https://doi.org/10.69760/jales.2025002011

Comparative Analysis of the Verb "to be" in Seven Indo-European Languages

https://orcid.org/0009-0007-0247-476X
[1]Hasan Alisoy
Nakhchivan State University, Azerbaijan

Abstract

This study investigates the morphological, syntactic, and semantic features of the verb "to be" in seven Indo-European languages: English, German, Russian, Persian, Sanskrit, Latin, and Lithuanian. As one of the most fundamental verbs across languages, "to be" serves critical grammatical functions such as copula, existential marker, and auxiliary. Using a comparative linguistic framework, the research examines the forms of "to be" in present, past, and future tenses, identifies shared roots and divergences, and highlights regular and irregular paradigms. Tables are provided to demonstrate inflectional patterns and the role of person, number, and tense across these languages. The diachronic analysis traces the evolution of the proto-Indo-European root *h₁es- and its reflexes, revealing both inherited structures and language-specific innovations. The study also contextualizes the verb's syntactic behavior in copular and existential constructions, providing examples in each language. The results illustrate both continuity and transformation within Indo-European verbal systems and offer insights into the historical development of grammatical categories. This paper contributes to comparative and historical linguistics, with implications for language teaching, typology, and philological studies.

Keywords

Indo-European languages, Verb "to be", Comparative linguistics, Copula, Morphological analysis, Historical linguistics

Introduction

The verb "to be" is one of the most fundamental and irregular verbs across human languages. In Indo-European linguistics, it holds special significance because a form of this copular verb is present in all Indo-European languages. As a copula, "to be" serves to link subjects with predicate complements (e.g., "she is a teacher") or to express existence (e.g., "there is a problem"). Its high frequency and essential grammatical function have caused "to be" to undergo extensive irregular developments in the Indo-European family. In many Indo-European languages, "to be" exhibits suppletion—different tense forms derive from historically distinct roots—making it an important case study for comparative and historical linguistics. Studying this verb's paradigms can illuminate sound changes, morphological evolution, and the reconstruction of Proto-Indo-European (PIE) forms.

Reconstructing PIE reveals that there was not a single verb root for "be"; instead, multiple PIE roots coexisted for the concept of being. The primary roots include *h₁es-* (to be), *bḱuH-* (to become, grow), and others like *h₂wes-* (to dwell, live) and possibly *h₁er-* (to move, arise). The daughter languages often formed

[1] Alisoy, H. Lecturer in English, Nakhchivan State University. Email: alisoyhasan@nedu.edu.az. ORCID: https://orcid.org/0009-0007-0247-476X.

suppletive paradigms by combining these roots, each used in different tenses or persons. This makes "to be" a mosaic of ancient linguistic heritage—for example, English "am/is" (< PIE *h₁es-*), "was/were" (< PIE *h₂wes-*), and "be" (< PIE *bʰuH-*) are etymologically unrelated internally, yet together function as the single verb "to be." Because these forms are cognate with forms in Sanskrit, Latin, and other Indo-European tongues, the verb "to be" has been a cornerstone for Indo-European comparative studies (Watkins, 2000; Ringe, 2006). By examining it across multiple languages, we can trace the phonological and morphological changes from the proto-language and gain insight into how each branch of Indo-European developed its verbal system.

In this study, we present a comparative analysis of the verb "to be" in seven Indo-European languages: English, German, Russian, Persian, Sanskrit, Latin, and Lithuanian. These languages were selected to represent major branches of the family (Germanic, Slavic, Indo-Iranian, Italic, and Baltic, respectively) and to include both ancient languages (Latin, Sanskrit) and modern languages. Each language's paradigm for "to be" is examined in the present, past, and future tenses (where applicable). By comparing these side-by-side, we highlight the deep historical connections as well as divergences caused by thousands of years of language change. The analysis sheds light on how a single conceptual verb can evolve into a complex patchwork of forms, reflecting both shared ancestry and individual language histories.

Methodology

This research employs the comparative-historical method of linguistics, using a descriptive and contrastive framework to analyze verb paradigms. We began by collecting the conjugation forms of the verb "to be" in the present, past, and future tenses for each of the seven selected languages. Authoritative sources were used for each language's morphology (e.g., standard grammar references and historical linguistics publications). For ancient languages like Sanskrit and Latin, classical grammar documents were consulted to obtain paradigms in transliterated form. For modern languages (English, German, Russian, Persian, Lithuanian), standard grammatical descriptions provided the forms, with transliteration applied for the non-Latin scripts (Cyrillic for Russian, Perso-Arabic for Persian) for ease of comparison. Each paradigm was double-checked against historical linguistics sources to identify the Proto-Indo-European root corresponding to each form.

The comparative approach involves aligning equivalent grammatical categories across languages. We therefore structured the data by person (first, second, third) and number (singular, plural) for each tense. This allows a side-by-side paradigm comparison in tabular form, so that, for example, all first-person singular forms ("I am/was/will be") can be viewed together. Such alignment makes it easier to spot cognates—forms descended from the same proto-form—and to observe where languages have innovated or diverged (Beekes, 2011). Given that some languages do not have a synthetic future tense or drop certain copula forms (e.g., Russian has no present-tense "to be" in most contexts), we note these as special cases rather than leaving blank slots. We focused on one primary past tense per language for comparison (usually the simple past or imperfect, analogous to English "was"), even though some languages have multiple past tenses; this keeps the comparison consistent. For the future tense, we included the typical way to express future being in each language, whether as an inflected form or a periphrastic construction.

In analyzing results, we applied principles of historical linguistics to explain correspondences and differences. Sound change laws (like Grimm's Law for Germanic or rhotacism in Latin) are referenced to connect forms to their PIE etyma. We also refer to Indo-European reconstruction (proto-forms marked with

an asterisk) from standard etymological sources (e.g., Watkins, 2000; Pokorny, 2007) to discuss how each attested form can be derived from ancestral roots. By combining synchronic description (the paradigms as they exist in each language now or in historical record) with diachronic explanation (how those forms came to be), our methodology highlights both the descriptive facts of each language and the historical development linking them. All examples and claims are supported with references to established linguistic research and comparative data.

Results

Present Tense Paradigms

Table 1 below displays the present tense of the verb "to be" in English, German, Russian, Persian, Sanskrit, Latin, and Lithuanian. For consistency, forms are given in each language's usual transcription, with transliteration provided for Sanskrit and Persian. The English and German forms are from Modern Standard varieties; Sanskrit forms represent the active voice of Classical Sanskrit; Latin forms reflect Classical Latin usage; and the Russian and Persian forms are from contemporary standard language.

The Persian forms included are the "long copula" variants (those that contain *hast*), which are typically used for formal or emphatic expression. Although Persian often employs short enclitic forms of the copula, these are noted separately in the analysis for comparison. Russian presents a unique case: in present-tense usage, the verb "to be" is typically omitted, leaving no explicit conjugated form. This omission is indicated by a dash (—) in the table. The third-person singular form *есть* (*est'*, "is") exists but appears only in archaic or emphatic contexts.

The paradigms in Table 1 are organized by person (first, second, third) and number (singular, plural), enabling a clear visual comparison of equivalent grammatical categories across the seven languages.

Table 1: Present Tense Forms of "to be" (1st, 2nd, 3rd person; singular and plural)

Person	*English*	*German*	*Russian*	*Persian*	*Sanskrit*	*Latin*	*Lithuanian*
1sg "I am"	am	bin	— (Ø)	hastam	ásmi	sum	esu
2sg "you are"	Are	bist	— (Ø)	hastí	ási	es	esi
3sg "he/she is"	is	ist	(есть) est'	hast (ast)	ásti	est	yra
1pl "we are"	are	sind	— (Ø)	hastím	smás	sumus	esame
2pl "you are"	are	seid	— (Ø)	hastíd	sthá	estis	esate
3pl "they are"	are	sind	— (Ø)	hastánd	sánti	sunt	yra

In English, the present tense of "to be" has three forms: *am* (1st person singular), *is* (3rd singular), and *are* (2nd singular and all plurals). All persons in the plural use *are*, reflecting the merger of original second-person plural "ye are" and third-person plural forms in Modern English. German similarly has an irregular present: *bin*, *bist*, *ist* for 1st–3rd singular, and *sind*, *seid*, *sind* for 1st–3rd plural. German maintains a distinct second-person singular *bist* ("thou art") and second-person plural *seid* ("you all are"), whereas English uses *are* for both. The forms *bin* and *bist* begin with *b*-, unlike *ist*, hinting at their different origin (discussed in the Discussion section). Both English and German exhibit suppletion in the present paradigm (multiple stems in one tense).

Russian stands out by typically lacking an overt present tense form of *быть* (*byt'*, "to be") in the indicative. In modern Russian, one says "I student" (literally "I student") to mean "I am a student," with no verb in

present-tense nominal sentences. The form *ecmь* (*est'*, cognate to *is*) exists and literally means "is/exists," but it is used mostly to indicate existence or emphasis (e.g., "God is") or in set phrases. For completeness, we list *ecmь* under 3rd singular in Table 1 in parentheses, but it is not commonly used as a copula for all persons in modern Russian. Essentially, Russian has a zero-copula in the present tense: the copular meaning is understood without a verb. This is a later syntactic development in Slavic; Old Church Slavonic had present forms like *jesmĭ* "I am," but these fell out of use in East Slavic.

In Persian, the modern present copula is usually attached as personal suffixes to predicates. For example, *man azādam* ("I am free") uses *-am* to mean "am." The full forms with *hast* are shown in Table 1 for clarity: *hastam, hasti, … hastand*, corresponding to "I am, you are, … they are." In everyday use, Persian often drops the initial *h-* (especially after vowels) or even omits the copula in certain registers, but the literary form *hast* (3rd sg.) or its reduced form *ast* is retained for "is." Notably, Persian *ast* (است) is a direct cognate of Sanskrit *asti* and Latin *est*, all reflecting the Proto-Indo-European 3rd singular form *h₁ést(i)*. The Persian 1st singular *am* (ـم) and 2nd *-i* (ـی) are actually cliticized pronoun forms historically derived from the same root *h₁es-* (through Old Persian *ahmi* > Middle Persian *om* > *-am*). Thus, although Persian's method of conjugation (using enclitic pronouns) differs from the fusional endings of Sanskrit or Latin, the etymology of its present copula is equally ancient. Persian, like a few other Indo-European languages (e.g., Spanish, Irish), even has multiple forms of "to be" for different contexts—for instance, Persian uses *ast/hast* for essential being and *mi-tavān-ad būd* for potential being ("can be"), but that is beyond our current scope.

Sanskrit preserves the full inherited paradigm of the verb √*as* ("to be") in the present tense with minimal change from Proto-Indo-European. The forms are *ásmi, ási, ásti, smás, sthá, sánti* for 1st–3rd singular and plural. These correspond exactly to the patterns reconstructed for PIE: e.g., *h₁ésmi* > *asmi* "I am," *h₁ésti* > *asti* "he is," and the plural *sánti* "they are" from PIE *h₁sénti*. Sanskrit even retains the distinctive 2nd person plural form *sthá*, which comes from the PIE root *steh₂-* ("to stand") appended to the *as-* stem (a relic of an old injunctive form). However, by the Classical Sanskrit period, *sthá* was simply understood as part of the paradigm of *as-*. Lithuanian, interestingly, has a very similar form *(jū)s esate* for 2pl "you are," which is cognate to Sanskrit *sthá* after regular sound changes (the *sth* > *s* in Baltic).

Latin present forms are *sum, es, est, sumus, estis, sunt*, showing some innovations. Latin *sum* ("I am") and *sunt* ("they are") appear to derive from the PIE root *h₁es-* but underwent early Latin-specific sound changes. Some scholars interpret Latin *sum* as coming from a reduplicated or o-grade form of the root (earlier *esom* < *h₁ésmi*), while others think a different root might be involved; however, it is widely accepted that Latin *est, estis, sunt* align with the standard PIE *es-* forms (compare *est* = Skt. *asti*, *sunt* = archaic Latin *sont* < *h₁sónti*). Latin *es* ("you are") is a direct continuation of PIE *h₁ési*. One anomaly is that Latin uses *-nt* for the 3rd plural *sunt*, whereas Sanskrit has *-nti* (an older ending). Latin dropped the final *-i*, a regular sound change in the Italic branch. Overall, Latin's present conjugation of *esse* closely parallels the Sanskrit one (minus the 1sg form *sum*, which is irregular).

Lithuanian present forms of *būti* ("to be") combine two different stems: one from *es-* and one unique form. The 1st and 2nd persons are *esu* ("I am") and *esi* ("you are"), transparently cognate with Sanskrit *asmi, asi* (the *-mi* ending became *-u* in Baltic languages). However, the 3rd person *yra* ("is/are") is irregular. It does not obviously descend from *esti*; historical linguists suggest it comes from a Proto-Baltic form *yes-* or a fusion of *es-* with a demonstrative. Old Lithuanian texts also had *esti* or *yesti* for "is" in earlier times, and even an obsolete *esmi* form for "I am." In modern usage, *yra* serves for both singular and plural 3rd person ("he/she is" and "they are"), thus *jie yra* = "they are." This is a Baltic innovation; Lithuanian has effectively

leveled the 3rd person forms to one invariant form in the present. It is notable that Lithuanian retains a very conservative morphology overall, yet even here the copula shows some irregularity. In addition to these, Lithuanian can use an alternative present *bū́na* (from the *bū-* stem) to express habitual or repeated actions (literally "be occasionally"). This is not the ordinary present but rather a special iterative aspect form. The *bū-* stem is related to the verb for "to become" (as in *būti* itself), reflecting again the interplay of two roots *es* and *bhu* in Indo-European copulas.

To summarize the present tense: all seven languages have forms traceable to the PIE root *h₁es-*. English *is*, German *ist*, Latin *est*, Persian *ast*, Sanskrit *asti*, Lithuanian (Old) *esti/yra*, and even the seldom-used Russian *est'* are all descendants of PIE *h₁ésti* "(he) is." The first-person forms likewise reflect PIE *h₁ésmi* > English *am*, Sanskrit *asmi*, Persian *am*, Russian (Old Church Slavonic) *jesmĭ*, Lithuanian *esu*. These cognates illustrate the remarkable preservation of the verb "to be" across millennia. At the same time, suppletion is already evident: English and German require a different stem for "I am" vs. "he is," and Lithuanian uses an unexpected form for "is." The high frequency of this verb tends to preserve ancient forms (hence the cognates), but also invites analogical leveling and irregular simplifications (such as Russian dropping it in the present).

Past Tense Paradigms

Table 2 presents the past tense forms of "to be" in the seven languages. For consistency, we use the simple past in each case: the English preterite (*was/were*), the German preterite (*war*, etc.), the Russian past tense (which is based on a past participle in Russian), the Persian past simple (*būd-*, "was"), the imperfect in Sanskrit and Latin (which correspond to a past state "was"), and the Lithuanian preterite. These forms generally translate to English "was" or "were." Note that in some languages the past tense of "to be" does not inflect for person the same way as the present. Russian, for example, indicates the past by gender and number, not person—a trait inherited from Proto-Slavic participial forms. We indicate Russian masculine forms for singular (*byl*) and plural (*byli*) in the table for simplicity, with a note on gender. Sanskrit and Latin have fully inflected past tenses (imperfect) for person/number. Persian and Lithuanian have one set of past endings for all persons (which we show). English and German have two forms each (singular vs. plural) in the past, due to partial leveling of older distinctions.

Table 2: Past Tense Forms of "to be" (equivalent to English "was/were")

Person	*English*	*German*	*Russian (m.)*	*Persian*	*Sanskrit*	*Latin*	*Lithuanian*
1sg "I was"	was	war	býl (masc.)	búdam	āsáṁ	eram	buvau
2sg "you were"	were	warst	býl (masc.)	búdī	āsīḥ	erās	buvai
3sg "he/she was"	was (he) / was (she)	war (er) / war (sie)	býl (he) / býla (she)	búd	āsīt	erat	buvo
1pl "we were"	were	wáren	býli (pl.)	budím	āsma	erāmus	buvome
2pl "you were"	were	wárt	býli (pl.)	budíd	āsta	erātis	buvote
3pl "they were"	were	wáren	býli (pl.)	budánd	āsan	erant	buvo

In English, the simple past of "to be" shows a singular/plural split: *was* (for *I/he/she/it*) versus *were* (for *you/we/they*). Old English actually distinguished second-person singular (*wāere*, "thou wert") from others, but Modern English leveled all plural and the singular *you* to *were*. Both *was* and *were* come from a different root than the present forms. These are historically from the Proto-Germanic verb *wesaną* ("to remain, dwell"), which in turn comes from PIE *h₂wes-* ("to reside, live"; *Indo-European copula*, n.d.). English *was*

reflects the PIE o-grade *wos-*, and *were*, with /r/, is from a different ablaut grade *wēs-* (the alternation *s/r* is due to a Germanic sound law—Verner's Law—or analogical leveling; *Indo-European copula*, n.d.). Thus, the English past tense *was/were* is not cognate with Latin *erat* or Sanskrit *āsīt*; instead, it shares origin with, for example, German *war/waren* (which clearly resemble English *was/were*) and Old Norse *var*. German's past *war*, *warst*, *waren*, *wart* likewise come from the *wes-* root, as indicated by the *w-* in the present and the *r* ~ *s* alternation in some forms (*Indo-European copula*, n.d.). This Germanic-specific development means that the Germanic languages use a separate PIE root for the past of "to be," distinct from the one used in the present. In contrast, many other Indo-European branches used the main *es-* root even for the past in their older stages.

Latin and Sanskrit exemplify the use of the *es-* root for the past through their imperfect tense of the verb "to be." Latin forms are *eram, erās, erat, erāmus, erātis, erant*, which translate to "I was, you were, ... they were." These can be derived from PIE as well: Latin *eram* < *es-m̥*, with the characteristic Latin change of *s* to *r* between vowels, called rhotacism. In fact, one theory posits a PIE root *h₁er-* used for an archaic past or stative verb (possibly meaning "to arise, appear"), to explain the *r* in Latin *eram* (*Indo-European copula*, n.d.). However, an alternative explanation is that Latin simply took the present stem *es-* and applied a sound change (PIE *ésm̥* > *esom* > Proto-Latin *esom* > *erom* > *eram*, where *s* became *r* and final *m* became a nasalized vowel). The consensus leans toward Latin *eram* being an innovative imperfect formed within Latin (or Italic) rather than a direct inheritance from PIE *h₁er-* (*Indo-European copula*, n.d.). Regardless, Latin *erat* ("he was") is cognate with Sanskrit *āsīt*. Sanskrit's imperfect conjugation of *as-* is *āsam, āsīḥ, āsīt, āsma, āsta, āsan*, which corresponds to PIE *h₁és-m̥, h₁és-s* (or *h₁és-ti* with augment), etc. The Sanskrit forms show the augment *a-* prefixed (a past tense marker in Indo-Iranian), and lengthened initial vowel (*ā* instead of *a*), which is typical in the imperfect. For example, Sanskrit *āsan* ("they were") = PIE *h₁és-ent* > *ás-ent* (augment) > *āsan*. These forms are clearly related to the present-tense forms (just with past augment and secondary endings), meaning Sanskrit did not need a new root for the past. Lithuanian behaves more like Sanskrit and Latin in this respect: its simple past *buvo* ("was") is built on the *bū-* stem but with the usual past tense suffix *-o/au*. Actually, Lithuanian's *buvau, buvai, buvo*... comes from the root *bū-* (from PIE *bʰuH-*, "become") plus a past-tense morpheme *-v-* and endings. In Lithuanian, as in many Indo-European languages, the PIE perfect or aorist of *bhu-* was repurposed as the past of "to be." So while Latin and Sanskrit used *es-* for the imperfect, Baltic and Slavic used *bhu-*.

Persian uses the verb *budan* ("to be, exist") for the past tense. The past simple conjugation is *budam, budi, bud, budim, budid, budand*, meaning "I was, you were, … they were." These forms clearly derive from the root *bud-*, which is the Persian reflex of PIE *bʰuH-* (with regular *bh* > *b* sound change in Indo-Iranian). We can see the connection with other languages: e.g., Persian *bud-* ~ Sanskrit *abhūt* ("he was" or literally "became" in Classical Sanskrit perfect) ~ Latin *fuī* ("I was"). Indeed, Latin's perfect *fuī* (and infinitive *fuisse*, future participle *futūrus*) comes from the same *bhu̯ā* root (*Proto-Indo-European Roots*, n.d.). The *bhu-* root was originally a verb meaning "become, come into being" rather than a stative "be." Persian has basically replaced the old *h₁es-* forms with *bhu-* forms in the past tense: Old Persian would have used forms of *ah-* (*ahata* = "was"), but those have vanished in Modern Persian, leaving only *bud-*. It is worth noting that Persian *budan* is also the source of the English loanword *Buddha* (meaning "enlightened one," literally "awakened, has become aware"), illustrating how *bud-* carries the sense of "become" (to attain a state).

In Russian (and other Slavic languages), the past tense of "to be" is formed from the old participle. The form *byl* (masculine), *byla* (feminine), *bylo* (neuter) for the singular and *byli* (plural) correspond to "was."

These come from the Proto-Slavic *bylŭ*, originally the past active participle of *byti* ("to be"). In Russian, as in all East Slavic, this participle is used as the past tense in combination with person and must agree in gender and number with the subject. In Table 2, we list *byl* for singular and *byli* for plural as representative forms (assuming a masculine subject). Importantly, *byl* is built on the stem *by-*, which is from the PIE *bʰu-* root, just like Persian *bud-* and Lithuanian *bū-*. The Slavic languages long ago lost the old *es-* past forms (if they ever had a separate imperfect, it disappeared), and the participle of *byti* took over the role of past tense. So, Russian *byl* is cognate with Sanskrit *bhūtaḥ* (a past participle meaning "been, become") and with the *fu-* series in Latin. Another relic of *bʰu-* in Slavic is the verb *бывать* (*byvat'*, "to frequent, happen to be"; *Proto-Indo-European Roots*, n.d.), which contains an iterative suffix, analogous to the Lithuanian *būna* mentioned above (*Proto-Indo-European Roots*, n.d.).

Comparing across the seven languages, we observe that two main PIE roots account for most past-tense forms of "to be": *h₁es-* and *bʰuH-*. Sanskrit and Latin use *es-* (augmented) for their imperfect "was," preserving the older pattern of conjugating *es-* in the past. The Germanic languages innovated a past based on *wes-* (with *was*, *were* in English, *war* in German)—a unique Germanic development from PIE *h₂wes-* ("dwell"; *Indo-European copula*, n.d.). Meanwhile, Baltic, Slavic, and Persian turned to the *bhu-* root to supply the past tense (*Lith. buvo*, *Rus. byl*, *Pers. bud-*), indicating a convergence on *bhu-* across these geographically and temporally separated groups. Notably, Latin also employs *bhu-* in its perfect tense (*fuī*), and Sanskrit uses it in the perfect (*babhūva*, "has become") and aorist (*abhūt*, "became"). In essence, PIE *bhu-* ("become") eventually filled in past or perfect roles in many descendants. The suppletive nature of the copula is evident: English present *am/is* vs. past *was* vs. participle *been* are all from different roots; the same goes for German *ist* vs. *war* vs. *gewesen*. Even in languages that seem more uniform, like Sanskrit, there were multiple strategies—using *as-* for the imperfect and *bhū-* for other past tenses. We also see analogical simplification in some cases: Russian and Persian do not use *es-* at all in the past (dropping it entirely in favor of *bhu-* forms), whereas English and German dropped the *es-* in their past in favor of *wes-* . These differences underscore how each branch of Indo-European resolved the coexistence of multiple "to be" roots in different ways.

Future Tense Paradigms

Not all Indo-European languages have a synthetic (single-word) future tense for the verb *"to be."* Proto-Indo-European itself did not have a dedicated future tense; futures were later innovations in the daughter languages, often formed periphrastically or by modal forms. However, for completeness, Table 3 shows the way to express the future "will be" in each of the seven languages. In English, German, Persian (modern), and Russian, the future is periphrastic – meaning it uses an auxiliary verb plus an infinitive. In contrast, Sanskrit, Latin, and Lithuanian developed inflected future forms for "to be" (though Sanskrit's is somewhat rare/literary). We list the usual future tense form in each language: English with auxiliary will, German with werden + *sein*, Persian with x̱āh-… bud construction, Russian with the conjugated bud- forms, Sanskrit with the future conjugation of *bhū-* "become," and Latin and Lithuanian with their synthetic futures of *esse* and *būti* respectively.

Table 3: Future Tense Forms of "to be" (equivalent to "will be")

Person	*English*	*German*	*Russian*	*Persian*	*Sanskrit*	*Latin*	*Lithuanian*
1sg "I will be"	will be	werde sein	budu	xâham bud	bhaviṣyāmi	erō	būsiu
2sg "you will be"	will be	wirst sein	budesh′	xâhī bud	bhaviṣyasi	eris	būsi
3sg "he will be"	will be	wird sein	budet	xâhad bud	bhaviṣyati	erit	būs
1pl "we will be"	will be	werden sein	budem	xâhim bud	bhaviṣyāmaḥ	erimus	būsime
2pl "you (pl) will be"	will be	werdet sein	budete	xâhid bud	bhaviṣyatha	eritis	būsite
3pl "they will be"	will be	werden sein	budut	xâhand bud	bhaviṣyanti	erunt	būs

In **English**, the future is formed with the auxiliary "will" (or "shall" in a more archaic or formal usage) followed by the bare infinitive be. Thus all persons say "will be" (I will be, you will be, etc.). This construction developed in Middle English and Early Modern English; Old English actually had no dedicated future tense, often using present tense or modal verbs to indicate futurity. The use of *will* (itself originally meaning "want" or "wish") as a future marker is a later grammaticalization. German similarly does not have a unique future inflection for *sein*; it uses the auxiliary werden ("become"/"will") with the infinitive sein. E.g. *ich werde sein* "I will be," *wir werden sein* "we will be." In everyday German, just like English, the present tense can also serve a future meaning given the right context, but formally *werden + infinitive* is the future tense construction. It is worth noting that werden itself is an old Germanic verb from PIE *werdh- ("to turn, become"), not related to the *es* or *bhu* roots; thus, German future uses *yet another* verb to denote future being, reinforcing the suppletive character of the copula across time.

Persian forms the future analytically by combining the conjugated auxiliary xâstan ("to want," pronounced *khāstan*, often written as خواه- *xwāh*-) with the infinitive budan. For example, xâham bud (modern Persian *khāham bud*) literally means "I will want to be," which idiomatically is just "I will be." In Table 3, *xâham bud, xâhī bud,* etc. correspond to *man xâham bud (I will be), to xâhī bud (you will be),….* Colloquially, Persian often drops *bud* when context allows (saying simply *xâham* "I will" can imply "I will be"), but the full form with *bud* is the explicit future of "to be." Historically, Old Persian did not have a distinct future, much like PIE; the modern periphrastic future in Persian emerged in the past few centuries and is analogous to English *will*. Thus, Persian's future tense uses an entirely different verb (want) plus the infinitive, further showcasing how Indo-European languages tend to recruit other verbs (*will, shall, become, want*) to express futurity for "be."

Russian (and Slavic languages in general) has a synthetic future for imperfective verbs formed by the verb budet (from *byt'*) plus an infinitive. However, *"to be"* itself (*byt'*) is usually considered an imperfective verb, and its future is conjugated form of the perfective aspect по быть (in some discussions) or just treating *буду* as the future of *быть*. In practice, Russian says буду for "I will be," будешь for "you will be," up to будут for "they will be." These are the present-tense conjugations of budu/*budut'*, which historically is the old future (subjunctive) form of *biti* that replaced any older synthetic future. So, Russian буду (budu) is cognate with the auxiliary будет meaning "will" in other contexts, but here it directly translates to "will be." Unlike English/German, Russian doesn't need an extra infinitive after *budu* when expressing "will be" – *budu* itself encapsulates *"will be"* (since *byt'* is implied as its infinitive). This is somewhat unique: Russian's "will be" is a single word (e.g. *on budet* = "he will be"), making it closer to a synthetic future. The bud- root of these forms again is from PIE $b^{h}u$-; tellingly, the same root used for the past participle *byl* "was" doubles as the auxiliary for the future, emphasizing the centrality of *bhu*- in Slavic "be" forms.

Sanskrit developed several ways to express future time. Classical Sanskrit has a simple future tense formed by adding the suffix *-syá-* (or *-iṣya-*) to the verb root. For "to be," instead of using *as-*, Sanskrit typically uses the root bhū- ("become") in the future. The future conjugation of *bhū-* is: bhaviṣyāmi, bhaviṣyasi, bhaviṣyati, bhaviṣyāmaḥ, bhaviṣyatha, bhaviṣyanti, which literally mean "I will become/am going to be," etc. In usage, bhaviṣyati can mean "he will be." We list these forms in Table 3 as the Sanskrit future, since saying *as-** in the future is not common (there is a Vedic periphrastic future *āsati* form, but classical Sanskrit prefers *bhū-*). The choice of *bhū-* highlights a fascinating point: Sanskrit uses a different PIE root for the future than for the present, relying on *bhu* (like many languages do for past or future) and reserving *as* for present and imperfect. This again reflects a semantic distinction – *bhū* carries an inchoative sense ("come to be"), suitable for a future action. Other Indo-Aryan languages (like Hindi/Urdu *hoga*) also have distinct future copulas, but Persian, though Indo-Iranian, as we saw, now uses a periphrastic construction.

Latin has a synthetic future for *esse*: erō, eris, erit, erimus, eritis, erunt, translating "I will be," etc. These come from the Latin future suffix *-bi-/-*be-* attached to the verb *es-*. However, Latin futures of *esse* are irregular in that they use er- rather than es- (except second person singular *eris* looks like *es* + *i*). Latin *erō* is thought to come from an earlier *esō* (with *s* turning to *r* by rhotacism again). In any case, Latin *erunt* ("they will be") is clearly parallel to *sunt* ("they are"), just with *-u-/-*nt* vs *-nt*. So Latin's future is largely an internal development. It did not use *bhu* for the future (since *bhu* was already taken for the perfect *fuī*). Instead, it modified the present stem. The result is that Latin *erō/erit* superficially resembles the English word "are" or "art" (and indeed, one hypothesis connected *erō* with the PIE *h₁er- root for "arise"), but as mentioned, a simpler explanation is analogical *es-* > *er-*).

Lithuanian forms the future tense with a suffix *-s-*. For *būti*, the future conjugation is **būsiu, būsi, būs, būsime, būsite, būs** ("I will be," etc.). These are straightforwardly from the *bū-* stem plus *-s- + endings. The *-s-* is actually a common Baltic (and Slavic) future marker (cf. Old Church Slavonic *budǫ, buděši…* which also has *-d-*, a different formation though). In Lithuanian *būsiu* comes from earlier *būsyu* (with *-yu* first person ending). Again, **bū-** is from PIE *bʰu-*. So Lithuanian's future is transparently inherited from a Proto-Indo-European desiderative or future formation of the *bhu* root. This gives Lithuanian a neatly suppletive paradigm: present from *es-*, past and future from *bhu-*. Many other languages show a similar split between present vs. past/future stems for "to be."

In summary, the **future tense** across these languages reinforces the pattern of *"to be"* drawing from multiple sources:

- **Germanic (English, German):** use new auxiliary verbs (*will, werden*) meaning *want/shall* or *become* to form the future, rather than developing a new form of *be* itself.
- **Slavic (Russian):** uses the *bhu- root* (*bud-*) in a conjugated form to represent "will be," effectively extending the *bhu- paradigm* to cover the future (since present *"is"* was dropped).
- **Indo-Iranian:** Sanskrit and Persian diverge; Sanskrit employs *bhū-* in a true future tense (*bhaviṣyati*), whereas Persian uses a periphrastic want+be construction (*x̂hāhad bud*). Both avoid the *as-* root for future.
- **Italic (Latin):** innovated a future *er-/* based on *es-*, possibly influenced by phonological processes like rhotacism, but kept it mostly within the original root's morphology.

- **Baltic (Lithuanian):** uses *bū-* with a future suffix, aligning with the widespread use of *bʰu* for non-present tenses.

The **implication** is that Proto-Indo-European likely did **not** have a single unified future for "to be" – each branch solved it differently. However, the strong presence of *bhu- in future or subjunctive roles (Sanskrit bhav- future, Slavic bud-, Lithuanian būs-) suggests PIE *bʰuH-* had a modal or prospective sense ("become, come to be") that made it a natural choice to express future existence. Meanwhile, the root *h₁es-* was more stative and present-oriented, and when used in a future sense (Latin er-), it needed extra markers. The **suppletion** in the future thus often mirrors that in the past: *es-* for present, *bhu-/wes/others* for future/past. This pattern is a key piece of evidence in Indo-European reconstruction – it indicates that the proto-language probably used different verbs for the concept of "be" depending on grammatical context (tense, aspect, mood).

Discussion

The comparative data from the seven Indo-European languages reveal both **remarkable similarities** and **branch-specific differences** in the verb *"to be."* These findings can be interpreted in light of historical linguistics and Proto-Indo-European (PIE) reconstruction. In this section, we discuss the paradigms' etymological origins, phonological and morphological changes, and what they imply for our understanding of Indo-European language development.

Shared Features and Cognate Forms

All languages in our sample ultimately derive their forms of *"to be"* from a **common pool of PIE roots**. The most important of these roots are traditionally reconstructed as:

- **PIE *h₁es-*** "to be (stative existence)" – the root of Sanskrit *as-* (ásmi), Latin *es-* (est, sunt), English "is," German "ist," Lithuanian *es-* (esu), Persian *ast*, Russian *est'*. This root yields the vast majority of present tense forms across Indo-European. The widespread cognates (e.g. Skt *asti*, Lat *est*, Eng *is*, Rus *est'*) illustrate how conservatively this part of the paradigm was preserved. The PIE conjugation was athematic (no thematic vowel), with endings like *-mi, -si, -ti* in the singular. Many of those endings can still be seen: Sanskrit **-mi/-si/-ti**, Old Church Slavonic **-mĭ/-sĭ** (jesmĭ, jesĭ), etc.. Even English *am* < Old English *eom* < Proto-Germanic *ezmi* shows the first person *-mi* (with regular consonant changes). This continuity is a classic example in comparative linguistics of how an irregular but essential word preserves its lineage.

- **PIE *bʰuH-*** "to become, grow" – the source of English **be** (and **been**), German **bin/bist** (via Proto-Germanic *beun-/*bi-), Sanskrit **bhavati** ("becomes") and future **bhaviṣyati**, Latin **fuī** ("was") and **futūrus** ("about to be"), Lithuanian **būti** ("to be") and **būsiu** ("I will be"), Russian **budu** ("I will be"), Persian **budan** ("to be/to become"). This root is almost as ubiquitous as *es-* but tends to appear in **non-present** contexts – either as the infinitive, the past, or a future/subjunctive. Many linguists theorize that *bhu-* in PIE functioned as an **"inactive" or dynamic verb of being**, complementing *h₁es-* which was stative. Essentially, *h₁es-* meant "to be (as a stable state)" and *bhu-* meant "to become/come into being." Over time, as the daughter languages grammaticalized tense and aspect, *bhu-* was often recruited for expressing future states or completed states (past), fitting its "coming into being" semantics. Our data strongly reflect this: languages like Persian, Russian, Lithuanian use *bhu- forms* for past or future (or both), and even Latin and Sanskrit use *bhu-* for perfect and

future respectively. The **convergence** on *bhu-* across distant branches suggests that this division of labor between *es* and *bhu* goes back to Proto-Indo-European itself. The fact that English "be" and German "bin" are cognate with Sanskrit **bhavatu** (imperative "let it become") and Russian **budet** ("will be") is striking evidence of a shared linguistic heritage.

- **PIE *$h_2$wes-*** "to reside, dwell" – this root is less general, but in **Germanic** it became the past tense of "to be" (as discussed, Eng was/were, Ger war, etc.). Outside Germanic, *$h_2$wes- may be reflected in a few words (perhaps Sanskrit *vasati* "dwells"), but it was not generally used as a copula. Its use in Germanic past tense is considered an **innovation**: Proto-Germanic created a suppletive past *was-* from this root. This was likely motivated by the semantics (to stay, remain could describe a state in the past) and by analogy with strong verb patterns. It's notable that Gothic, an East Germanic language, also had this (e.g. Gothic *was, wēsun* = "was, were"), confirming it was in Proto-Germanic. So, while *wes- forms* are not shared by the non-Germanic languages, they illustrate how *all* IE branches had multiple roots available for "be" and could pick different ones.

- **PIE *$h_1$er-*** "to move, set in motion (possibly to arise)" – this root is hypothesized to have contributed to some "to be" forms. The clearest case is the **English** archaic **"art"** (2sg *thou art*), which was borrowed from Old Norse *ert* and corresponds to a Proto-Germanic form *ir- or *ar- for second person). Old Norse *em, ert, es* ("am, art, is") suggests North Germanic had an *er- stem in present alongside *im-* (from *esmi*). Some scholars trace this *er-/*ar-* to PIE *$h_1$er-*, giving an early suppletive present paradigm (Ringe, 2006, as cited in, argues however that Germanic *ert* might be explained via sound laws from *es-). Additionally, **Latin** imperfect *eram, erās…* and future *erō, eris…* contain *er-*. These could potentially come from an Proto-Indo-European root like *$h_1$er-, meaning the idea "I arose" = "I was," though mainstream opinion favors them being internal Latin developments with *s->r* change. If *$h_1$er-* was indeed involved in Indo-European copulas, it would show that yet another verb was at play. The evidence is suggestive but not conclusive. In any case, English **are** (plural present) actually comes historically from an Old English form *aron* (in Northumbrian dialect) which was influenced by Old Norse and possibly goes back to the same *ar/er-*. Meanwhile, the standard West Saxon OE had *sind(on)* for "are" (from *s-enti*), which German retains as **sind**. This is a fascinating instance where **dialect mixing** introduced a different suppletive form into English – effectively, modern English "are" is a Viking contribution, whereas "am/is" are native West Germanic heritage. Thus, even within one language, the copula can be a patchwork of pieces from various sources.

Beyond roots, we see common **inflectional patterns** that point to shared origin. For example, many languages distinguish singular vs plural in the present (sometimes using completely different stems, as in English was vs were, but at least grammatically marking number). The Indo-European languages originally had not only singular and plural, but also **dual** forms for "two." Sanskrit preserves dual forms of *as-* (e.g. 1 dual *svaḥ* "we two are", 3 dual *stāḥ* "they two are"), and Old Church Slavonic had dual *věste* "you two are," etc. Lithuanian, up until the early 20th century, also had a dual (e.g. *esva* "we two are"), now obsolete. The fact that dual forms existed for "to be" in Sanskrit and Old Slavic indicates PIE *es- was fully inflected for dual as well. Another shared feature is the use of special endings in certain tenses: Latin *eram* vs *sum* shows a shift to secondary endings (the *-m* in *eram* is actually a secondary ending, as opposed to primary *-ō* in *erō*). Sanskrit *āsīt* vs *asti* similarly shows primary vs secondary endings (the *-t* vs *-ti*). These technical details reflect how the **tense-aspect system** was structured in PIE and carried into daughters: primary

endings for present/future, secondary for past (imperfect, aorist). The copula, as an athematic verb, fits into this system like other verbs did.

Branch-Specific Developments and Differences

Despite the common inheritance, each language (or branch) has shaped the copular verb in unique ways. These differences are instructive about the processes of language change:

- **Phonological evolution:** Regular sound laws have altered the phonetic shape of cognate forms. For instance, Grimm's Law in Proto-Germanic turned PIE *bh > b, *dh > d, *gh* > *g* (among other shifts), which is why PIE *bʰuH-* yields **be-** in English and **bu-** in Germanic forms. Similarly, PIE *s* between vowels became **r** in Latin (rhotacism), explaining Latin **erant** vs Sanskrit **āsan** (from *-s-/-*nt). Indo-Iranian languages turned PIE *s* at word beginning into **h** in many cases: compare Sanskrit **ás-ti** with Avestan (Old Persian) **haṣṭi** and Modern Persian **hast** – the Persian **h-** in **hast** is the continuation of an Indo-Iranian sound change (though Persian often drops it in pronunciation). Lithuanian changed PIE *sm* > *m* in first person: PIE *esmi* > Lith. **esu** (where *-mi* became *-u*). Russian and other Slavs vocalized the *-mi* ending to *-mь* and eventually dropped it (Old Russian *jesmĭ* > *es'* > Ø). These sound changes sometimes obscure the relationships – e.g. one might not guess **yra** is related to **est** without historical analysis, because Lithuanian *y* corresponds to an older *e*. Through comparative reconstruction, linguists have mapped these correspondences, which allowed us to align forms like Persian *am* ~ Sanskrit *asmi* ~ Lithuanian *esu* (all from PIE *h₁esmi*). In the Germanic forms, a notable phonological detail is Verner's Law: Proto-Germanic *was* vs *war-* (in *waren*) differ by the consonant *s vs r* due to a stress shift in PIE (unstressed *-wés- > *-wér-*). Hence English **was** (with *s*) and **were** (with *r*) reflect the same root in different phonological conditions. Such alternations are fossilized in these paradigms, making "to be" an exhibit of ancient sound laws.

- **Morphological restructuring:** Several languages show **analogical leveling** or replacement in the paradigm of "to be." English lost the distinct **thou art, he is** in everyday speech (except in archaic or dialectal usage), simplifying to a uniform **"you are"** for singular and plural. Russian's loss of a present-tense verb (zero copula) is a morphological simplification at the sentence level, though historically it was a syntactic change. Persian radically simplified the entire conjugation by using enclitic pronouns for the present and adopting a single past stem *bud-* for all persons – essentially doing away with the old personal endings except in frozen forms like *ast*. These changes often stem from general linguistic tendencies: very frequent irregular verbs tend to become **more irregular** (through suppletion) but sometimes also **abridge** (shorten) their forms. Persian's enclitics could be seen as a phonological erosion of full verbs. In contrast, Lithuanian, being conservative, kept a fairly complete set of personal endings in all tenses; yet even Lithuanian allowed an alternative present *būna* to develop for habitual aspect, showing that the verb *būti* is not entirely static.

- **Suppletion patterns:** Each language exhibits a particular pattern of suppletion, choosing different pieces from the PIE grab-bag. **English** is extreme with three roots in one tense (am/are vs is vs (archaic) art) and another root for past, plus another for participle. **German** combines two roots in present (b- forms and s- forms) and one root for past, plus a past participle *gewesen* from yet another root (*wes-* plus participial ge-). **Latin** uses *es-* for present, *er(a)-* for imperfect, *fu-* for perfect – three roots across its full paradigm. **Sanskrit** has *as-* for present/imperfect, *bhū-* for future and

perfect, and even a third root *sam-* in Vedic for aorist (an old form *āsid* sometimes from *√ah "to be"?). **Russian** arguably uses primarily *bū- (by-)* for everything (since *es- *dropped out), but it still has a relic *est'*. **Persian** split by tense: *hast* for present, *bud* for past/future. **Lithuanian** split by mood/aspect: *es-* for present, *bū-* for past/future, plus *būna* for iterative present. Despite these differences, the underlying **trend** is consistent: **no single PIE daughter relies on only one root for all forms of "to be."** This strongly supports the reconstruction that PIE itself had multiple verbs that later coalesced into a single paradigm (Watkins, 2000). The copula's suppletion is not a later accidental development in each language, but a **common inheritance** in the sense that speakers of the proto-language already had a habit of using different verbs for different inflectional categories (Hackstein, 2013). As languages evolved, they **merged** those verbs into one conceptual category ("the verb to be"), but the seams of that merger are still visible as suppletion.

- **Semantic distinctions:** Some languages maintain multiple copular verbs to differentiate contexts. While our focus is on the primary verb "to be," it's worth mentioning briefly that **Persian** historically had two verbs: *ast* vs *hast* (some grammarians differentiate them as one being more existential), though in modern Persian this is a minor nuance. **Spanish** famously has *ser* vs *estar*, and **Irish** has the substantive *bí* vs the copula *is*. These are parallel outcomes of the same phenomenon – the copular function was distributed among multiple roots (*estar* < *stare* "to stand" vs *ser* < *esse* in Latin). In our Indo-European context, Sanskrit had something analogous in early Vedic: *ásmi* vs *bhávāmi* (the latter meaning "I become" for more temporary states). Although we did not include Spanish or Irish in the seven languages, it's notable that **Indo-European languages repeatedly show a tendency to use different verbs for essence vs state**. This suggests the PIE ancestors might themselves have had subtle aspectual or semantic distinctions using *es-*, *bhu-*, *wes-*, etc. The fact that *h₂wes-* meant "dwell/live" might indicate it was used for continuous states, and *bhu-* for coming-to-be or inchoative aspects. Over time, these nuances either were lost (merging into one verb) or amplified (split into separate verbs as in Spanish). Our comparative table hints at this: e.g. Sanskrit using *bhū-* for future (implying a change of state) but *as-* for present (a state). Understanding these choices enriches our understanding of how speakers conceptualize existence vs identity vs occurrence.

Conclusion

The verb *"to be"* in Indo-European languages is irregular precisely because it fuses together several ancient verbs from Proto-Indo-European. Our comparative analysis demonstrates that English, German, Russian, Persian, Sanskrit, Latin, and Lithuanian share a set of cognate forms (for example, **s-** forms like *is/ist/est/asti* and **b-** forms like *be/bin/bud-*) inherited from PIE, even as they differ in which forms are used in which tense. These similarities have allowed linguists to confidently reconstruct the PIE paradigm (e.g. *h₁ésmi, h₁ésti, h₁sénti*; *bʰúHmi, bʰúHtis* etc. for other tenses) and understand how suppletion and sound change have shaped each daughter language. The differences, on the other hand, illustrate typical processes of language change: phonological evolution (like *s→r*, *bh→b*), analogical leveling (dropping or merging forms), and periphrastic replacement of complex morphology. In essence, the patchwork paradigm of *"to be"* in each language today is a living record of both the **shared ancestry** of Indo-European languages and the **individual developments** those languages underwent over millennia. The comparative approach, therefore, not only elucidates the etymology of each form but also highlights the enduring legacy of Proto-Indo-European in the core grammar of its descendants.

References

Ajotikar, T., Kulkarni, M., & Bhattacarrya, P. Verbs in Sanskrit Wordnet. In *GWC 2012 6th International Global Wordnet Conference* (p. 30).

Beekes, R. (2011). *Comparative Indo-European Linguistics: An Introduction* (2nd ed., revised by M. de Vaan). John Benjamins.

Bopp, F. (1862). *A comparative grammar of the Sanskrit, Zend, Greek, Latin, Lithuanian..* (Vol. 3).

Bopp, F. (2022). *Comparative Grammar of the Sanskrit, Zend, Greek, Latin, Lithuanian, Gothic, German, and Sclavonic Languages: Vol. II*. BoD–Books on Demand.

Deshpande, M. M. (1992). Justification for verb-root suppletion in Sanskrit. *Historische Sprachforschung/Historical Linguistics*, *105*(1. H), 18-49.

Fortson, B. (2010). *Indo-European Language and Culture: An Introduction* (2nd ed.). Blackwell.

Hock, H. H. (2015). Proto-Indo-European verb-finality: Reconstruction, typology, validation. In Proto-Indo-European syntax and its development (pp. 51-78). John Benjamins Publishing Company.

Lambton, A. K. S. (1953). *Persian Grammar*. Cambridge University Press.

Luraghi, S. (1995). The pragmatics of verb initial sentences in some ancient Indo-European languages. *Word order in discourse*, *30*, 355.

Luraghi, S., Inglese, G., & Kölligan, D. (2021). The passive voice in ancient Indo-European languages: inflection, derivation, periphrastic verb forms. *Folia Linguistica*, *55*(s42-s2), 339-391.

Ostler, N. D. M. (1979). *Case-linking: a theory of case and verb diathesis applied to classical Sanskrit* (Doctoral dissertation, Massachusetts Institute of Technology).

Remys, E. (2007). General distinguishing features of various Indo-European languages and their relationship to Lithuanian. *Indogermanische Forschungen*, *112*, 244.

Ringe, D. (2006). *From Proto-Indo-European to Proto-Germanic*. Oxford University Press.

Watkins, C. (2000). *The American Heritage Dictionary of Indo-European Roots* (2nd ed.). Houghton Mifflin.

Received: 10.04.2025
Revised: 12.04.2025
Accepted: 13.04.2025
Published: 18.03.2025