

Linguistic Divergences Between Angika and Hindi

Jyoti Kumari^{#1} Praveen Gatla^{*2}

[#]Research Scholar, Department of Linguistics, Faculty of Arts, Banaras Hindu University, U.P, India

^{*}Assistant Professor, Department of Linguistics, Faculty of Arts, Banaras Hindu University, U.P, India

¹jyoti0502yadav@gmail.com, ²praveengatla@gmail.com

Abstract—Angika is an Eastern Indo-Aryan language spoken primarily in the eastern region of Bihar and adjoining areas of Jharkhand. Despite its long history and a substantial speaker community, it remains one of the least documented languages of the Indo-Aryan family. Owing to geographical proximity and historical classification, Angika has often been regarded as a regional variety of Hindi or Maithili. However, a closer examination of its grammatical structure reveals systematic linguistic characteristics that distinguish it from Standard Hindi. This paper presents a comparative analysis of the major linguistic differences between Angika and Hindi with particular emphasis on morphology and syntax. Using naturally occurring examples accompanied by phonetic transcription and interlinear glossing, the present study examines differences in verbal agreement, tense and aspect formation, pronominal systems, honorificity, case marking, pluralization, possessive constructions, conjunctive participles and emphatic particles. The analysis demonstrates that Angika exhibits little or no grammatical gender agreement in verbal morphology, employs an elaborate honorific-based agreement system and displays agglutinative morphological processes that differ substantially from the inflectional patterns found in Hindi. The study contributes to the linguistic documentation of an under-resourced Indo-Aryan language and provides a descriptive foundation for future comparative, typological and computational research. A systematic understanding of these linguistic differences is also expected to support the development of more accurate natural language processing tools and machine translation systems for Angika.

Keywords—Angika; Hindi; Linguistic divergence; Morphosyntax; verbal agreement; Agglutination.

I. Introduction

The Indo-Aryan language family comprises a large number of genetically related languages spoken throughout the Indian subcontinent. Although these languages share a common historical origin, they differ considerably in their phonological, morphological, syntactic and lexical characteristics. Comparative studies of closely related Indo-Aryan languages are essential for understanding language change, grammat-

ical variation and typological diversity. Such studies also provide valuable linguistic resources for developing computational tools for under-resourced languages.

Angika is an Eastern Indo-Aryan language belonging to the Bihari subgroup of the Indo-Aryan family. It is spoken predominantly in the historical Anga region of eastern Bihar, particularly in the districts of Bhagalpur, Banka, Katihar, Purnia, Madhepura, Saharsa, Munger and Sheikhpura. The language is also spoken in parts of Sahibganj and Godda districts of Jharkhand, the Malda district of West Bengal and neighbouring areas of Nepal (Regmi, 2017). According to the Mother Tongue Survey of India (2023)¹, Angika continues to serve as an important medium of everyday communication across these regions. It is written primarily in the Devanagari script, while historical sources indicate that Anga Lipi and later the Kaithi script were also used for writing the language. George A. Grierson classified Angika within the Bihari languages in the Linguistic Survey of India, where it was referred to as Chika Chiki. Since the 1961 Census of India, the language has been recorded under the name Angika. Although it has not yet been included in the Eighth Schedule of the Constitution of India, it possesses a well-established linguistic identity and continues to be used by a large speech community. Because of its geographical proximity to the Hindi-speaking region and its historical association with other Bihari languages, Angika has frequently been described as a variety of Hindi or Maithili. Such classifications, however, do not adequately reflect its grammatical structure. Earlier linguistic studies have shown that Angika differs substantially from Standard Hindi in several domains, including verbal agreement, case marking, pronominal systems, honorificity, nominal morphology and clause structure. Many of these features are shared with other Eastern Indo-Aryan languages, while others are distinctive to Angika itself.

The present paper presents a detailed comparison of the major linguistic divergences between Angika and Standard Hindi i.e. Khadiboli using naturally occurring examples drawn from Angika texts. The discussion focuses on verbal agreement, tense and aspect, pronominal systems, honorific distinctions, case marking, pluralization, possessive constructions, emphatic particles and conjunctive participles. By documenting these grammatical differences, the study contributes to the linguistic description of Angika and provides a useful reference for comparative Indo-Aryan linguistics, lan-

¹<https://language.census.gov.in/infographics/data/showMTSIdata>

guage documentation and computational research involving under-resourced Indian languages. These structural differences have important implications not only for descriptive and comparative linguistics but also for computational applications such as morphological analysis, syntactic parsing and machine translation.

The paper is divided into five sections. The section two is Review of Literature, section three deals with methodology, section four explains linguistics divergence between Angika and Hindi. Lastly we conclude the paper in section five.

II. Review of Literature

Although research on Angika remains limited compared with other Indo-Aryan languages, the existing literature has progressively enhanced our understanding of its linguistic structure and sociolinguistic status. Previous studies have examined its historical classification, language use, phonetic characteristics, morphosyntactic features and emerging computational resources, establishing Angika as a distinct and structurally rich Eastern Indo-Aryan language.

To establish the sociolinguistic baseline of the language, Regmi(2017) provides a comprehensive, data-driven profile of the Angika speech community. Utilizing participatory methods and wordlist comparisons, Regmi's report challenges previous classifications that labeled Angika as a "threatened" language. Instead, the study empirically demonstrates that Angika possesses a "vigorous" vitality, characterized by near-universal intergenerational transmission and deep community loyalty, firmly establishing it as an independent and actively spoken language in the region.

Kumar(2025) taking a diachronic sociolinguistic approach examines the variation in Angika copular verbs (chh versus chhik). The study reveals a significant generational shift: while older speakers tend to use the Maithili-associated chh, the younger generation predominantly utilizes chhik. Kumar argues that this morphological variation is not random but represents a conscious assertion of a distinct, modern Angika identity, further separating it from Maithili influence.

Chandrana and Pandey(2022) details the micro-syntactic structure of Angika. The authors highlight that Angika lacks a distinct category of definite articles, relying instead on demonstratives and numeral classifiers. Furthermore, the paper illustrates how Angika employs unique, agglutinative genitive markers and specific word orders for quantifiers and adjectives, distinguishing its nominal architecture from the inflectional patterns seen in Central Indo-Aryan languages.

Chandrana and Garg (2021) examine agreement patterns in Angika and Maithili, focusing on gender, number, person and honorificity. Their analysis shows that verbal agreement in Angika is primarily governed by person and honorific distinctions rather than grammatical gender and number. The study highlights important differences in the agreement systems of the two languages and contributes to the understanding of Angika verbal morphology. However, its scope is limited to agreement phenomena and does not examine other aspects of Angika grammar or its comparison with Standard Hindi.

Lahiri (2021) investigates the case systems of several Eastern Indo-Aryan languages from a cognitive linguistic perspective. The study demonstrates that case markers in Angika are multifunctional and exhibit semantic variation depending on their grammatical and discourse contexts. By analysing case marking within a broader cognitive framework, the book provides valuable insights into the functional behaviour of Angika case markers. However, the analysis is not intended as a direct comparison between Angika and Standard Hindi and therefore leaves broader morphosyntactic differences unexplored.

Lahiri(2012) provides a specific cognitive explanation for case syncretism in Angika. She explores why Angika uses the exact same postpositional marker (*se* or *sā*) for both the ablative and instrumental cases. Lahiri argues that Angika speakers conceptually map the "instrument" of an action as its foundational "source," providing a logical, semantic resolution to morphological overlaps that do not exist in Hindi or Bangla.

Syntactic distinctiveness in Angika extends into its negative constructions, as analyzed by Abbasi and Sulaiman(2020). The study highlights a sharp morphological contrast: while Urdu (and Hindi) relies on a context-dependent, tripartite system of negative particles (*mat*, *nahī*, *nā*), Angika utilizes a single, versatile negative particle (*ne*). The authors also demonstrate how Angika uses this single particle to efficiently handle complex pragmatic functions, including pre-suppositional negation and conveying positive implicit meanings.

Dorr(1994) proposed a formal framework for addressing lexical-semantic translation divergences using Lexical Conceptual Structure (LCS) as an interlingua. The study categorized divergences into seven classes and introduced Generalized Linking Routines (GLR) and Canonical Syntactic Realizations (CSR) to map language-independent structures to surface syntax. Implemented in the UNITRAN system, the parameterized approach effectively handled both individual and interacting divergences, demonstrating the value of LCS for robust and compositional machine translation.

Behera, Maurya, and Pandey(2016) analyzed linguistic divergences in English-to-Bhojpur machine translation using Dorr's Lexical Conceptual Structure (LCS) framework. They categorized syntactic and lexical-semantic divergences and showed that Generalized Linking Routines (GLR) and Canonical Syntactic Realizations (CSR) can effectively address such variations. The study highlighted the limitations of heuristic rules and emphasized the importance of universal structural frameworks for accurate MT of less-resourced languages.

III. Methodology

In the present study The Angika data analyzed were selected from a manually curated parallel corpus comprising 10,065 Angika-Hindi Parallel corpus. The Angika texts were collected from six printed literary works, digitized through OCR, manually corrected and translated into Hindi to create a high-quality Angika-Hindi parallel corpus. Examples representing

major grammatical phenomena were extracted from this corpus for qualitative comparative analysis.

IV. Linguistic Divergences Between Angika and Hindi

While Angika and Standard Hindi share historical roots within the Indo-Aryan language family, they exhibit profound morphosyntactic, phonological and lexical divergences. Hindi relies heavily on inflectional morphology with strict gender-number agreements, whereas Angika demonstrates agglutinative structure, gender-neutrality in verbs and a complex honorific-based pronominal system. Understanding these structural differences is critical for explaining the computational challenges faced by machine translation models.

The paper systematically examines seven distinct morphosyntactic phenomena that structurally differentiate the two languages. The analysis begins with (1) the complete lack of gender agreement in verbs, demonstrating how Angika maintains gender neutrality across the past, continuous, present, and future tenses, unlike Hindi's strict gender inflection. The section then explores divergences in (2) the pluralization system, highlighting Angika's use of uninflected nouns with agglutinative markers, and (3) case markers and postpositions, showing how Angika fuses reduced markers directly onto noun stems. Furthermore, the text details the differences in (4) emphatic particles and enclitics, (5) the morphophonological vowel mutations used to create conjunctive participles, and (6) the highly regular, agglutinative formation of possessive and reflexive pronouns. Finally, it concludes by outlining the contrasting syntactic formulas used for (7) the present continuous tense construction.

The following subsections detail the primary linguistic divergences between Angika and Hindi.

A. Lack of Gender Agreement in Verbs

In Standard Hindi, verbal morphology is heavily dependent on grammatical gender, requiring strict agreement between the verb and its subject (or object, in ergative constructions). For instance, a masculine subject takes a masculine verb form (rāmā d̥ʒa:t̪a: fi: [masculine], 'Ram goes'), while a feminine subject necessitates a feminine verb form (si:t̪a: d̥ʒa:t̪i: fi: [feminine], 'Sita goes'). In contrast, Angika exhibits complete gender neutrality in its verbal paradigm. The verb form remains entirely invariant regardless of whether the subject is masculine or feminine. This is evident in Angika where both rāmā d̥ʒa:j t̪ʰe: 'Ram goes' and si:t̪a: d̥ʒa:j t̪ʰe: 'Sita goes' utilize the identical verb phrase d̥ʒa:j t̪ʰe: 'goes'.

1) Past Tense

- (1) so:nəmə: ē:gənfi: sē: gərəd̥ʒi:
Sonma.[Fem] courtyard.EMPH from roaring
ke: bə:ləle:
having spoke

'Sonma roared from the courtyard itself and spoke.'

Hindi: so:nəmə: ā:gənə se: fii: gərəd̥ʒə kərə bə:li:

- (2) bə:esərə go:ssa: mē: lə:lə fi:t̪ē: bə:ləle:
Batesar.[Masc] anger in red becoming spoke
'Batesar spoke with anger.'

Hindi: bə:esərə gusse: mē: lə:lə fi:t̪ē: fii: bə:li:

A comparative analysis of Examples (1) and (2) clearly illustrates that typological divergence in the past tense. In Angika, the past tense verb bə:ləle: ('spoke') surfaces in an identical form for both the feminine subject (so:nəmə:) and the masculine subject (bə:esərə). The morphological structure of Angika prioritizes tense and person markers over gender, resulting in a unified, gender-agnostic verbal suffix.

Conversely, Standard Hindi enforces gender concord on the main verb. To express the exact same narrative event, Hindi alters its verbal roots, requiring the feminine past tense inflection -i: for the feminine subject Sonma (bə:li:) and the masculine past tense inflection -a: for the masculine subject Batesar (bə:lə:). This fundamental lack of gender inflection in Angika significantly distinguishes it from the gender-sensitive inflectional rules governing Central Indo-Aryan languages like Hindi.

2) Past Continuous Tense

- (3) d̥məjənt̪i: t̪ʰu:lfi: t̪ʰijā: bē:t̪həli: so:t̪fi:
Damayanti.[Fem] stove near seated thinking
rəfiəlo: t̪ʰe:le:
remaining was

'Damayanti was thinking sitting near the stove.'

Hindi: d̥məjənt̪i: t̪ʰu:lfi: ke: pə:sə bē:t̪h̄i: so:t̪fə rəfi: t̪ʰi:

- (4) əbə t̪ā:jə rəmuə: jəfi: so:t̪fi:
now until Ramua.[Masc] this.only thinking
rəfiəlo: t̪ʰe:le:
remaining was

'Ramua was thinking only this till now.'

Hindi: əbə t̪əkə rəmuə: jəfi: so:t̪fə rəfi: t̪ʰa:

Examples (3) and (4) demonstrate complete gender neutrality in the continuous aspect. In Standard Hindi, continuous and progressive states mandate active gender agreement on both the main verb's aspectual marker and the auxiliary verb. Thus, Hindi uses the feminine rəfi: t̪ʰi: for Damayanti and the masculine rəfi: t̪ʰa: for the masculine subject Ramua. In stark contrast, Angika relies on a regularized, gender-agnostic construction. Regardless of whether the subject is masculine or feminine, Angika maintains the invariant continuous marker and auxiliary sequence rəfiəlo: t̪ʰe:le:.

3) Present Tense

- (5) sukəri: k̪h̄e:t̪ō: mē: əsəkəlli:je: kāmā
Shukri.[Fem] fields in alone.EMPH work
kərə: t̪ʰe:
doing is

‘Shukri works alone in the fields.’

Hindi: sukəri: k^he:ṭō: mē: əke:li: fi: kɑ:mə
kərəṭi: fi:

- (6) visənɑ: ḍinə-ḍinə b^həri: k^he:ṭō: mē:
Vishana.[Masc] day-day full fields in
kɑ:mə kərə: t^hε:
work doing is

‘Vishana works all day in the fields.’

Hindi: visənɑ: ḍinə-ḍinə b^hərə k^he:ṭō: mē: kɑ:mə
kərəṭɑ: fi:

The divergence in present tense constructions, as seen in Examples (5) and (6), further highlights the fundamental morphological differences between Angika and Hindi. In Hindi, the present imperfective (habitual) suffix directly encodes the gender of the subject, mutating the verb into the feminine kərəṭi: fi: or the masculine kərəṭɑ: fi:. In Angika, this distinction is entirely absent. The present tense verb phrase kərə: t^hε: acts as a universal, gender-agnostic marker applied uniformly to both the feminine subject Shukri and the masculine subject Vishana.

4) Future Tense

- (7) ɑ:be: kɑ:səljɑ: ke: ḍʒinəgi: nifiɑ:lə
now Kaushalya.[Fem] GEN life fulfilled
fi:ʒə ɛ:ṭε:
happen will.go

‘Kaushalya’s life will be fulfilled now.’

Hindi: əbə kɑ:səljɑ: ki: ḍʒimḍəgi: nifiɑ:lə fi:
ḍʒɑ:ɛ:gi:

- (8) ḍəsə sɑ:lə: sē: kərdʒɑ: sē: ləḍlə:
ten years from debt from loaded
məḡələ ɑ:pəno: pəriɑ:rə səme:ṭə məri:
Mangal.[Masc] own family with die
ḍʒɛ:ṭε:
will.go

‘Loaded with debt for ten years, Mangal will die along with his family.’

Hindi: ḍəsə sɑ:lō: se: kərdʒə se: ləḍɑ: məḡələ
əpəne: pəriɑ:rə səme:ṭə mərə ḍʒɑ:ɛ:ga:

Finally, Examples (7) and (8) illustrate that gender inflection is equally absent in the Angika future tense. Standard Hindi maps future tense gender using a strict -gɑ: (masculine) / -gi: (feminine) / -ge: (plural) suffix paradigm, visibly alternating between the feminine ḍʒɑ:ɛ:gi: and the masculine ḍʒɑ:ɛ:ga:. Angika bypasses gender entirely, utilizing a person-based future paradigm. Consequently, it applies the identical third-person future suffixes, whether ɛ:ṭε: or ḍʒɛ:ṭε:, uniformly, irrespective of whether the subject is the feminine Kaushalya or the masculine Mangal.

B. Pluralization System

To pluralize nouns, Standard Hindi generally modifies the terminal vowel of the noun (e.g., lərəkɑ: → lərəkε:). Cru-

cially, in Hindi, when the subject noun is pluralized, the verb must also inflect to agree with the plural number. Conversely, Angika utilizes an agglutinative system where the noun itself remains unchanged, but a separate pluralization marker (such as səbə or si:ni:) is attached directly to the noun. Furthermore, Angika verbs do not inflect for number; the verb form remains completely invariant whether the subject is singular or plural.

- (9) t^hɔ:rɑ: k^he:li: rəfiələ: t^hε:
boy playing remaining is

‘The boy is playing.’

Hindi: lərəkɑ: k^he:lə rəfiɑ: fi:

- (10) t^hɔ:rɑ: si:ni: k^he:li: rəfiələ: t^hε:
boy.PL playing remaining is

‘The boys are playing.’

Hindi: lərəkε: k^he:lə rəfiε: fi:

Examples (9) and (10) clearly demonstrate this typological difference. In (10), the noun t^hɔ:rɑ: (boy) is not inflected to show plurality; instead, the agglutinative marker si:ni: is appended. More importantly, notice that the Angika verb phrase k^he:li: rəfiələ: t^hε: remains exactly the same in both the singular (9) and plural (10) constructions in Angika. In Hindi, however, both the noun and the entire verb phrase must undergo inflection, shifting from the singular lərəkɑ: ... rəfiɑ: fi: to the plural lərəkε: ... rəfiε: fi:.

- (11) bε:tɑ: səbə pərə bəfiute: b^həro:sɑ: t^hε: ɑ:kəɑ:
son.PL on much trust is him

‘He has a lot of trust in all his sons.’

- (12) mənε:mənə pəṭuɑ: bəkəriɑ:sini: kē: bəcɑ:bɑ:i
in-mind Patua goat.PL ACC save.INF
ke: ḍʒugutə k^ho:ḍʒɑ:i lɑ:gələ:
GEN trick searching started

‘In his mind, Patua started looking for a trick to save the goats.’

Examples (11) and (12) further illustrate the productivity of these plural markers. In (11), the marker səbə attaches to bε:tɑ: (son) to mean “sons.” In (12), the marker sini: attaches to bəkəriɑ: (goat) to indicate plurality, yielding bəkəriɑ:sini: (“goats”). In all cases, the root noun remains structurally unaltered, confirming Angika’s agglutinative approach to pluralization.

C. Case Markers and Postpositions

In Angika the functional case markers (postpositions) differ from those in Standard Hindi. While Hindi generally employs stand alone postpositional words, Angika markers frequently exhibit agglutination, attaching directly to the noun stem, or they fundamentally alter the phonetic shape of the root word.

- (13) mǝgǝlə kǝ fǝmɛ:sǝ:lɛ:li: sǝdʒǝ: d̪ɛ: kɛ:
Mangal ACC always for punishment giving of
sǝbǝsɛ: ǝtʃtʃʰǝ: t̪ɛ:ri:kǝ:..
best method
'The best way to punish Mangal forever...'
Hindi: mǝgǝlə kɔ: fǝmɛ:sǝ: kɛ: liɛ: sǝdʒǝ: d̪ɛ:ne:
kǝ: sǝbǝsɛ: ǝtʃtʃʰǝ: t̪ɛ:ri:kǝ:...

In example (13), demonstrating the accusative/dative case, the Standard Hindi standalone marker kɔ: is entirely replaced by reduced Angika marker kǝ.

- (14) bǝ:bʊ: rɔ: niqʰǝnǝ d̪u: mǝfi:nǝ: pɛ:nfiɛ:
Father GEN demise two months before
fiɔ:jǝ ge:lɔ: rǝfiɛ:
happen went was
'Father's demise had happened two months ago.'
Hindi: bǝ:bʊ: dʒi: kǝ: niqʰǝnǝ d̪o: mǝfi:ne:
pǝfiǝle: fiɔ: gǝjǝ: tʰǝ:

In example (14), demonstrating the genitive case, Angika utilizes the distinct marker rɔ: to indicate possession or relationship. This completely replaces the inflected Hindi genitive paradigm of kǝ:/ki:/ke:.

- (15) sɔ:nǝmǝ: ɛ:gǝnfɛ: sɛ: gǝrǝdʒi: kɛ: bɔ:lǝlɛ:..
Sonma courtyard from roaring having spoke
'Sonma roared from the courtyard itself and spoke...'
Hindi: sɔ:nǝmǝ: ǎ:gǝnǝ se: fi: gǝrǝdʒǝ kǝrǝ
bɔ:li:..

Example (15) highlights the ablative/instrumental case. The Standard Hindi marker se: is rendered in Angika as sɛ:, heavily relying on nasalization.

- (16) bǝtɛ:sǝrɛ: kǝfiǝlǝkɛ: bʰǝgǝvǝ:nɛ: dʒǝmi: kǝ
Batesar.ERG said God.ERG heavy ACC
bǝ:risǝ d̪ɛ:lɛ: tʃʰɛ:lɛ:
rain gave was
'Batesar said God had given heavy rain to the land.'
Hindi: bǝtɛ:sǝrǝ ne: kǝfiǝ: bʰǝgǝvǝ:nǝ ne: dʒǝmǝ
ke: bǝ:risǝ d̪i: tʰi:

Example (16) highlights a structural divergence in the ergative case (used for past tense transitive subjects). In Hindi, the ergative case is marked by the separate, standalone postposition ne: (bǝtɛ:sǝrǝ ne: and bʰǝgǝvǝ:nǝ ne:). In contrast, the Angika ergative case is agglutinative; the marker fuses directly onto the subject noun as a nasalized -ɛ: suffix, yielding single tokens like bǝtɛ:sǝrɛ: and bʰǝgǝvǝ:nɛ:.

- (17) t̪ɔ:ro: gfiǝro: mɛ: rɔ:dʒɛ: jǝfi: rǝ dʒʰǝgǝrǝ:
Your house.LOC in daily this like fight
fiɔ:jǝ tʃʰɔ:
happens is
'Fights happen like this daily in your house.'
Hindi: t̪umfiǝ:re: gfiǝrǝ mɛ: rɔ:dʒǝ isi: t̪ǝrǝfiǝ
dʒʰǝgǝrǝ: fiɔ:tǝ: fiɛ:

Example (17) demonstrates the locative case and its associated phonological rounding. In Standard Hindi, the locative marker mɛ: (in) simply follows the uninflected noun gfiǝrǝ (house). In Angika, the application of a locative marker (mɛ: or pǝrǝ) frequently forces a phonological shift in the root noun, altering its terminal vowel to an -o: ending, thereby changing gfiǝrǝ to gfiǝro: (gfiǝro: mɛ:).

D. Emphatic Particles and Enclitics

Standard Hindi utilizes separate, standalone lexical words to indicate emphasis, such as fi: (only/exactly, indicating exclusivity) and bʰi: (also/even, indicating inclusivity). In contrast, Angika relies heavily on morphological enclitics. Rather than using separate words, Angika merges these emphatic markers directly into the preceding noun, pronoun, or adverb by altering its terminal vowel. Typically, exclusivity is marked by a shift to -e/je, while inclusivity is marked by a shift to -ɔ:.

- (18) sukǝri: kʰɛ:t̪ɔ: mɛ: ǝsǝkǝlli:je: kǝ:mǝ kǝrɛ:
Shukri fields in alone.EMPH work doing
tʃʰɛ:
is
'Shukri works in the fields completely alone.'
Hindi: sukǝri: kʰɛ:t̪ɔ: mɛ: ǝkɛ:li: fi: kǝ:mǝ
kǝrǝti: fiɛ:

In Example (18), the Hindi construction uses the separate particle fi: to emphasize exclusivity (ǝkɛ:li: fi:, 'alone only'). In Angika, this emphasis is fused directly into a single word, ǝsǝkǝlli:je:, where the -je: suffix intrinsically carries the exclusive emphatic meaning without requiring an additional lexical particle.

- (19) suni:t̪ɔ: dʒɛ:t̪ɛ:
Sunita.EMPH will.go
'Sunita will also go.'
Hindi: suni:t̪ǝ: bʰi: dʒǝt̪ɛ:gi:

Example (19) illustrates inclusive emphasis. Standard Hindi uses the standalone inclusive marker bʰi: (suni:t̪ǝ: bʰi:). Angika dispenses with the separate marker entirely. Instead, the terminal vowel of the proper noun mutates; the base name suni:t̪ǝ: is altered to end in the -ɔ: suffix, resulting in the fused emphatic form suni:t̪ɔ:. This agglutinative enclitic system is a defining phonological and morphological characteristic of Angika.

E. Conjunctive Participles

Conjunctive participles are used to link sequential actions (e.g., "having done X, Y happened"). Standard Hindi forms conjunctive participles by appending the standalone invariant markers kǝrǝ or kǝrǝke: directly to the uninflected verb root (e.g., dʒǝ:kǝrǝ, kʰǝ:kǝrǝ). Angika utilizes a highly distinct agglutinative and morphophonological process to achieve this. It appends the reduced particles ke: or kǝ (often written

with an Avagraha as *kā*) and crucially, it mutates the terminal vowel of the verb root, typically shifting it to an *-i* or *-y* ending.

- (20) *rəmuɑ: pəmo: e:kə kəna: mē: p^hē:ki: kə*
Ramua stick one corner in throwing having
ut^htʃo: a:va:dʒo: mē: bəte:sərə ke: sə^hijare:
high voice in Batesar ACC explaining
tʃ^hε:
is
‘Ramua, having thrown the stick in a corner, is explaining to Batesar in a loud voice...’
Hindi: *rəmuɑ: dānda: e:kə kəne: mē: p^hē:kə kərə*
ū:tʃi: a:va:dʒə mē: bəte:sərə ko: sə^hari: d̪e:ʈa:
fiε:

In Example (20), the Hindi conjunctive participle *p^hē:kə kərə* (having thrown) translates to Angika as *p^hē:ki: kə*. The base verb *p^hē:kə* takes an *i* vowel suffix (*p^hē:ki:*) and the Hindi *kərə* is replaced by the uniquely reduced Angika particle *kə*.

- (21) *təb^hiε: bəte:sərə ləpəki: ke: o:kəro: fiɑ:t^hə*
Then.only Batesar leaping having his hand
pəkə:ɪ: lε: tʃ^hε:
grabbing takes is
‘Right then Batesar, having leaped, grabs his hand...’
Hindi: *təb^hi: bəte:sərə ləpəkə kərə usəka: fiɑ:t^hə*
pəkəə le:ʈa: fiε:

Example (21) further demonstrates this vowel mutation. The Hindi verb root *ləpəkə* (leap/grab) remains uninflected before *kərə*. In Angika, the root mutates to *ləpəki:*, followed by the particle *ke:*.

- (22) *biʈtɑ: b^həri: g^ho:g^ho: sē: dʒ^hā:ki: kə*
Hand-span full veil from peeking having
vē: d̪e:k^hənē: tʃ^hε:le: dʒi:təna: kē:
she seeing was Jeetna ACC
‘She had seen Jeetna by peeking through a hand-span long veil.’
Hindi: *biʈtɑ: b^hərə g^hū:g^hətə se: dʒ^hā:kə kərə*
usəne: d̪e:k^hɑ: t^hɑ: dʒi:təna: ko:

In Example (22), the Hindi *dʒ^hā:kə kərə* (having peeked) maps to the Angika *dʒ^hā:ki: kə*. The use of *kə* represents a highly authentic phonetic feature of rural Angika, characterized by an exceptionally light, almost swallowed *k* articulation compared to the hard Hindi *kar*.

- (23) *so:nəma: ke: səmədʒ^hɑ:jə-budʒ^hɑ:jə kə*
Sonma ACC explaining-consoling having
dʒəbe: u: t^həkki: ge:le:
when he tired went
‘When he got tired having consoled and explained to Sonma...’
Hindi: *so:nəma: ko: səmədʒ^hɑ:-budʒ^hɑ: kərə*
dʒəbə vəfiə t^həkə gəja:

Example (23) illustrates a paired conjunctive participle. While Hindi uses *səmədʒ^hɑ:-budʒ^hɑ: kərə*, Angika applies the *-y* mutation to verbs ending in *a* vowels, yielding *səmədʒ^hɑ:jə-budʒ^hɑ:jə*, followed by the *kə* particle.

- (24) *nε nε kərtēñ sərbəʈ e:kε: sa:ñsoñ sē:*
No no doing.while syrup one.EMPH breath in
sədsəda:j kə pi:vi: ge:liε...
slurping having drink went
‘Saying no-no, he drank the syrup in one single breath, slurping it down...’
Hindi: *nɑ: nɑ: kərtε: fue: fərbəʈ e:k fi: sa:ñs*
mē: sədsəda: kərə pi: gəe:...

Example (24) further corroborates the *-y* mutation rule. When the verb root terminates in an *-aa* vowel (such as the onomatopoeic *sədsəda:*), Angika avoids the *-i* suffix. Instead, it appends a *-y* suffix to create the conjunctive form (*sədsəda: + j = sədsəda:j*), resulting in *sədsəda:j kə*.

- (25) *vē: bina: kut^hə so:tʃələ: pəfi:lē:*
He without anything thinking.having first
dʒile:bi: ke: a:pəno: mūfiə: mē: d̪a:ləna: suru:
jilebi ACC own mouth in putting start
kəraləkē:
did
‘Without thinking anything, he first put the jilebi in his mouth...’
Hindi: *usəne: bina: kut^hə so:tʃe: pəfi:le:*
dʒile:bi: ko: əpəne: mūfiə mē: d̪a:ləna: suru:
kija:
- (26) *gfiəro: mē: bma: kutʃfi bəʈε:nē:*
House.LOC in without anything telling.having
o:krəñ sa:tʃi tʃəli: d̪e:liε.
him with walk gave
‘...without telling anything in the house, [he] left with him.’
Hindi: *gfiər mē: bma: kutʃfi bəʈa:ε: uske: sa:tʃi*
tʃəl d̪ie:

Examples (25) and (26) demonstrate a highly distinct, zero-particle conjunctive structure. In these constructions, Angika circumvents the use of *ke:* or *kə* entirely. Instead, a conjunctive or participial meaning (“having thought” or “having told”) is generated strictly through agglutinative suffixes. The verb *so:tʃ* (think) takes the suffix *-lē:* (*so:tʃələ:*) and *bəʈa:* (tell) takes the suffix *-nē:* (*bəʈε:nē:*).

F. Possessive and Reflexive Pronouns

In Standard Hindi, the root forms of first and second-person pronouns undergo a complete, irregular morphological transformation when inflected for the possessive genitive case (e.g., *mē:* [I] → *me:ra:* [my]; *tū:* [you] → *te:ra:* [your]). Angika, however, utilizes a highly regular, agglutinative approach. It retains the original pronominal root and directly attaches the genitive marker (such as *ro:* or *ke:*), preserving the base morphology of the pronoun without irregular stem changes.

- (27) fəməro: gudʒa:ra: je: məɾd̪ə ke: sɑ:tʃiē: nɛ
My survival this man GEN with not
fi:ɾɛ:
will.be
'My survival will not happen with this man.'
Hindi: mə:ra: guza:ra: isə məɾd̪ə ke: sɑ:tʃiə
nəfi:n fi:ga:

In Example (27), the Angika possessive fəməro: (my) is formed simply by attaching the genitive particle ro: directly to the first-person root fəmə (I/we). This contrasts sharply with the irregular Hindi shift to mə:ra:.

- (28) ʈo:ro: ɡʃəro: mē: ro:dʒe: jəfi: rə dʒʃəgəɾa:
Your house.LOC in daily this like fight
fi:ɾə ʈʃfi:
happens is
'Fights happen like this daily in your house.'
Hindi: ʈumʃa:re: ɡʃəɾə mē: ro:zə isi: ʈəɾəʃə
dʒʃəgəɾa: fi:ʈa: fi:

Similarly, in Example (28), the second-person possessive ʈo:ro: (your) is formed by agglutinating the genitive ro: to the base root ʈo:.

- (29) ko:jə o:kəro: bɑ:ʈə nɛ uθʃa:bɛ: ʈʃfiē:
Nobody his/her matter not overturns is
'Nobody overrides his/her word/decision.'
Hindi: ko:ɪ: usəki: bɑ:ʈə nəfi:n ʈa:ləʈa: fi:

Example (29) demonstrates the third-person possessive o:kəro: (his/her). Notice that due to Angika's inherent gender neutrality, o:kəro: applies universally to both masculine and feminine subjects, whereas Hindi actively shifts its possessive pronouns between usəka:, usəki:, and usəke: depending entirely on the grammatical gender of the possessed object.

G. Present Continuous Tense Construction

The syntactic formula for constructing the present continuous (progressive) tense diverges significantly between the two languages. Standard Hindi utilizes the structure: [*Verb Root* + rəfiɑ:/rəfiɪ/rəfiē: + fiē:/fiē:]. Angika, however, utilizes a mutated verbal root structure paired with an invariant continuous marker: [*Verb Root* + i: + rəfiəlo: + ʈʃfiē:].

Crucially, in Angika, the continuous aspect marker (rəfiəlo:) is entirely static and gender-neutral, unlike the highly inflected Hindi markers (rəfiɑ:/rəfiɪ/rəfiē:).

- (30) rəmuɑ: nɑ:dʒɑ:jəɾdʒə bə:li: rəfiəlo: ʈʃfiē:
Ramua. illegitimate speaking remaining is.[NH]
'Ramua is speaking nonsense.'
Hindi: rəmuɑ: nɑ:dʒɑ:jəzə bə:lə rəfiɑ: fi:

In Example (30), the Angika continuous tense mutates the base verb root bə:lə (speak) into bə:li: by appending an -i suffix. The continuous state is denoted by rəfiəlo:, followed by the non-honorific auxiliary ʈʃfiē:.

- (31) sukəri: jəfiə: səbə so:tʃi: rəfiəlo: ʈʃfiē:lɛ:
Shukri. this also. all thinking remaining was
'Shukri was thinking all this too.'
Hindi: sukəri: jəfiə səbə bʰi: so:tʃə rəfi: ʈʃfi:

Example (31) illustrates this phenomenon seamlessly in the past continuous tense. The verb root so:tʃə (think) mutates to so:tʃi:. Despite the narrative subject being strictly feminine, Angika utilizes the gender-neutral continuous marker rəfiəlo: instead of a feminine equivalent. In contrast, the Hindi output explicitly enforces feminine gender agreement via both the continuous marker rəfi: and the auxiliary ʈʃfi:.

V. Conclusion

This study presents a systematic comparison of the grammatical structures of Angika and Standard Hindi, demonstrating that they differ across several phonological, morphological and syntactic domains. The analysis shows that, unlike Standard Hindi, Angika exhibits little or no grammatical gender agreement in verbal morphology, employs an extensive honorific-based pronominal system and displays agglutinative tendencies in case marking, plural formation and conjunctive participle constructions. The findings contribute to the linguistic documentation of Angika by providing a unified description of its major grammatical features in comparison with Standard Hindi. Such comparative analyses are important for understanding linguistic diversity within the Indo-Aryan language family and for improving the documentation of under-resourced languages. The study also provides a descriptive foundation for future research in comparative linguistics, language documentation and computational linguistics, particularly in the development of language-specific resources and Natural Language Processing tools for Angika.

References

- Abbasi, B. A., Sulaiman, Z. M., Kashyap, R. (2020). An inquiry into negation(s) in Urdu and Angika: A comparative study. *Purakala*, 31(18), 912–924.
- Asad, M. (2022). Acoustic space and formant patterns of vowels in Maithili and Angika: A comparative and descriptive study. *Interdisciplinary Journal of Linguistics*, 15, 127–138.
- Chandana, A. K., Garg, N. (2021). Number and gender agreement: A comparative study of Angika and Maithili. *Anukriti*, 11(6), 47–54.
- Chandana, A. K., Pandey, A. (2022). Nominal phrases in Angika, Maithili, Awadhi, and Bhojpuri: A comparative study. *International Journal of Novel Research and Development*, 7(7), 458–474.
- Kumar, A., Deep, M. K. (2025). Comparative analysis of morphological approaches for low-resource Indo-Aryan languages: Case studies in Angika,

- Maithili, and Hindi. *International Journal of Multidisciplinary Research and Explorer*, 5(3), 104–117. <https://doi.org/10.70454/IJMRE.2025.05037>
6. Kumar, R. (2025). Linguistic variation and stability in Angika. *International Journal of Creative Research Thoughts*, 13(1), a547–a556.
 7. Kumar, S., Jyothi, P., Bhattacharyya, P. (2024). Part-of-speech tagging for extremely low-resource Indian languages. In *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 14422–14431). Association for Computational Linguistics.
 8. Lahiri, B. (2012). Ablative and instrumental markers in some of the EIA languages. In *Proceedings of the 5th Students' Conference of Linguistics in India (SCONLI-5)* (pp. 43–51).
 9. Lahiri, B. (2021). *The case system of Eastern Indo-Aryan languages: A typological overview*. Routledge.
 10. Regmi, A. (2017). *A sociolinguistic survey of Angika* [Report submitted to Linguistic Survey of Nepal]. Central Department of Linguistics, Tribhuvan University.
 11. Grierson, G. A. (1903). *Indo-Aryan Family: Eastern Group: Specimens of the Bihārī and Oṛiyā languages* (Linguistic Survey of India, Vol. V, Pt. II). Calcutta: Office of the Superintendent of Government Printing, India.
 12. Jeffers, R. J. (1976). *The Position of the Bihari Dialects in Indo-Aryan*. *Indo-Iranian Journal*, 18(3/4), 215–224.
 13. Verma, S. (1985). *The Structure of Magahi verbs*. New Delhi: Bahri Publications.
 14. Yadav, R. (1996). *A Reference Grammar of Maithili*. Berlin/New York: Mouton de Gruyter.
 15. Office of the Registrar General Census Commissioner, India. (2023). *Mother Tongue Survey of India*. Language Division, Ministry of Home Affairs, Government of India.
 16. Office of the Registrar General Census Commissioner, India. (2023). *Mother Tongue Survey of India: Angika*. Language Division, Ministry of Home Affairs, Government of India.
 17. Dorr, B. J. (1994). Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20(4), 597–633.
 18. Sinha, K., Mahesh, R., Thakur, A. (2005). Translation divergence in English-Hindi MT. In *Proceedings of the 10th EAMT conference: Practical applications of machine translation*.
 19. Behera, P., Mourya, N., Pandey, V. (2016). Dealing with linguistic divergences in English-Bhojpuri machine translation. In *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)* (pp. 103–113).