey the understanding of the phenomenon under study. Furthermore, the data set was authenticated by an expert in the Urdu

language who is also a native speaker. The expert serves as associate head in the linguistic department of a university. He also has expertise in the research on Urdu morphology.

Technique of Data Analysis:

The data set in this study is analyzed through the framework of distributed morphology (DM) that is grounded in the findings of generative linguistics. The theoretical framework under consideration was proposed by Alec Marantz and Morris Hale in 1993, later used by Arregi and Nevins (2006), E. A. Ofori (2016), and Marinaccio (2020) in their articles. It intermingles syntax, morphology, and phonology to study the structure of a word and its explication. DM views the morphological processes as a blend of various intricate systems which do not exist in isolation. The important characteristics of DM are late insertion, syntactic hierarchy, under-specification, competition, and distributed representation. It segregates the hypothetical features such as ([+Past], [+Plural]) to explain the symmetrical process of word formation. The suppletive forms' analysis is done through competition and vocabulary insertion features of DM. These characteristics make DM a flexible process and advantageous to study lexical productivity. The analysis through DM is a five-step process involving morphosyntactic decomposition, syntactic structure formation, vocabulary insertion, morphophonological adjustments, and surface form generation. An example can be taken into consideration to explain how DM explains the past form synthesis. The regular past tense form "talked" of the verb *talk* is created as follows:

Morphosyntactic Decomposition:

- Root: √TALK
- Tense Feature: [+Past]

Syntactic Structure:

- [TP [vp √TALK [T +Past]]]

Vocabulary Insertion:

- The root √TALK remains *talk*.
- The [+Past] feature is realized as *-ed*.

Morphophonological Adjustment:

- No adjustments are necessary for regular past form verbs like *talked*.

Surface Form:

- The resulting item is "*talked*".

Data analysis:

Word productivity in the Urdu language is majorly carried out by derivational and inflectional morphology and compounding. The mentioned procedures are followed by various subtypes which are discussed further.

Derivational morphology in Urdu:

Derivational morphology, in the Urdu language, helps create new words with a grammatical category different than the root word. In some cases, the meaning is altered as well. This is done by affixation: prefix and suffix, or combining the roots of the different words. Urdu has multiple productive derivational morphemes which the speakers use to form new words frequently.

Examples:

a. -ی (ī): This suffix can be added to nouns or adjectives to form an adjective that expresses a characteristic. For instance:

Root Word	Derivation	DM Analysis
علم (Ilm, "knowledge")	علمی(ilmi,"knowledge-based")	[√ILM + DER (-i)] Smooth concatenation; no significant adjustment is needed.
انسان (insaan, "human")	*انسانی* (insaanee, "human-related")	[√INSAN DER (-i)] Vowel elongation (-i → -ee) for phonological fluency.
کتاب (kitab "book")	کتابی (kitabee "bookish")	[√KITAB + DER (-i)] Vowel elongation (-i → -ee) for phonological fluency.

The words on the left side are the root words that can take various types of endings and transform them into plurals or adjectives. This ending is also a gender marker for feminine, as in:

- لڑکا (larka "boy") → لڑکی (larki "girl")
- دہی

- لکڑی

Nouns like لکڑی, and دہی are taken as feminine in the speech and writing due to their ی ending.

However, this rule does not apply to all Urdu words altogether. There are Urdu words that have **ی- (ī)** ending but are not nouns transformed into adjectives. For example,

کمیٹی - kameti (committee)

نئ - naye (new)

روٹی- roti (flatbread)

روتی – roti [crying (for female)]

پانی – Paani (water)

The above-mentioned words are nouns except نئی and روتی which are the adjective and the verb respectively. This type of exception can lead to the potential false analysis by the non-native speakers.

b. آنہ - (ānā): this Persian borrowed suffix is added with nouns and adjectives to form new quality describing adjectives.

Root Word	Suffixation	DM Analysis
دوست (dost "friend")	دوستانہ (dostaana "friendly")	[√DOST + DER (-ana)] No phonological adjustment; regular concatenation.
ظالم (zaalim "cruel")	ظالمانہ (zaalimaana "cruel")	[√ZLM + DER (-ana)] No phonological adjustment; regular concatenation.
احمق (ahmaq "stupid")	احمقانہ (ahmaqana "stupid")	[√AHMAQ + DER (-ana)] No phonological adjustment; regular concatenation.

The first two words are nouns while the rest are adjectives. In Urdu, the آنہ- ending is used with adjectives when these words complement the nouns and verbs. Unlike English, where the same form is used. For example, حرکت ظالمانہ فیصلہ, احمقانہ etc. These adjectives compliment the verbs and a noun respectively.

There is also a potential chance for false analysis with words like سکینہ,زمانہ,دانا,بہانہ,جانا etc. The given words are nouns and verb, can only be changed into plurals.

Inflectional morphology in Urdu:

Inflectional morphology in Urdu is used to create a different word class to represent tense, case or number rather than creating new words with different meanings. Urdu has a rich scope of morphological productivity displayed through the ways by which verbs and nouns are used for multiple reasons.

4.2.1. Pluralization*: Urdu language uses the suffixes ے (ẽ) and*یں *(õ̃) to make plurals:*

Root Word	Inflection	DM Analysis
کتاب (kitaab, "book")	کتابیں (kitaabein, "books")	[√KTB + INFL (-*õ̃*)] Nasalization of -õ̃ in plural formation to mark number.
بچہ (bacha, "child")	بچے (bachay, "children")	[√BCHY + INFL (-*ẽ*)] Smooth concatenation with nasalized -ẽ.
پودا (poda, “plant”)	پودے (poday, "plants")	[√PODA + INFL (-*ẽ*)] Smooth concatenation with nasalized -ẽ.

However, the ے ending is used more in day-to-day speaking. Whereas, یں ending is used in a formal setting. Also, it depends upon the speaker and his choice of words according to the situation.

Infixation, a borrowed technique from Arabic, is also used in Urdu to make plurals, as in:

حالت(haalat “condition”)	حالات(halaat “conditions”)	[√HALAT + INFL(-I infix)] Infixation (-aat) with vowel elongation for pluralization.

In some cases, the -s suffix is also used for pluralization, like English, with the Urdu lexical items. For example, *mullah(s), taliban(s), jihadi(s).*

It is to be noted that words like چنیں, دیکھیں, سکھائیں, بتائیں, لکھیں which have the syntactic makeup of plurals, are used to give respect to the listener. These words are used mostly while talking to adults and have a cultural association to display solidarity.

4.2.2. Verb Conjugation: Verb roots are combined with the word endings that reflect the mood, tense, politeness, and aspect.

Examples:

Root	Inflection	Morphological change	DM Analysis
آنا (aana, "to come") (infinitive)	آیا (aaya, "came") (past tense, third person, masculine singular)	- یا suffix	[√AA+ PST (-YA)] No phonological adjustment; regular suffix addition.
کرنا (karna, "to do") (infinitive)	کرتا ہے (karta hai, "does") (present tense, third person, masculine singular)	- تا (suffix + auxiliary ہے for present	[√KAR + CONT (-TA) + AUX (HAI)] Smooth concatenation of suffix -ta and auxiliary verb hai without modification.
پڑھو (parho "read" informal) (infinitive)	پڑھیں (parhen “read”) (formal/polite imperative, second person plural)	و (informal) or - یں (formal) suffix	[√PRHO + IMP (-AIN)] Suffix -ain adds politeness with elongation of root vowel a for phonological harmony

These changes of form in Urdu exhibit its productive nature and rules are applied to gain wider purposes from the limited words.

4.2.3. Negatives :

-لا(laa-), -بے, -نا, and -ان prefix is added with words to make their negatives or to mark the absence of the phenomenon to which this prefix is added. For example:

Root Word	Inflection	DM Analysis
حاصل (haasil “attain”)	لاحاصل (laa haasil “unattainable”)	[√HSL + INFL (-LA)] No phonological adjustment; regular concatenation.

قابو (qaabu "control")	بے قابو (beqaabu "out of control")	[√QABU + INFL (-BE)] Smooth concatenation; no significant adjustment.
لائق(laaiq"worthy, intelligent")	نالائق (nalaaiq "unworthy, incompetent")	[√LAIQ + INFL (-NA)] No phonological adjustment; regular concatenation.
دیکھا (dekha "to see")	اندیکھا (andekhaa "uncertain, unforseen")	[√DKHA + INFL (-AN)] Addition of nasal prefix an- without phonological alteration.

4.2.4. Compound words:

Another method to enhance word productivity in the Urdu language is by compounding. In this process, multiple words are grouped and assigned different meanings. There are various types of compounding as discussed below:

4.2.4.1. Noun + Noun Compounds:

These compounds are formed by combining two nouns, often to express a unified concept.

Examples:

First item of compound	Second item of compound	DM Analysis
گھر (ghar "home)	بار (baar, "work")	[√GHR + √BAR]. Household. Smooth concatenation; no phonological adjustment needed.
شور (shor, "noise")	شرابہ (sharaba, "uproar")	[√SHR + √SHRBA]. Commotion. Smooth concatenation; no adjustment needed.
ریلوے (railway)	سٹیشن (station)	[√RAILWAY+ √STATION]. No phonological adjustment; borrowed words retain original structure.

4.2.4.2. Adjective + Noun Compounds:

These compound words are created by combining an adjective with a noun to form a new word group with a modified meaning. These words are commonly used in everyday conversations in Urdu language.

Examples:

First item of compound	Second item of compound	DM Analysis
کالے (kaalay, "black")	بادل (baadal, "clouds")	[√KALY + √BADL] No phonological adjustment; regular concatenation of adjective and noun.
گرم (garam, "hot")	پانی (paani, "water")	[√GRM + √PNI] No phonological adjustment; regular concatenation of adjective and noun.
سفید (safed, "white")	پوش (posh, "wear")	[√SFD + √POSH] No phonological adjustment; smooth concatenation.

4.2.4.3. Verb + Noun Compounds:

In this word structure, a verb is combined with a noun to mostly give the noun an active quality or a purpose. These are colloquial expressions used by working people to express their daily struggles.

Examples:

First item of compound	Second item of compound	DM Analysis
کام (kaam, "work")	کاج (kaaj, "work")	[√KAM + √KAJ] No phonological adjustment; reduplication of synonymous elements.
دوڑ (daur, "run")	دھوپ (dhoop,"sunlight")	[√DOR + √DHUP] No phonological adjustment; smooth concatenation.
پانی (paani, "water")	پینا (peena, "drink")	[√PANI + √PEENA] No phonological adjustment; verb-noun combination.

4.2.4.4. Reduplicative Compounds:

Reduplication, in the Urdu language, involves repeating or using similar-sounding words to enhance or emphasize the meaning. These words also show the speaker's annoyance, frustration, excitement, and desire. This is a widely used practice among Urdu native speakers. Famous and commonly used reduplicated Urdu words are listed below:

Examples:

First item of compound	Second item of compound	DM Analysis
تھوڑا (thora, "less")	تھوڑا	[√THRA + √THRA]. No phonological adjustment is needed. Repitition for emphasis.
سچ (sach, "true")	سچ	[√SCH + √SCH]. No phonological adjustment is needed. Repitition for emphasis.
کالا (kaala, "black")	سیاہ (siyah, "dark")	[√KALA + √SIYAH] No phonological adjustment; near-synonymous adjective pairing.

4.2.4.5. Metaphorical or Idiomatic Compounds

This category of compound words is often idiomatic and carries a metaphorical and non-literal meaning of the individual words. These words are used in highly formal, professional, and poetic discourse. These compound words are usually created by combining two words of the same nature which may or may not have *and* added in between.

Examples:

First item of compound	Second item of compound	DM Analysis
ہوش (hosh, "consciousness")	حواس (hawaas, "senses")	[√HSH+ (O) +√HWS] Smooth concatenation; no adjustment needed.
جسم (jism, "body")	جان (jaan, "life")	[√JSM + (O) √JAN] Smooth concatenation; no adjustment needed.
آب (aab, "water")	ہوا *(hawa, "air")*	[√AAB + (O) √HAWA] Smooth concatenation; no adjustment needed.

As the English translation of the words as mentioned above displays, these are metaphorical and idiomatic expressions.

Base isolation in the Urdu language:

Another misconception that can arise in the morphological analysis by the beginner level or non-native speaker of the Urdu language is base isolation. The study above shows how affixation and compounding are performed in Urdu to get various results. However, several words have a lexical makeup like the affixed words. The words may carry a few properties of the affixed or compound words, but they are actually single entities. This is explained with the help of given examples:

Word	Synthesis	DM Analysis
نایاب (nayaab): rare	Na (no) + Yab(available)= Not Available (rare).	[-na + √YAAB] No phonological adjustment; regular concatenation.
نابود (nabood): Obliterate	Na(no)+bood(existence)= Nonexistent (obliterate).	[-na + √BOOD] Smooth concatenation without phonological changes.
ناگزیر (naguzeer): inevitable	Na(no)+Guzeer(avoid)= Not avoidable (inevitable).	[-na+ √GUZEER] No phonological adjustment; regular concatenation.
کتابچہ (kitaabcha): booklet	Kitab(book)+Cha(small)= small book (booklet).	[√KTB + -cha] No phonological adjustment; suffix -cha smoothly integrates with root.
بندہ (banda): Worshipper	Band (close)	[√BND + -a] No phonological adjustment; suffix -a smoothly combines with root.
آنسو (aansu): Tear	Aan(eye)+Su(from)= Coming from eyes (tear).	[√AAN + -su] No significant phonological adjustment; suffix -su

		integrates smoothly with root.
بیوقوف (beywaquf): Stupid	Bey(without)+Waquf(consciousness)= unconscious (stupid)	[-bey + √WAQUF] No significant phonological adjustment; prefix be- integrates smoothly.
دستیاب(dastyaab): Available	Dast(hand) + Yab (available)= At hand (Easily available).	[√DAST + -yaab] No phonological adjustment; suffix -yaab integrates smoothly with root.
بوسیدہ (boseeda): Decayed	Bos(decayed)+yada(state)=decayed state (decayed).	[√BOS + -eeda] Vowel elongation in -eeda for phonological harmony with root.

The first three words with **نا-** prefixes are not the negatives of their respective counterparts. In base isolation, the roots are borrowed from other languages, mainly from Persian in Urdu. This derivation followed by the morphological attachment, stands as a unit. These instances demand expertise and careful observation of the text under study to avoid misinterpretation and false analysis. Thus, these words are unbreakable into morphemes and stand as one lexical unit.

Discussion:

The findings of this study shed light on the morphological productivity and base isolation in the Urdu language, contributing to the broader linguistic discourse and addressing the research questions outlined in the study. This discussion contextualizes the findings within the theoretical framework of Distributed Morphology (DM) (Halle & Marantz, 1993) while linking them with previously produced research and their practical implications.

The study's first research question on "how are morphemes structured and distributed in the Urdu language productivity?" was addressed by the analysis of the derivational and inflectional morphology of Urdu. The findings provide an evidence that Urdu, much like Persian and Arabic, relies heavily on affixation processes, particularly suffixation, to generate new words (Islam, 2011). The -ی (-ī) suffix, as evidenced in examples such as علمی → علم (ʿilm → ʿilmī) and انسانی → انسان (insān → insānī), turns nouns into adjectives, showing a productive morphological pattern.

Furthermore, these findings align with those of Hardie (2003), who brought the readers' attention to the role of suffixes in determining word categories in Urdu. The آنہ (-ānā) suffix, which forms adjectives like ظالمانہ → ظالم (ẓālim → ẓālimānā) and احمقانہ → احمق (aḥmaq → aḥmaqānā), exemplifies how borrowed Persian morphological structures have been integrated into the Urdu lexicon (Moizuddin, 1989). The study also found that prefixation plays a crucial role in forming negatives, as seen in بے قابو (beqaabu, "out of control") and نالائق (nalaaiq, "unworthy"), a process that aligns with Shafi et al. (2023), who emphasized the complex role of affixation in Urdu word formation.

The second research question of this study titled as *what does base isolation manifest in the Urdu language?* was substantiated through the analysis of morphological decomposition. The study established the fact that certain lexical items appear to be affixed but are monomorphemic units, causing false morphological analysis. This was spotted in نایاب (nayaab, "rare"), نابود (nabood, "obliterated"), and ناگزیر (naguzeer, "inevitable"), where the نا (nā-) prefix does not serve as a negative marker but as an integral part of the word.

This observation validates the findings of Islam (2011) and Matras (2009), who provided evidence in their study that Urdu has adopted Persian and Arabic morphological features, which makes its base isolation more detailed. Furthermore, the findings of this research show the potential risk of misinterpretation in machine translation and computational linguistics, as previously depicted in the study of Shafi et al. (2023). This displays that a thorough comprehension of the Urdu language's base isolation can lead to more precise lexical parsing in NLP models which can definitely refine the machine translation to Urdu as well as the corpus development.

The third research question in this study *how can the identification of these characteristics potentially aid in the development of linguistic resources?* is specifically helpful in the field of natural language processing (NLP) and Urdu lexicography. Identification of productive morphemes and the isolated bases provides structured linguistic data which can potentially enhance the morphological analyzers and spell-checking algorithms.

Additionally, the morphological variations in Urdu affect the seamless understanding for a language learner, as mentioned earlier in the study of Batool & Saleem, 2023; Saleem & Ahmad, 2025). The differences between compound words (e.g., شور شرابہ [shor sharaaba, "commotion"]) and idiomatic constructions (e.g., آب و ہوا [aab-o-hawa, "climate"]) expresses the need of the systematic approach to Urdu morphology. This innovation can benefit the language learners as well as the computational models.

This study emphasizes on the morphological richness of the Urdu language through its word productivity structure and the base isolation intricacies. The findings of this research emphasize

on the findings of previous morphological theories while offering practical interpretations for Urdu's linguistic applications. This research adds to the profound understanding of the morphology of Urdu language. It also plays a significant role in linguistic advancement and technological evolution by addressing leading issues in lexical analysis and computational linguistics.

Conclusion:

In a nutshell, the dynamic and adaptable nature of the Urdu language is shown by explaining the native and borrowed morphological processes. It is noteworthy that the Urdu language is indeed the camp language because of its Arabic, English, and Persian adaptations. It has explained the word formation process while also providing a thorough understanding of how root words can be segregated from the available lexical items. These findings not only serve the theoretical purpose, but can potentially lend service in the field of teaching, lexicography, and the development of translation software. This research also serves as the basis for future studies on the morphological patterns of Urdu as well as other South Asian languages. Also, it can majorly improve the machine translation, automated word recognition, and the development of language resources adapted to the Urdu's distinct linguistic landscape.

References:

Algeo, John (1980). "Where Do All the New Words Come from?". American Speech 55: 264-277

Arregi, K., & Nevins, A. (2006). *A Distributed Morphology analysis of present tense auxiliaries in Zamudio.* University of Chicago.

Azim, A. (1975) "Urdu - A Victim of Cultural Genocide?" In Zafar Imam (ed.) Muslims in India. New Delhi: Orient Longman, pp. 256-272.

Emmanuel Amo Ofori. Distributed Morphology of Akan-Twi Plurals. International Journal of Language and Linguistics. Vol. 4, No. 2, 2016, pp. 57-60. doi: 10.11648/j.ijll.20160402.12

Grimes, B. 2000. "Pakistan". In Ethnologue. Languages of the World. 14th ed. Dallas, Texas Summer Institute of Linguistics.

Lehrer, A. (1996). Identifying and Interpreting Blends: An Experimental Approach. Cognitive Linguistics 7, 359-390.

Halle, M., & Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111–176). MIT Press.

Hardie, A. 2003. The Computational Analysis of Morpho-Syntactic Categories in Urdu, PhD Thesis, Lancaster University.

Islam, R. A. (2011, October). The morphology of loanwords in urdu: The Persian, Arabic ... https://core.ac.uk/download/pdf/153776959.pdf.

Khan, Mohsin (2013). Word Formation in Urdu. *Core.Ac.Uk*.

Marinaccio, M. (2020). *Something from nothing: A Distributed Morphology analysis of null morphology in feminine plural verbal inflections in Modern Standard Arabic* (Honors thesis, New York University). Retrieved from https://as.nyu.edu/content/dam/nyu-as/linguistics/documents/Michael%20Marinaccio%20Honors%20Thesis%2020201.pdf

Matras, Y. 2009. "Formal and Functional Aspects of Verb Borrowing". In Morphologies in Contact. An International Symposium Financed by the Volkswagen-Foundation. Universitat Bremen.

Moizuddin, M. 1989. Word Forms in Urdu. Islamabad: National Language Authority.

Nida, Eugene (1948). The Identification of Morphemes. Indianapolis: Bobbs-Merrill.

Rehman, T. 2004. "Language Policy and Localization in Pakistan: Proposal for a Paradigmatic Shift, Crossing the Digital Divide". SCALLA Conference on Computational Linguistics 5th to 7th January, Katmandu.

Saleem, T., & Ahmad, S. (2025). From structure to meaning: a lexical semantic framework for Urdu compounding. Humanities and Social Sciences Communications, 12(1), 676. https://doi.org/10.1057/s41599-025-04982-x