

## A Study of Compound Formation and Semantic Opacity in Punjabi

Isra Irshad <sup>a</sup> Musarat Yasmin <sup>b</sup>

**Abstract:** This research aims to investigate how Punjabi compounds are formed and in what way the formation of some of these compounds contributes to semantic opacity. This is qualitative research, where data have been analyzed descriptively. The study identifies and categorizes various types of compounds, including endocentric, exocentric, copulative, and verbal compounds. The study reveals that exocentric compounds are formed without the interfixes and follow patterns such as A+N, N+A, N+V, and A+V patterns. Endocentric compounds are found in N+N, A+N, and N+A combinations. These are found with and without interfixes, though compounding with interfixes is noticed as a rare morphological process in Punjabi. Copulative compounds are observed in N+N, A+A, and V+V patterns. These are also formed without the interfixes. Verbal compounds are seen in N+dv and A+dv. Notably, semantic opacity is found in exocentric and copulative compounds. The research underscores a need for detailed work on semantic opacity in Punjabi compounds.

**Keywords:** Punjabi, Derivational Morphology, Compound Formation, Semantic Opacity

### Introduction

In any language, new words are formed by three major word formation processes: derivation by compounding, derivation by affixation, and derivation by modification of base (Haspelmath & Sims, 2010; Lieber, 2009). The derivation by compounding is a very regular phenomenon in all languages (Bauer, 2003; Scalise & Guevara, 2006). It is also frequent in Punjabi, an Indo–Aryan language (Abbi, 1992; Butt & King, 2006). The present research focuses on the formation of Punjabi compounds.

Punjabi is classified as an agglutinative language having a combination of discrete morphemes to express grammatical relationships (Baart, 2006). It is spoken in Pakistan and India. It is written with two scripts: *Gurmukhi* and *Shahmukhi*. *Gurmukhi* is used in India, and *Shahmukhi* is used in Pakistan. It has 49 consonants and 16 vowels (Malik, 2006). It has received a little attention in the field of linguistics and particularly in morphology (Shackle, 2003; Masica, 1993). This research brings to light some features of Punjabi compounds.

The primary question of the research is concerned with the formation of Punjabi compounds, and the secondary question is to investigate how the formation of some of these compounds creates semantic opacity. Thus, the present research specifically addresses the following research questions:

- In what way are Punjabi compounds formed?
- How does the formation of these compounds contribute to semantic opacity?

### Literature Review

A compound is considered a complex combination of two words called constituents, which generally function as head and modifier. Plag (2000) states that the left-hand constituent modifies the right-hand constituent, called the head. Semantic and syntactic information can be derived from the head. The compound will be a verb if its head is a verb, and if its head is a noun, it will be a noun accordingly. These compounds are formed with the combination of different lexical categories.

Haspelmath and Sims (2010) have given the following list of possible combinations of compounds in the English language: N+N, V+N, A+N, N+V, A+A, N+A. The most productive pattern is N+N, whereas the V+N pattern is quite unproductive. The compounds are mostly right-handed.

<sup>a</sup> Lecturer, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan.

<sup>b</sup> Associate Professor, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan.

In the Indo-Aryan languages, most research has focused on Hindi and Urdu, while Punjabi has received little scholarly attention. For example, Masica (1991) and Kachru (1980) discussed the complex morphology of North Indian languages but did not discuss compounding in detail. Sharma (1998) found compound patterns such as Noun Noun, Adjective Noun, and Noun+Verb.

Moreover, Punjabi shares morphological features with Urdu. Islam (2012) states that Urdu compounds may be left-branching or right-branching. The same can be seen in Punjabi also. Left branching means the modifier comes on the left hand side, e.g. *pokh* 'hunger' + *hātāl* 'strike' → *pokh hātāl* (N) 'hunger strike', and right branching means the modifier comes on the right side, e.g. *qətl* 'massacre' + *am* 'general/common' → *qətl- e- am* (N) 'massacre'. Interfix plays a very significant role in the formation of the latter. In Punjabi, compounding with interfixes is less frequent, which is similar to what is seen in South Asian languages. Shackle (2003) highlighted the relatively analytic structure of Punjabi compared to languages with complex derivational morphology, which may reflect the lower frequency of interfixation. Many compounds are formed with the help of distinctive compounding words. For example, *mar* 'beat/kill' forms various compounds, e.g., *mar + dhar* 'roar' → *mar dhar* 'fighting', *lut* 'plunder' + *mar* → *lut mar* 'plundering and killing', and *mayhi* 'fish' + *mar* → 'hunter of fish'. There are different kinds of compounds.

There are a few works that discuss the structure of Punjabi compounds (e.g., Bhatia, 1993), but little attention has been given to their meaning. The present study fills the gap by semantically categorizing the types of compounds and focusing on the cases of semantic opacity in these compound types.

## Methodology

This is qualitative research, focusing on the morphological structure and semantic interpretation of compound formation in Punjabi. The study aims to investigate the patterns and types of compounds and to explore the examples of semantic opacity in these compounds. The data was collected from various authentic linguistic sources. These include spoken Punjabi from informal conversations, print and digital media, Punjabi grammar references, and contemporary Punjabi literature. Thus, a purposive technique was used to collect instances of Punjabi compounds from these sources. The data was analyzed descriptively.

The data was analyzed according to the following steps. 1. Classification of compounds based on syntactic patterns 2. Identification of interfix used in compounds; 3. Interpretation of compound meanings to detect opacity or transparency.

This study is limited in scope to descriptive observations and does not employ statistical analysis due to the qualitative nature of the data. Further research with a broader dataset and psycholinguistic methods could enhance the understanding of semantic processing in compound structures.

## Data Analysis

This section provides a descriptive analysis of the Punjabi compound structures collected during the study. The objective is to investigate the morphological patterns underlying compound formation and explore how various structures give rise to semantic opacity. The data reveal various types of compounds: endocentric, exocentric, copulative, and verbal. Particular consideration is given to the presence or absence of interfixes and their role in the semantic transparency or opacity of the compounds. The discussion begins with the endocentric compounds.

### Endocentric Compound

In endocentric compounds, the meaning of a compound is the subset of the meaning of the head. The head, from the semantic point of view, is called the semantic head. The element that modifies the semantic head is known as the dependent. The semantic head is inside the compound (Haspelmath and Sims, 2010).

In Punjabi, *pokh hātāl* 'hunger strike', the head is the *hātāl* 'strike' and the *pokh* 'hunger' is the modifier. The compound *pokh hātāl* is the hyponym of the head of the compound. Like that in Urdu, the formation of Punjabi includes N + N, A + N, and N + A patterns. Consider the following configurations:

**a) N + N**

- i)  $pok^h$  'hunger' +  $h\dot{a}rt\dot{a}l$  'strike' →  $pok^h$   $h\dot{a}rt\dot{a}l$  (N) 'hunger strike'
- ii)  $q\dot{s}d\dot{r}$  'worth' +  $qim\dot{s}t$  'value' →  $q\dot{s}d\dot{r}$ -o-  $qim\dot{s}t$  (N) 'worth'
- iii)  $t\dot{o}r$  'end' +  $t\dot{o}r$  'end' →  $t\dot{o}r$ -o-  $t\dot{o}r$  (N) 'very end'

**b) A + N**

- i)  $qabil$  'worthy/capable' +  $y\dot{o}r$  'consideration' →  $qabil$ -e- $y\dot{o}r$  'considerable'

**c) N + A**

- i)  $q\dot{s}t\dot{l}$  'murder' +  $am$  'general/common' →  $q\dot{s}t\dot{l}$ -e-  $am$  (N) 'massacre'

The above-mentioned examples illustrate that there are three structural patterns with some variations in Punjabi. Native Punjabi compounds rarely have interfixes. Some compounds of the Urdu language, with interfixes, are also used in Punjabi. N+N and A+N compounds have two variations i.e. with and without the interfixes -o- and -e- respectively.

Interestingly, the compounds having interfixes are only present in the endocentric class. Islam (2011) has found five functions of two interfixes -e- and -o-. The interfix -e- conveys three interpretations: the genitive marker *ka*, agentive particle *vala*, and phonological links between the constituents. On the other hand, -o- shows two functions: conjunctive particle *or*, and an emphasis. In Punjabi, four functions of the interfixes are observed. For example:

- (i)  $q\dot{s}d\dot{r}$  (N<sub>1</sub>) 'worth' +  $qim\dot{s}t$  (N<sub>2</sub>) 'value' →  $q\dot{s}d\dot{r}$ -o-  $qim\dot{s}t$  (N) 'worth'  
(-o- describes the *or* 'and', a conjunctive particle)
- (ii)  $t\dot{o}r$  'end' +  $t\dot{o}r$  'end' →  $t\dot{o}r$ -o-  $t\dot{o}r$  (N) 'very end'  
(-o- shows an emphasis)
- (iii)  $q\dot{s}t\dot{l}$  (N) 'massacre' +  $am$  (Adj) 'general/common' →  $q\dot{s}t\dot{l}$ -e- $am$  (N) 'a general massacre'  
(-e- is linking the two constituents phonologically)
- (iv)  $qabil$  (Adj) 'worthy/capable' +  $y\dot{o}r$  'consideration' (N) →  $qabil$ -e-  $y\dot{o}r$  'worthy of consideration'  
(-e- describes the genitive marker *ka*)

So, the interfixes in the above-mentioned examples depict various interpretations. In example 1(i), the two constituents are nouns. N<sub>1</sub> is a head and N<sub>2</sub> is a modifier, and interfix, -o-, describes the conjunctive marker *or* 'and'. In 1(ii), -o- merely conveys an emphasis. The infix -e- also conveys two functions. In 1(iii), it links the constituents phonologically, and in 1(iv), it describes the genitive marker *ka*.

In sum, these compounds are easy to understand and clear. Exocentric compounds, on the other hand, are not simple.

**Exocentric Compound**

In exocentric compounds, the semantic head is not inside the compound, e.g., *redhead* is not the type of head, but 'someone who has red hair'. Head-modifier relation is present morphologically, but no head is present from the semantic point of view, e.g., Urdu compound  $m\dot{u}t\dot{h}\dot{i}$   $t\dot{j}\dot{o}ri$  'charming words' (N),  $m\dot{u}t\dot{h}\dot{i}$  means 'sweet' (A), and  $t\dot{j}\dot{o}ri$  means 'knife' (N). In this example, the modifier is an adjective  $m\dot{u}t\dot{h}\dot{i}$  'sweet', and the morphological head is  $t\dot{j}\dot{o}ri$  'knife', but semantically, there is no head.

In Punjabi, these compounds are mostly in A + N structure, but they can be found in N + V and N + A combinations as well. Like that in Urdu, there are no interfixes in Punjabi exocentric compounds. Consider the following patterns:

**a) N + A**

- i)  $h\dot{a}th$  'hands' +  $s\dot{o}r\dot{e}$  'narrow' →  $h\dot{a}th$   $s\dot{o}r\dot{e}$  'poverty' (N)
- ii)  $lu:n$  'salt' +  $h\dot{s}ram$  'forbidden/unlawful' →  $lu:n$   $h\dot{s}ram$  'a dishonest person' (N)

**b) A + N**

- i) *pāndz tən* 'five' *tən* 'bodies' → *pāndz tən* (N) 'Holy Prophet (PBUH), Hazrat Ali, Hassan, Hussain and Hazrat Fatima' / Holy Prophet (PBUH) and his family members
- ii) *meli* 'dirty' + *əkh* 'eye' → *meli əkh* 'bad intentions' (N)
- iii) *qo* 'two' + *tok* 'stop' → *qo tok* 'straightforward' (Adj)

**d) N+V**

- i) *sir* (N) 'head' + *phir* (V) 'turn' → *sir phira* 'mad/crazy' (Adj)

So, the above-mentioned examples show that if a compound is an AN pattern, the resulting compounds can be an adjective or a noun, e.g., *sabit* 'firm' (A) + *qəḍəm* 'feet' (N) → *sabit qəḍəm* 'firm/faithful' (A), and *meli* 'dirty' + *əkh* 'eye' → *meli əkh* 'bad intentions' (N). If the compound is with N+A combination, the resulting compound is always a noun, e.g., *hathth* 'hands' + *soṛe* 'narrow' → *hathth soṛe* 'poverty' (N).

In exocentric compounds, the headedness is not semantic, but morphological. As shown:

- 2a. *hathth* 'hands' (N) + *soṛe* 'narrow' (Adj) → *hathth soṛe* 'poverty' (N)
- 2b. *meli* 'dirty' (Adj) + *əkh* 'eye' (N) → *meli əkh* 'bad intentions' (N)

In the above-mentioned example 2 (a), the Adjective is the modifier, and the noun is the head. But this compound is headless semantically. The noun *hathth* means 'hands' and the adjective *soṛe* means 'something physically narrow'. But the compound *hathth soṛe* refers to 'poverty'. In the same way, the adjective *meli* in 2(b) means 'dirty' and *əkh* means eye, but the compounds refer to the 'bad intentions' (N).

Some compounds in Punjabi reflect merely a culturally specific metaphorical sense, e.g., *pāndz tən* is a commonly established compound in the Punjabi language. *Pāndz* refers to 'five' and *tən* refers to 'bodies'. *Pāndz tən* refers to 'Holy Prophet (PBUH) and his family'. This compound expression carries an intertextual and culturally specific interpretation. It depicts the ideology of *Shia'* community.

Exocentric compound also contains a distinctive constituent. Consider the following configuration:

- 2b) *lu:n* 'salt' + *həram* 'forbidden/ unlawful' → *lu:n həram* 'a dishonest person' (N)
- 2c) *həd* 'bone' + *həram* 'forbidden/ unlawful' → *həd həram* 'a lazy person' (N)

In examples 2(b) and 2(c), *həram* 'forbidden/unlawful' is a distinctive compounding word and forms various compounds. When it is used with the *lu:n* 'salt', it refers to a dishonest person, and when it is used with *həd* 'bone', it refers to a lazy person. Pollatsek and Hyona (2005) state that when the meaning of the compound is not consistent with the meaning of the constituents, then it is called a semantically opaque compound (Pollatsek and Hyona, 2005). So, the above-mentioned examples of exocentric compounds carry the semantic divergence.

In sum, the exocentric compounds exist in various combinations, e.g., A + N, N+ A, and A +V, and contain metaphorical meanings. Unlike the above-mentioned two types of compounds, copulative compounds do not have any head-modifier relation.

**Copulative Compounds**

A copulative compound has two semantic heads, and neither is subordinate to the other. There exists no head-modifier relation. Discussing these compounds of the Urdu language, Islam (2011) posits that these are few in number as compared to endocentric and exocentric compounds. In Punjabi, these are frequent, like the endocentric compounds. Unlike Urdu, these are present without any interfix. The formation of Urdu endocentric compounds includes: N + N, N + A, N + V. These compounds are found in N+ N, V + V, and A + A combinations. The resulting compound is a noun, a verb, and an adjective. Consider the following patterns:

### a. N + N

- i) *dokh* 'sorrow' + *sokh* 'happiness' → *dokh sokh* (N) 'sorrow and happiness'
- ii) *huqa* 'an Eastern smoking pipe designed with a long tube passing through an urn of water that cools down the smoke as it is drawn through' + *pani* 'water' → *huqa pani* 'social acceptability' (N)
- iii) *dal* 'split grain, a common foodstuff' + *roti* 'bread, especially a flat round bread cooked on a griddle' → *dalroti* 'livelihood' (N)

### b) V + V

- i) *aīa* 'coming' + *ḍaīa* 'going' → *aīa ḍaīa* (V) 'coming and going'
- ii) *lot* 'plundering' + *mar* 'killing' (V/N) → *lot mar* (N) 'plundering and killing'

### c) . A + A

- i) *maṛi* 'happy' + *ḡāNgi* 'sad' → *maṛi ḡāNgi* (N) 'happy and sad situation'
- ii) *lamma* 'tall' *ḡōra* + 'broad/ tall' → *lamma ḡōra* (Adj) 'tall and wide'

Semantically, these compounds can be transparent or opaque. Consider the following examples:

- 3 (i) *dokh* 'sorrow' (N) + *sokh* (N) 'happiness' → *dokh sokh* (N) 'sorrow and happiness'
- 3 (ii) *dal* (N) 'split grain, a common foodstuff' + *roti* (N) 'bread, especially a flat round bread cooked on a griddle' → 'livelihood/ hand to mouth' (N)

Both of the above-mentioned compounds contain an N + N structure, but the interpretation of the compound gives the idea that the compound in 3(i) is transparent in meaning, e.g., *dokh* means 'sorrow', *sokh* means 'happiness', and *dokh sokh* simply means 'sorrow and happiness'. However, 3(ii) contains semantic divergence, e.g., *roti* is a kind of bread, and *dal* is a foodstuff used in South Asian countries. These are considered cheap. The compound is used in the expression, e.g., *dal roti ḡādi pei e*, which gives the idea that someone is living from hand to mouth.

Unlike copulative compounds, there are many compounds in the Punjabi language in which the first constituent of the compound can be an adjective, noun, or verbal noun, and the other constituent is a dummy verb; these are termed verbal compounds.

### Verbal Compounds

Islam (2012) posits that there is a distinction between the light verb and dummy verb. Light verb forms the complex predicate. It is used in the V1 + V2 construction, where V1 is the main verb and V2 is a light verb. On the other hand, a dummy verb is combined with an adjective or a noun to form a compound verb, e.g., N/A + dv. In the Indo-Aryan languages, many verbs exist in their lexical form, e.g., *pi*.

On the other hand, because of the absence of the lexical verb, the adjective or noun takes a dummy verb to form a compound verb. A dummy verb can be described as a tool for the formation of verbal compounds in the absence of some base verbs. In Urdu, five verbs chiefly function as main verbs and secondarily as dummy verbs. These include *kār* 'do', *ḍā* 'go', *ho* 'be', *ḍē* 'give', and *le* 'take' (Versteegh 2001, 488). These are also present in Punjabi with many other dummy verbs. Some of these verbs end with a nasalized syllable. Consider the following patterns.

### a) N + dV

- i) *qorbani* 'sacrifice' + *ḍēṛa* 'give' → *qorbani ḍēṛa* (V) 'to sacrifice'
- ii) *qabu* 'power/hold/control' + *kārna* 'do' → *qabu kārna* (V) 'to subdue'
- iii) *baṇ* 'Azan' + *milli* 'to do' → *baṇ milli* (V) 'to crow'

### b) A + dV

- i) *qabul* 'acceptable' + *hoṛa* → *qabul hoṛa* (V) 'to accept'

The above-mentioned examples describe how dummy verbs *ḍēṛa*, *kārna*, *hoṛa*, and *milli* are used with nouns and adjectives to form the verbal compounds.



## Findings Drawn From the Formation of Punjabi Compounds

The description of the formation of Punjabi compounds draws the following findings:

1. Punjabi compounds can be left-branching as well as right-branching.
2. These can be formed with distinctive compounding words.
3. Most frequent types of compounds in Punjabi are exocentric compounds. These are noticed in AN, NA, AV, and NV patterns, and are formed without the interfixes.
4. Endocentric compounds are seen in NN, AN, and NA formations. These are present with and without interfixes, though compounding with interfixes is a rare morphological process.
5. Copulative compounds are observed in NN, AA, and VV patterns. These are also formed without the interfixes.
6. Verbal compounds are present in N+ dV and A+ dV.
7. There is semantic opacity not only in exocentric compounds, but also in copulative compounds.

## Discussion and Conclusion

The findings of the present study show that Punjabi compound words are both similar to and different from those in Urdu and English. The right-hand head rule is present in the English language (Bloomfield, 1933), but the above-mentioned findings show that Punjabi compounds, like Urdu, have right-branching and left-branching rules, as they share many structural features with Urdu.

One important finding is the existence of Semantic divergence in some of the exocentric and copulative compounds in Punjabi and Urdu. On the other hand, despite these similarities, Punjabi and Urdu diverge at certain morphological levels. For example, interfixation, a productive compounding mechanism in Urdu (Islam, 2011), is hardly observed in Punjabi compounds, highlighting a more restricted morphological role of interfixes in Punjabi.

Additionally, endocentric compounds are frequent in the Urdu language (Islam, 2011). These are also very frequent in the English language (Bloomfield, 1933). However, exocentric compounds are frequent in the Punjabi language.

The study concludes that the formation of compounds is very productive in the Punjabi language. The most frequent types of compounds are exocentric compounds. Many of the morphological features of Punjabi match with those of Urdu, as both languages belong to the same language family. The study suggests that there is a need for detailed work on semantic opacity in Punjabi compounds. Further, a comprehensive study on hybrid compounds, including those formed through contact with English and other regional languages, is also required.

---

Note: <sup>1</sup>Adherents of Shia Islam are called *Shias*. They are the "followers" of Muhammad's son-in-law and, Ali, whom the Shia believe to be Muhammad's successor in the Caliphate. *Panjten* is the ideology of *Shias* who consider that *pandžtən* are the Five Purified Ones of the Prophet's family: Prophet Muhammad (SAW), Hazrat Ali, Hazrat Fatima, Hazrat Hassan and Hazrat Hussain.

---

## References

- Abbi, A. (1992). *Reduplication in South Asian languages: An areal, typological and historical study*. New Delhi: Allied Publishers.
- Baart, J. L. G. (2006). *An introduction to syntax and morphology: A guide for field linguists*. SIL International.
- Bauer, L. (2003). *Introducing linguistic morphology* (2nd ed.). Edinburgh University Press.
- Bauer, L. (2005). Les composés exocentriques de l'anglais. In D. Amiot (Ed.), *La composition dans une perspective typologique* (pp. 35–47). Arras: Artois Presses Université.
- Benczes, R. (2004). On the analysability of English exocentric compounds. *Jezekoslovija*, 5(1–2).
- Bhatia, T. K. (1993). *Punjabi: A Cognitive-descriptive Grammar*. Routledge.
- Bloomfield, L. (1933). *Language*. Chicago: University of Chicago Press.
- Butt, M., & King, T. H. (2006). The status of case. In V. Dayal & A. Mahajan (Eds.), *Clause structure in South Asian languages* (pp. 153–198). Springer.
- Haspelmath, M., & Sims, A. D. (2010). *Understanding morphology* (2nd ed.). London: Hodder Education.
- Islam, R. A. (2011). *The morphology of loanwords in Urdu: The Persian, Arabic and English strand* (PhD thesis). School of English Literature, Language & Linguistics, Newcastle University.
- Kachru, Y. (1980). *Aspects of Hindi Grammar*. Manohar.
- Lieber, R. (2009). *Introducing morphology*. Cambridge University Press.
- Malik, M. (2006). Punjabi machine transliteration. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL* (pp. 1137–1144).
- Masica, C. P. (1993). *The Indo-Aryan languages*. Cambridge University Press.
- Plag, I. (2002). *Word-formation in English*. Cambridge University Press.
- Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20(1–2), 261–290. <https://doi.org/10.1080/01690960444000098>
- Scalise, S., & Guevara, E. (2006). Exocentric compounding in a typological framework. *Lingue e Linguaggio*, 5(2), 185–206. <https://doi.org/10.1418/23143>
- Shackle, C. (2003). Punjabi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan Languages* (pp. 581–621). Routledge.
- Sharma, D. (1998). *Word Formation in Hindi: A Study of Compounding*. Delhi: Bahri Publications.
- Versteegh, K. (2001). Linguistic contacts between Arabic and other languages. *Arabica*, 48(4), 470–508. <https://doi.org/10.1163/157005801323163825>