

MAREK STACHOWSKI
Jagiellonian University in Kraków
marek.stachowski@uj.edu.pl
ORCID: 0000-0002-0667-8862

HOW MANY TYPES OF ETYMOLOGY ARE THERE?

Keywords: etymology, folk etymology, perceptual etymology, lexicology, general linguistics

Abstract

The aim of this article is to attempt to determine how many types of etymology can be distinguished and what relations exist between them. The final results are presented at the end of the article in the form of a table and a diagram.

1. What will not be discussed

Before describing the notion “type of etymology” it is necessary to clarify which kinds of etymology will not be considered in the main part of this study.

First of all, we will confine our discussion to lexical etymology. Sometimes, albeit not very often, one can come across the term “structural etymology”, which is used for the explanation of the origins of idiomatic phrases or grammatical morphemes. Mailhammer (2013) presents the proceedings of a conference organized by the Australian National University, Canberra, in April 2010. The editor of the volume, the German linguist Robert Mailhammer, says in the introduction as follows:

Traditionally, etymology has been mainly perceived as being concerned with words, but there is no reason why this should necessarily be the case. In principle, anything can be studied etymologically, simply by asking the question, “where did that come from?” (Mailhammer 2013: 3)

As a matter of fact, however, the key question of etymology is not “where did that come from?” but rather “what was its original meaning?”, as the etymology of the word *etymology* clearly shows. Of course, the reconstruction of the original structure

is very important – some would even think it is crucial – but the truth of the matter is that the reconstructed structure is only a step towards the reconstruction of the original meaning.

If a Slavic etymologist reconstructs the etymon of the Polish word *krowa* ‘cow’ as Proto-Slavic **korva*, he is right, but he cannot be satisfied, because he still does not know what the Proto-Slavic word literally meant. Therefore, he leaves his own Slavic territory and moves on to Indo-European linguistics. Only after he discovers that the word **korva* originally meant ‘horned’ can he consider his task completed.

In the case of migratory words, the reconstruction of meaning is usually even more complex than that of structure (cf. K. Stachowski 2024a, 2024b, where various – and sometimes quite distant – meanings of English *pudding* and its European equivalents are discussed).

Non-lexical etymologies may or may not concern meaning. Sometimes they do – for instance, when the German idiom *unter aller Sau* ‘beneath contempt/criticism’ is explained by separating the German *Sau* from its homonym *Sau* ‘sow, hog’ and deriving it instead from Yiddish (ultimately < Hebrew) *sea* ‘measure’ (Olschansky 1996: 176), so that its original meaning would have been ‘beneath any measure’.

In short, structural etymologies lie beyond the scope of this study, unless they contribute to the explanation of the original meaning.

“Formal”, “comparative”, and “semantic” etymology are, in essence, not types of etymology. Rather, they represent three aspects of etymological research that should be taken into account in every etymological analysis.

Likewise, the terms “native” and “contact” etymology are not types either; they merely indicate the direction and scope of research. Similarly, the terms “regressive” and “progressive” etymology refer only to the direction of presentation. Both groups of terms will therefore be excluded from the discussion that follows.

2. Definition. Linguistic and perceptual etymology

A **type of etymology** can be defined as

*etymological explanations based on
the same methodological principles and procedures.*

For listed criteria, see section 6.

We are all generally aware of two distinct types of etymology: scholarly (or linguistic) etymology on the one hand, and folk etymology on the other. Various names are used for folk etymology,¹ but most of them suggest something inferior,

¹ Among others: English *associative etymology*, *etymological reinterpretation* || Polish *etymologia asocjacyjna*, *par(a)etymologia*, *analogia wyrazowa* || German *Paretymologie*, *Pseudesemantisierung*, *Volks(um)deutung*, *Sekundärmotivation*, *Scheinsemantisierung*, *metaphysische Etymologie* || French *étymologie seconde*, *attraction sémantique*, *étymologie populaire savante* || Turkish *yanlış kökenleme*, *köken yakıştırma* (sources: Olschansky 1996: 109–111; M. Stachowski 2022: 64, fn. 5; Parlar 2025: 315, fn. 3).

non-scholarly, marginal, and socially associated with a lack of genuine value. In short, linguistic etymology is good; folk etymology is bad.

But is it really so? We shall return to this question shortly. Before that, however, two other kinds of etymology should be mentioned.

The “misguided learned etymology”,² as Liberman calls it, can be assessed in different ways. Liberman (2005: 50) says that it “does not differ from folk etymology”. This opinion is accurate from the standpoint of its cognitive value. However, such etymologies are intended and published as serious linguistic proposals (both approaches are also reflected in the terminology; cf. German *fehlgehende gelehrte Etymologie* [i.e. a type of scholarly etymology] vs. *gelehrte Volksetymologie* [i.e. a type of folk etymology]). This is the case with Papp’s (1956: 287, 290) explanation of the Hungarian “imperative word” *gyere* ‘come! (2nd sg.)’. It stands isolated in the Hungarian lexicon, has no cognates, and possesses no other grammatical forms besides a secondary plural imperative.³ Papp suggests the following derivation:

- [2a] *ered* ‘to go/move away’ → 2nd sg. imperative *eredj* > **eregy*
- [2b] repetition similar to shouting: **eregy-eregy-eregy*
- [2c] apheresis, metanalysis and apocope: **[ere]gy-ere,gy-ere[gy]*
- [2d] reversal of meaning: ‘go away’ → ‘come (here)’
- [2e] extraction of a single word from the repeated sequence in [2c]: *gyere!* ‘come (here)’

Two things cannot be readily accepted. Firstly, steps [2b], [2c], and [2d] are anything but obvious and are not based on any material evidence – they are pure fantasy on the part of the author. Secondly, even though processes such as apheresis, metanalysis, apocope, reversal of meaning, and the extraction of a word from a syntagm are well known, their simultaneous occurrence is hardly credible. Besides, none of these steps is documented in philological sources.

All in all, one cannot resist the impression of uncontrolled imagination. Yet Papp’s study was published in an international Finnish linguistic journal and includes references to academic literature. That etymology should certainly be classified as “misguided”. It was intended as a scholarly explanation, but the author was unsuccessful, and the final result has the value of folk etymology. I do not think it should be accepted as a separate type of etymology but rather as an unfortunate linguistic proposal.⁴

² = Polish *etymologia natchniona*, German *fehlgehende gelehrte Etymologie* (Olschansky 1996: 150), *gelehrte Volksetymologie*.

³ “Wohl wurden danach auch andere konjugierte Formen des Imperativs gebildet (*gyerünk, gyertek*), doch haben dieselben kein Moduszeichen. Formen wie **gyer-jü-nk, *gyer-je-tek* gibt es nicht” (Papp 1956: 290).

⁴ Another Hungarian scholar, István Futaky (2001: 79), suggested a much more realistic etymology, based on the plausible conjecture that some of the so-called European Avars spoke a Tungusic dialect, and thus: Hungarian *gyere!* < Tungusic future imperative 2nd sg. **žiru* < *ži-* ‘to come, approach’. – It is impossible to decide nowadays whether the Tungusic borrowing was originally a slang (or perhaps a secret) word. This, however, seems quite likely – as is also

The last kind to be presented here is perceptual etymology. The term was introduced in 2021 and more precisely defined in 2022. For convenience, let us repeat that definition here: “*Perceptual etymology concerns itself with identifying the language which, in the eyes of the borrowing society, represents the culture most closely associated with the specific object or phenomenon*” (M. Stachowski 2022: 64).

Let us now look at a few examples:

- [2f] ‘Evergreen scrubland typical of the Mediterranean area, especially of Italy’, is called *macchia* in Italian and *maquis* in French. The Turkish reflex *maki* clearly goes back to the French word. Nevertheless, as Altun (2021: 62) shows, some Turkish dictionaries list it as a loanword from Italian. The authors of those dictionaries are evidently misled by a geographical and cultural fact: *maki* is typical of Italy.
- [2g] Slovenian *čorba* ‘soup’, *burek* ‘filled pastry’, and *čevapčič* ‘a grilled meat dish’ (Dapit 2021) are not Turkish loanwords, since Slovenian has never had direct contact with Turkish. Rather, they were borrowed from Croatian (*čorba*, *burek*, *čevapčić*, respectively). Thus, for instance: Slovenian *čorba* < Croatian *čorba* < Turkish *çorba* < Persian *šōrbā* ‘id.’. Why, then, have some authors chosen exactly the Turkish form from this chain? It is neither the immediate nor the ultimate source. The only plausible explanation for such a choice is an anthropological one: they correctly associated the Slovenian words with Turkish influence on Balkan cuisine.
- [2h] The Old Uyghur word *imga* ‘merchandising manager, fiscal director’ is usually regarded as a Chinese loanword, which seems perfectly convincing in view of the strong Chinese influence on Old Turkic – including Old Uyghur – administration. However, at the time of possible contact (approximately the 9th century, during the Tang dynasty), the Chinese word was pronounced *yaya* (< Late Middle Chinese *ap-nga* = [ap-ŋa] ~ [ʔja:p-ŋja:]⁵). Awareness of the extensive Chinese influence appears to have precluded consideration of any other possible source language for the Old Uyghur term. It was only in 2022 that an alternative (and better) etymology was proposed. A young Turkish linguist, Hüsni Ç. Arslan, after reading about the concept of “perceptual etymology” in my article (M. Stachowski 2022), managed to identify a more plausible match: the Late Middle Chinese word was first borrowed into Khotan Saka as *amäga*, and only later passed from Saka into Old Uyghur as *imga* (Arslan 2022: 87–88).

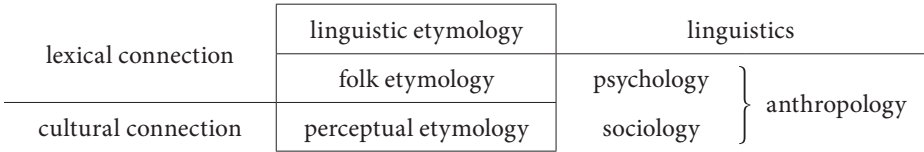
For some lexemes in this group (this is particularly evident in the case of the direct Middle Chinese etymology proposed for the Uyghur word *imga*), the paraphrase

the case with the following Slavic slang words: Czech *an* ‘bring it here!’ < Romani imperative 2nd sg. *an* < *anel* ‘to bring, to give’; Serbo-Croatian *ik* ‘look!’ < Romani imperative 2nd sg. *ikh* < *ikhel* ‘to look (at), to watch’ (Vučković 2017: 444, 447).

⁵ I thank T. Majtczak (Kraków) for his sinological consultation.

coined by Zaikovsky (2007: 182) aptly applies. The term *ετυμολογία* ‘etymology’ (< *ἔτυμος* ‘true’) can be reinterpreted as *έτοιμολογία* (< *έτοιμος* ‘ready’), that is, an approach based on the “application of ready-made patterns”.

The interrelationship between linguistic and perceptual etymology is illustrated in the following diagram (M. Stachowski 2022: 66):



3. Folk etymology – General remarks

Folk etymology concerns three groups of words: foreign loanwords, compound words, and derivatives from an unknown base. Their common denominator is unintelligibility – that is, their meaning cannot be deduced from their constituent morphemes unless these have been altered or explained. I therefore cannot fully share Durkin’s view on the nature of folk etymology: “Arguably, etymology is not really involved in the process at all: speakers are, unconsciously, altering word forms in order to create iconic connections with other words, rather than in an effort to explain their origins” (Durkin 2009: 204). There is a striking contradiction between “unconsciously” and “in order to create”. Moreover, “iconic connections with other words” do indeed result from “an effort to explain [the] origins”. I think, however, I understand what Durkin might have meant: the author of a folk etymology does not consciously inquire about the original meaning – rather, he creates it. Just as “[a]t all times some people believe that they can coax an etymology out of a word by looking hard at it” (Lieberman 2005: 4), others believe they can do so by freely altering sounds and syllables. In fact, there is no essential difference between them and *Ménage*, except that *Ménage* published etymological dictionaries.

One point should be emphasized in this context. The term “folk etymology” is sometimes used with reference to virtually any loanword that has undergone formal change. This is certainly incorrect. Words such as Polish (colloquial) *babington* (< English *badminton*) and Polish (archaic, dialectal) *okowita* ‘hard liquor, vodka’ (< Latin *aqua vitae* ‘id.’) are not etymologies at all. They are simply distortions, probably resulting from the inaccurate recollection of a foreign word. With regard to such words, Durkin’s position is correct and well founded. They will be disregarded in the remainder of this paper.

Another group consists of words that have undergone formal changes due to their negative connotations. This is the case with the surprising name of the Turkish town *Denizli*, whose literal meaning is ‘marine, maritime’ – a deceptively easy etymology (< *deniz* ‘sea’ + adjective suffix *-li*). The puzzling point, however, is that the town is almost two hundred kilometers from the sea. Originally, the town was

called *Domuzlu* (< *domuz* ‘pig, swine’ + adjective suffix *-lu*) (Korkmaz 2009: 248). The name was, therefore, morphologically transparent and perfectly understandable. From a linguistic point of view, no deformation was necessary. However, the disgust of the Muslim population towards pigs created the need to change the name. The modification yielded a new name, phonetically similar but free of negative associations, equally understandable, and therefore preferred by the population, even though it produces a confused geographical impression. This change cannot be called an “etymology”, because *Domuzlu* was semantically clear, while *Denizli* rather conceals than explains the original meaning.

A special case in this context is the Turkish city name *İstanbul* [-mb-]. As a Turkish reflex of the Middle Greek dialectal syntagm *stambóli* < *stan Póli(n)* ‘to/in the City’ (Stachowski, Woodhouse 2015: 241), it is essentially a distortion. However, its later descendant form (used primarily in the speech of lower classes), *İslambol*, lit. ‘full of Islam’, clearly results from the desire to understand the name – to find out what it really means. One can even imagine, albeit hardly prove, that people who called the city *İslambol* believed this was the correct form, whereas *İstanbul* was a disguised one. In other words, *İstanbul* results from distortion, *İslambol* from folk etymology.

4. Three kinds of folk etymology

We can now state that linguistic, perceptual, and folk etymology constitute three types of etymology. The next step is to consider the interrelations between them. However, before doing so, we need to present three kinds of folk etymology:⁶

[4a] Narration – The word itself remains intact, but a special story is invented to explain its origins. Examples include:

English *marmalade*. – “[...] the story is that when Mary, Queen of Scots was ill, her French maid would say *Marie est malade* (‘Mary is sick’). Then, someone would bring her preserved fruit to make her feel better” (M. Stachowski 2022: 61, fn. 2)

English *paper* < “*pay per sheet*” (Lieberman 2005: 12)

Polish *Okocim*, a town name. – One day the king was travelling through Poland in a carriage with his wife. When they arrived at the place where the town of Okocim stands today, the king looked out of the carriage window and was so delighted that he decided to settle there permanently, and said to his wife: “*Tu będziemy żyć i tu się okocim*” [= ‘Here we shall live, and here we shall kitten’].

Polish *szarytka* ‘a nun of the Daughters of Charity’ is a very special case. The word is said to be derived from the grey colour (Polish *szary* ‘grey’) of their habits. It is true that the Daughters of Charity, founded in 1633, originally wore grey habits and were therefore called *sœurs grises* (‘grey sisters’) in France. However,

⁶ As a matter of fact, various classifications of folk etymologies exist (cf. Olschansky 1996: 178–199). My classification proposed here is particularly suitable for comparing folk etymology with linguistic and perceptual etymology.

this fact concerns the period before 1792, when the congregation was suppressed in France and the nuns were ordered to remove their habits. The first nuns of the Daughters of Charity arrived in Poland in 1652, so it is quite possible that their name was associated with the colour of their habits (which would not be the case today, as they mostly wear dark-blue habits). Is, then, the Polish word *szarytka* a calque of French (*sœur*) *grise* and derived from *szary* ‘grey’? It may appear so, but in reality, we are dealing with a baffling coincidence. The crux is that a suffix *+tka* does not exist in Polish. There is, however, the suffix *+ka*, so the word should be divided as *szaryt+ka* (not *szary+tka*) and considered a derivative of Polish **szaryt+a* < French *charité* ‘charity’.

Russian *barsúk* ‘badger’ < Old Russian *borsúk* id. has a generally accepted Turkic etymology (cf. modern Turkish *porsuk* id.). In Russia, however, the word was associated with two Slavic words and interpreted as a compound of *bor* ‘forest’ and *súka* ‘bitch, female dog’. In this way, the Old Russian word *borsuk* was explained as meaning ‘forest dog’ (Iliadi 2005: 21).

- [4b] Partial deformation – Apart from unintended changes resulting from the incorrect hearing or recollection of a foreign or archaic word, some changes are intended to “validate” unintelligible or non-typical words. This group can be regarded as a type of folk etymology. Contamination (marked with ↔ here) is, more often than not, the method used to produce partial deformations. Their typical feature is that only part of the word becomes understandable, whereas the other part remains obscure. However, in some rare cases, the entire word becomes transparent because only one part of it was unclear and became clear after the alteration (Dutch *zondvloed* and German *Süntflut*). Examples include:

German *Alkoven* (m) ‘alcove’ < 17/18th century *Alcove* (f) id. (< French *alcôve* < Spanish *alcoba* < Arabic *al-qubba* ‘cupola’) ↔ Lower German *Koven* = High German *Koben* ‘a demarcated zone; piggery, hoghouse’ (Unger 2013: 39) || *al-* remains unclear.

Polish dialectal *kondoktór* ‘conductor (bus, train)’ < Polish literary *konduktor* id. ↔ dialectal *doktór* ‘doctor’ (Cienkowski 1972: 86) || *kun-* remains unclear.

Dutch *zondvloed* ‘deluge, flood’ and German *Süntflut* id. represent a special case. Their common original etymon is Old High German *sinvluot* (< *sin-* ‘all-, omni-’ + *vluot* ‘flood’), which was altered by unetymological *-t-* epenthesis into Middle High German *sintvluot*. This form was then borrowed into Dutch as *sintvloed* in the 16th century (de Vries 1967: 262). After the prefixal *sin(t)-* had become obsolete and, thus, incomprehensible, the word came to be associated with *Sünde* ‘sin’ in German and *zonde* id. in Dutch. The result was the emergence of new forms: German *Sündflut* and Dutch *zondvloed*. While in modern German the etymologically correct form *Süntflut* is recognized as the standard variant and *Sündflut* is accepted as a popular alternative, the folk-etymologically modified and etymologically incorrect form *zondvloed* has become the only literary variant in Dutch.

- [4c] Complete deformation – Here, the entire word becomes altered and its meaning appears perfectly transparent to users. Examples include:

- ① scholarly knowledge
- ② focus on linguistic features
- ③ diachronic/historical attitude
- ④ unchanged form of discussed words (= no deformation)
- ⑤ explanation of the original meaning

The following table shows the results:

Type of etymology	①	②	③	④	⑤
linguistic	+	+	+	+	+
perceptual	+	-	+/-	+	-
narrative	-	+	+/-	+	+
partially deformative	-	+	-	-	+/-
completely deformative	-	+	-	-	+

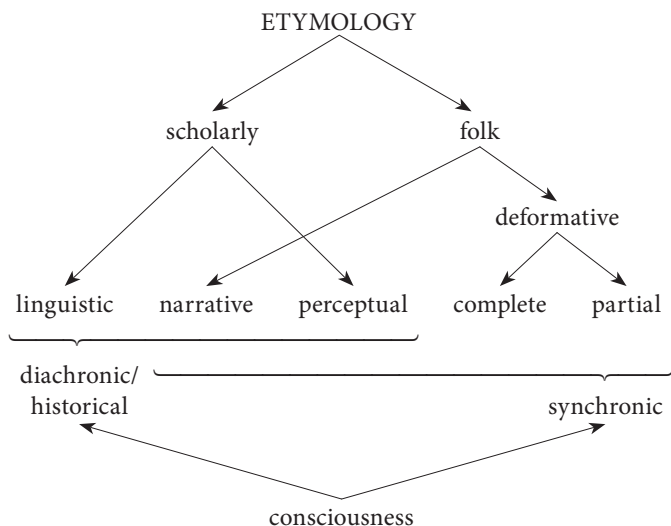
If a criterion of a non-linguistic etymology is fully concurrent with the same criterion of linguistic etymology, it receives one point; if it is partially concurrent, it receives half a point. The results are as follows:

linguistic:	5
perceptual:	2,5
narrative:	3,5
part.deform.:	1,5
compl.deform.:	2

It is interesting to see that narrative folk etymology is located closest to linguistic etymology. The only feature that separates them is the use (or non-use) of scholarly knowledge.

In the diagram below, an additional criterion has been added, viz., diachronic/historical or synchronic consciousness, as demonstrated in the results of the etymological explanations (cf. section 5):

deform the word; the partial type, by definition, explains only part of the word and often makes use of contamination; the complete type explains the entire word, frequently replacing foreign words entirely with native ones of similar sound.



References

- Altun H.O. 2021. So-called Italianisms and their etymological status in modern standard Turkish dictionaries. – *Occhiali* 8: 57–70.
- Arslan H.Ç. 2022. Algısal etimoloji ve yeni düşüncelerle Eski Türkçe *ımg*a ile *il ımg*a'nın kökeni. – *Türkologia* 4.112: 74–91.
- Cienkowski W. 1972. *Teoria etymologii ludowej*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Dapit R. 2021. Vicende de alcuni turchismi della gastronomia nella lingua slovena. – *Occhiali* 8: 31–56.
- Durkin Ph. 2009. *The Oxford guide to etymology*. Oxford: Oxford University Press.
- Futaky I. 2001. *Nyelvtörténeti vizsgálatok a Kárpát-medencei avar-magyar kapcsolatok kérdéséhez. Mongol és mandzsu-tunguz elemek nyelvünkben*. Budapest: Universitas Könyvkiadó.
- Iliadi A.I. 2005. *Osnovy slavjanskoj ètimologii*. Kiev: Dovira.
- Korkmaz Z. 2009. *Türkiye Türkçesi grameri. Şekil bilgisi*. Ankara: Türk Dil Kurumu.
- Liberman A. 2005. *Word origins... and how we know them. Etymology for everyone*. Oxford: Oxford University Press.
- Mailhammer R. (ed.) 2013. *Lexical and structural etymology. Beyond word histories*. Boston, Berlin: Walter de Gruyter.
- Olschansky H. 1996. *Volksetymologie*. Tübingen: Max Niemeyer.
- Papp I. 1956. Probleme des Imperativ-Zeichens im Ungarischen. – *Finnisch-Ugrische Forschungen* 32.1–2: 282–301.
- Parlar Z. 2025. Denizcilik dilinde yer alan İtalyanca sözcüklerin Türkçeye alıntılanma yolları ve halk etimolojileri. – Robbeets M., Akalın Ş.H. (eds.). *Geçmişten günümüze Türkçe dil ilişkileri*. Ankara: Erkmen Yayıncılık: 311–325.
- Stachowski K. 2024a. O pochodzeniu pol. *budyń*. – *Polonica* 43: 109–127.
- Stachowski K. 2024b. On the spread and evolution of *pudding*. – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 141: 117–137.
- Stachowski M. 2022. Perceptual etymology. A social aspect of etymological research. – *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 139.1: 61–67.

- Stachowski M., Woodhouse R. 2015. The etymology of *İstanbul*: Making optimal use of the evidence. – *Studia Etymologica Cracoviensia* 20: 221–245.
- Unger A. 2013. *Von Algebra bis Zucker. Arabische Wörter im Deutschen*. Stuttgart: Philipp Reclam jun.
- Urban M. 2008. *Secretary bird*, or how an etymological dictionary should be written and by whom. – *Studia Etymologica Cracoviensia* 13: 191–199.
- de Vries J. 1967. *Etymologisch woordenboek. Waar kommen onze woorden vandaan?* Utrecht: Aula-Boeken.
- Vučković M. 2017. Accommodation of Romani loan verbs in Czech and Serbo-Croatian. – Janyšková I., Karlíková H., Boček V. (eds.). *Etymological research into Czech*. Praha: Nakladatelství Lidové noviny: 441–454.
- Zaikovsky V. 2007. Ètimologija, ètnolingvistika i jazykovye kontakty slavjanskich s grečeskim, drugim balkanskimi i tjurkskimi jazykami. – Loma A. (ed.). *Slovenska etimologija danas / Etymologie slave aujourd'hui*. Beograd: Institut za srpski jezik SANU: 171–208.

