UDC 81-139; 81'32
DOI https://doi.org/10.24919/2308-4863/97-1-33

Anatolii BEZPALENKO,
orcid.org/0000-0002-2389-7442
Doctor of Philological Sciences,
Professor at the Department of Foreign Languages for Mathematical Faculties
Educational and Research Institute of Philology of Taras Shevchenko National University of Kyiv
(Kyiv, Ukraine) proftolik@ukr.net

APPLICATION OF BAYES' THEOREM TO DETERMINE THE QUANTITATIVE STATUS OF VOCALISM IN A PROTO-INDO-EUROPEAN LANGUAGE

This article is the latest in a series of articles by the author in which he applies mathematical approaches to the analysis of linguistic phenomena (see Humanities Science Current Issues. Issues 35, Volume 1, 2021; and Issue 40, Volume 1, 2021).

In particular, this article draws on Bayes' theorem, a concept from probability theory, which the author uses to determine the quantitative status of the vowel system of the Proto-Indo-European language. As is well known, the question of the number and nature of vowels in Proto-Indo-European is a matter of debate. The prevailing view in linguistics is that the number of vowels ranges from one phoneme to ten. It seems logical to us that the set of vowels in Proto-Indo-European can be determined using three factors: 1) the known sets of vowels in modern and ancient languages of the Indo-European area, including extinct languages; 2) the alternations of these vowels; and 3) Bayes' theorem. To achieve this aim, the author proposes using the vowel systems of 13 Indo-European languages (10 living languages – Ukrainian, English, Greek, German, French, Spanish, Bulgarian, Turkish, Armenian, Hindi – and 3 extinct languages – Latin, Sanskrit, Hittite).

Applying Bayes' theorem to phonetic phenomena for the first time in linguistics, the author introduces into this theorem a special coefficient $\sum$***ph* x *2****, which represents the sum of vowel phonemes in the language, multiplied by two. The author explains the nature of this coefficient by the fact that, with every act of articulating a particular vowel phoneme, the speaker faces a dilemma: the need to consciously or unconsciously choose one of ***TWO*** possible variants of the same phoneme, or one of two adjacent phonemes. With good training, i.e. when the correct pronunciation is established, this may be a conscious choice of one correct (standard) phoneme or its variant; or, as happens with insufficient training, an unconscious erroneous choice of a phoneme variant or an adjacent phoneme, i.e. when the articulatory intention becomes replaced with its actual execution. Thus, drawing on Bayes' theorem and etymological research, the author demonstrates that the Proto-Indo-European language could not have had more **than three vowel phonemes**. Furthermore, the author derives a general universal: the number of vowel phonemes in a language is inversely proportional to the magnitude of their alternation index.*

Key words: *Probability theory, vowel alternation. Proto-Indo-European language, etymology, Bayes' theorem.*

Анатолій БЕЗПАЛЕНКО,
orcid.org/0000-0002-2389-7442
доктор філологічних наук,
професор кафедри іноземних мов математичних факультетів
Навчально-наукового інституту філології
Київського національного університету імені Тараса Шевченка
(Київ, Україна) proftolik@ukr.net

ЗАСТОСУВАННЯ ТЕОРЕМИ БАЙЄСА ДЛЯ ВИЗНАЧЕННЯ КВАНТИТАТИВНОГО СТАТУСУ ВОКАЛІЗМУ ПРАІНДОЄВРОПЕЙСЬКОЇ МОВИ

Пропонована стаття є продовженням серії статей автора, в яких він застосовує математичні підходи до аналізу мовних явищ (див. Актуальні питання гуманітарних наук. Вип. 35, том 1, 202; а також Вип. 40, том 1, 2021).

Зокрема, в цій статті застосовується така категорія теорії ймовірності, як теорема Байєса, яку автор використовує для визначення квантитативного статусу вокалічної системи праіндоєвропейської мови).

Як відомо, питання про кількість і якість голосних у праіндоєвропейській мові є предметом дискусій. У лінгвістиці переважають погляди, згідно з якими кількість голосних коливається від однієї фонеми до десяти. Нам видається логічним, що множину голосних праіндоєвропейської мови можна визначити за допомогою трьох факторів: 1) відомих множин голосних сучасних і давніх мов індоєвропейського ареалу, включаючи вимерлі мови; 2) чергувань цих голосних; а також 3) за допомогою теореми Байєса. Автор пропонує для досягнення цієї мети використати голосні системи 13 індоєвропейських мов (10 живих – українську, англійську, грецьку, французьку, іспанську, болгарську, турецьку, вірменську, грецьку, хінді та 3 вимерлі – латинську, санскрит, хетську).

Вперше в лінгвістиці застосовуючи теорему Байєса до фонетичних явищ, автор вводить до цієї теореми спеціальний коефіцієнт ***∑ph x 2****, який виражає суму голосних фонем у мові, помножену на два. Природу такого коефіцієнта автор пояснює тим, що при кожному акті артикуляції тієї чи іншої голосної фонеми перед мовцем кожного разу стоїть дилема, яка полягає в необхідності свідомого чи несвідомого вибору одного з* ***ДВОХ*** *можливих варіантів однієї й тієї ж фонеми, або ж однієї з двох суміжних фонем. При хорошому тренінгу, тобто при постановці правильної вимови це може бути свідомий вибір однієї правильної (стандартної) фонеми чи її варіанта, або ж, що буває при недостатньому тренінгу, несвідомий помилковий вибір варіанта фонеми чи суміжної фонеми, тобто коли артикуляційний намір замішається реальним його виконанням. Таким чином, спираючись на теорему Байєса та етимологічні дослідження, автор доводить, що праіндоєвропейська мова* ***не могла мати більше трьох голосних фонем.*** *Крім того, автор виводить універсалію: кількість голосних фонем у мові обернено пропорційна величині індексу їх альтернацій.*

Ключові слова: *теорія ймовірностей, чергування голосних, праіндоєвропейська мова, етимологія, теорема Байєса.*

It is well known that the reconstruction of the Proto-Indo-European language (PIE) through the comparison and analysis of common and distinctive features of living Indo-European languages (I-E) is one of the greatest achievements of linguistics during its existence as a science.

Statement of the problem. The proposed discourse draws attention to a **contradiction** that has been taking place in linguistics during the 19th and 20th centuries, and which for some reason no one pays attention to. On the one hand, if we believe etymological studies and etymological dictionaries, we can observe that **in every single PIE root, without exception, there is an alternation of vowels.**

Examples.

*PIE root *g**E**l– means cold, to freeze – in Eng ch**I**lly (causing a sensation of cold) – Old English ciele, c**E**le (coolness) – c**O**ld – Eng cool [k**U:**l] (Harper, 2000) – Ger k**A**lte (coldness). All 6 vowels of the Hellwag triangle are involved in alternation. (This example demonstrates that the PIE proto-sound *E alternated with *I, *A, *O, *U, which could be either allophones or fixed phonemes, which later in different languages definitely turned into separate phonemes).

*PIE root *s**A**l– in Eng silt [s**İ**lt] (fine sand, clay, or other material carried by running water) – Eng s**E**ll, Ukr c**Е**ль – Ukr c**А**льдо (balance), Lat s**A**larium (an allowance, a stipend, a pension, said to be originally "salt-money, soldier's allowance for the purchase of salt" – Eng sal [s**Æ**l] (salt) – Eng salary ['s**Æ**ləri] – Eng salt [sɔ:lt], to solder ['sɔldə] (to braze) – Sp s**O**ldar (to solder) – Eng soldier ['səʊldʒə] – Sp s**U**eldo (the 1st person from 'soldar') – Lat s**U**lphur (salt of sulfur) (Harper, 2000). All 6 and more vowels of the Hellwag triangle are involved in alternation.

*PIE root ***O**yno-, **O**yko– means Eng *one* [w**E**n] – in Pre-Sl **İ**nŭ – Gr h**E**īs, Eng alone [ə'lƏʊn] – Got Ains, Toch s**A**s – Lat **Ū**nus (Harper, 2000). All 6 vowels of the Hellwag triangle are involved in alternation etc.

On the other hand, it is noted that Proto-Indo-European had 10 vowel phonemes. However, given the total alternation of vowels, such a number of phonemes is impossible.

So, the **aim of this article** is therefore to resolve the contradiction mentioned above and to determine the exact number of vowel phonemes in Proto-Indo-European**.**

Review of the literature. In linguistics, it is widely accepted that Proto-Indo-European had 10 vowel phonemes: five short ones (i, e, a, o, u) and five long ones (ī, ē, ā, ō, ū) (Saussure, 1887; Meier-Brügger. 2003: 75). This contradiction leads to the conclusion that either all etymologists and etymological dictionaries are wrong, which is impossible to believe, or there were no 10 vowel phonemes in PIE, as is traditionally believed.

Presentation of the main material. There is obviously a certain correlation between the number of vowel phonemes in a language and their alternation index, which is logical. **The smaller the depot of vowel phonemes in a language, the higher the index of their alternation.** After all, if the vowel alternation index in PIE is equal to 100%, then the existence of 10 vowel phonemes here is impossible. Each PIE root did not remain unchanged until it was fixed in any I-E language. In other words, when any PIE root is identified in any I-E language, it always turns out that its vocalic component was constantly changing at the PIE stage. The vowel alternations were stable and constant, involving several neighbouring vowels, or even the entire number of vowels in the Hellweg's triangle. Ch. Hellwag depicted his triangle in a mirror image, which is unusual in comparison to its later images in phonetics textbooks:

u ü i

o ö e

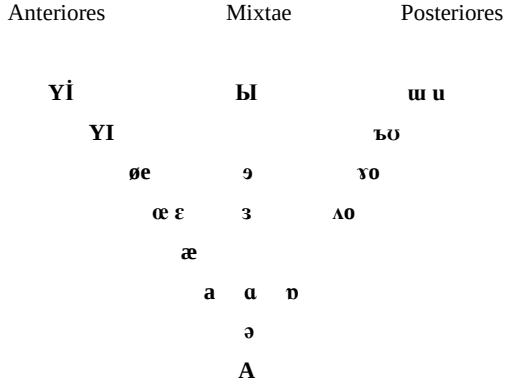
å ä

a

(Hellwag, 1781)

In all variations of the image of the Hellwag triangle, which are found in other researchers, for example, A. Bell in 1894, and Daniel Jones (Jones, 1997: IV) the fixation of the cardinal, i.e. the most distant vowel phonemes in the triangle remains unchanged. These three phonemes form a field of phonetic events and between them in this field the all variety of vowels that occurring in different languages can be gradually placed.

Later the Hellwag triangle was significantly refined by other linguists.



It is noteworthy that in the interpretations of later authors, not three but six degrees of tongue elevation are distinguished, and these degrees are not separated by any lines. The absence of lines indicates that from 'I' to 'A', just as from 'A' to 'U', there is a series of transitional vowels arising during the slow lowering or raising of the tongue. (See the bibliography review Безпаленко, 2009: 317 – 319).

We believe that Bayes' theorem can be applied not only to determining future facts, but also to retrospective determinations; in particular, it can be used to accurately determine the number of vowel phonemes in Proto-Indo-European. Here is Bayes' theorem, which was first applied to phonetics, in particular to determine the probability of vowel alternation:

$$P\,(E/I) = \frac{P\,(I/E) \times P\,(E)}{P\,(I)},\ \text{where}$$

P (E) is the a priori probability of occurrence of phoneme (E) or any other phoneme in the root that exists in the synchrony of a particular language;

P (E/I) is the a posteriori probability of the occurrence of phoneme (E) or any other phoneme in the root, provided by existence of a variant of the root with phoneme (I);

P(I/E) is the probability of occurrence of phoneme (I) or another phoneme in the root provided by existence of a variant of the root with phoneme (E);

P (I) is the probability of occurrence of phoneme (I) or another phoneme in the root.

Our calculations have shown the expediency of introducing a special coefficient

∑ph x 2 to Bayes' theorem in its application to phonetics.

Let's explain the nature of our coefficient. Here, the component ∑ph is the sum of all vowel phonemes in a language. Where does multiplication by 2 come from? Since language is a specific **intrahuman phenomenon**, where in the process of sound production the speaker faces a conscious or unconscious dilemma of choosing between **two sounds** – either a standard phoneme or its variant, or another phoneme. At each specific moment of speech, a pattern phoneme can turn only into a **2nd allophone** or **phoneme**. That is, two sound elements (1st – image of reference sound which is kept in mind as the standard sound pattern, and 2nd – the real sound) are involved in act of pronunciation, not three or four! Every time a sound is produced, the speaker's mind struggles between **two options** – the intention and its performance. In the terminology of I. Mennen and co-authors, these two components of sound production are defined as 1 – "articulatorily-specified settings" and 2 – "the acoustic output." (Mennen I., Scobbie J., and others, 2010:34). In other words, when the intention to articulate is replaced by the actual act of articulation.

Let us try to establish a correlation between the index of vocalic alternations and the number of vowel phonemes in ten modern living I-E languages (Ukrainian Armenian, Bulgarian, German, French, Hindi, Turkish, Spanish, Greek and English) and three dead languages (Latin, Hittite, Sanskrit) whose vocalic system is well described, and on the basis of the data obtained, derive the number of vowels in PIE.

For Gr, Hit, Sp (5 phonemes): $P\,(E/I) = \frac{\frac{1}{4} \times \frac{1}{5}}{1} \times (5 \times 2) \approx 0{,}50\ (50\%)$

For Arm, Bul, Ukr, Lat (6 phonemes): $P\,(E/I) = \frac{\frac{1}{5} \times \frac{1}{6}}{1} \times (6 \times 2) \approx 0{,}39$

For San, Tur (8 phonemes): $P\,(E/I) = \frac{\frac{1}{7} \times \frac{1}{8}}{1} \times (8 \times 2) \approx 0{,}28$

For Hin (9 phonemes): $P\,(E/I) = \frac{\frac{1}{8} \times \frac{1}{9}}{1} \times (9 \times 2) \approx 0{,}25$

For Eng (12 phonemes): $P\,(E/I) = \frac{\frac{1}{11} \times \frac{1}{12}}{1} \times (12 \times 2) \approx 0{,}18$

For Ger, Fr (15 phonemes): $P\,(E/I) = \frac{\frac{1}{14} \times \frac{1}{15}}{1} \times (15 \times 2) \approx 0{,}143$

For PIE, if we assume the existence of 10 phonemes, then

$$P\,(E/I) = \frac{\frac{1}{9} \times \frac{1}{10}}{1} \times (10 \times 2) \approx 0{,}22\ (22\%).$$

But this assumption cannot be correct, because it contradicts the above-mentioned 100% vocalic alternation in PIE.

Conclusions. Thus, the only number of vowels in Proto-Indo-European that is 100% comes from Bayes' theorem is **three phonemes**: $P\,(E/I) = \frac{\frac{1}{2} \times \frac{1}{3}}{1} \times (\mathbf{3} \times 2) \approx 1{,}0\ (100\%)$, which is realistic, because it is consistent with the fact that there is 100% vowel alternation here. And these three phonemes were the cardinal ones in the Hellwag triangle – **I, A, U**.

BIBLIOGRAPHY

1. Безпаленко Анатолій. Принцип суміжності у мові. Слово у дзеркалі ґештальт-теорії. Монографія. Київ. Видавничо-поліграфічний центр "Київський університет". 2009. 424 с.
2. Harper Douglas. Online Etymology Dictionary, 2000. URL : https://www.etymonline.com/ (Date of addressing: 31.10.2025).
3. Hellwag Christoph Friedrich. Dissertatio de formatione loquelae. Heilbronn: 1781. URL: https://www.digitale sammlungen.de/de/view/bsb10582927?page=9 (Date of addressing: 21.06.2025).
4. Jones Daniel. English Pronouncing Dictionary. Originally completed by D. Jones. Extensively revised and edited by P. Roach and J. Hartman. 15th edn. Cambridge : Cambridge Unvs. Press. 1997. 559 p.
5. Meier-Brügger Michael. Indo-European Linguistics. 2003. 384 p. URL: https://books.google.com.ua/books?id=49xq3UlKWckC&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
6. Mennen I., Scobbie J., Esther de Leeuw, Schaeffler S., Schaeffler F. Measuring language-specific phonetic set tings. *Second Language Research. 26, 1*. Published by SAGE. 2010. p.13–41. DOI: 10.1177/0267658309337617. URL: http://slr.sagepub.com/cgi/content/abstract/26/1/1
7. Saussure Ferdinand de. Mémoire sur le système primitif des voyelles dans les langues indo-européennes. Paris. 1887. F. Vieweg, (Original from University of Michigan. Digitised 19 July 2007). URL: https://play.google.com/books/reader?id=5eo6AAAAMAAJ&pg=GBS.PA1&hl=uk

REFERENCES

1. Bezpalenko A. (2009) Pryntsyp Sumizhnosti u movi. Slovo u dzerkali geshtalt-teorii. Monohrafia. [Bezpalenko A. The Principle of Adjacency in Language. The word in the mirror of Gestalt Theory. Monograph]. Kyiv. Vydavnycho-polihrafichnyi tsentr "Kyivs`kyi Universytet". 424 p. – Kyiv University Publishing and Printing Centre. 2009. 424 p. [in Ukrainian].
2. Harper Douglas (2000) Online Etymology Dictionary. URL: https://www.etymonline.com/ (Date of addressing: 31.10.2025).
3. Hellwag Christoph Friedrich (1781) Dissertatio de formatione loquelae. Heilbronn.. URL: https://www.digitale sammlungen.de/de/view/bsb10582927?page=9 (Date of addressing: 21.06.2025).
4. Jones Daniel (1997) English Pronouncing Dictionary. Originally completed by D. Jones. Extensively revised and edited by P. Roach and J. Hartman. 15th edn. Cambridge: Cambridge Unvs. Press. 559 p.
5. Meier-Brügger Michael (2003) Indo-European Linguistics. 384 p. URL: https://books.google.com.ua/books?id=49xq3UlKWckC&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
6. Mennen I., Scobbie J., Esther de Leeuw, Schaeffler S., Schaeffler F. (2010) Measuring language-specific phone tic settings. *Second Language Research.* 26, 1. Published by SAGE, p.13–41. DOI: 10.1177/0267658309337617. URL: http://slr.sagepub.com/cgi/content/abstract/26/1/1
7. Saussure Ferdinand de (1887) Mémoire sur le système primitif des voyelles dans les langues indo-européennes. [A paper on the primitive vowel system in Indo-European languages] Paris. F. Vieweg, (Original from University of Michigan Digitised 19 July 2007). URL: https://play.google.com/books/reader?id=5eo6AAAAMAAJ&pg=GBS.PA1&hl=uk [in Fench]

Дата першого надходження статті до видання: 04.03.2026
Дата прийняття статті до друку після рецензування: 25.03.2026
Дата публікації (оприлюднення) статті: 19.05.2026