

Albanian Language in Light of Diverse Studies

Joana Lapa¹

¹ Department of Education, Faculty of Humanities, University "Ismael Qemali" ,Vlorë, Albania.
Corresponding author: joana.lapa@univlora.edu.al

Article Type: Review Article

Received: November 2025 | Accepted: December 2025 | Published: December 2025

Abstract

The Albanian language represents a unique branch within the Indo-European family, distinguished by its historical depth, linguistic autonomy, and ongoing evolution through contact with diverse cultures and languages. This study explores the Albanian language through multiple disciplinary perspectives, including historical linguistics, sociolinguistics, comparative philology, and digital humanities. By integrating empirical and theoretical analyses, the paper investigates phonological developments, lexical borrowings, dialectal variations, and the influence of globalization on linguistic identity. It also considers the role of technology and corpus-based research in the preservation and promotion of Albanian in academic and digital spaces. Particular emphasis is placed on the impact of globalization, migration, and multilingual environments on linguistic identity and language use. In addition, the study considers the growing role of technology, digital corpora, and computational tools in the documentation, preservation, and promotion of the Albanian language within academic and digital spaces. The findings highlight the importance of

interdisciplinary collaboration in understanding the complexity of Albanian as both a national symbol and a dynamic linguistic system shaped by history, modernization, for capturing the complexity of Albanian as both a national symbol and a living, dynamic linguistic system. Ultimately, the paper underscores the relevance of continued research in understanding how historical heritage and modern influences interact in shaping the present and future of the Albanian language.

Keywords: *linguistic diversity, sociolinguistics, historical linguistics, language contact, linguistic identity.*

Introduction

The Albanian language is an Indo-European language forming its own branch (Paçarizi, 2008) ¹.

Its study offers valuable insights into linguistic evolution, cultural identity, and language contact. Throughout history, Albanian has absorbed elements from Latin, Greek, Slavic, and Turkish, while maintaining its grammatical independence

Review

(Çoçka, 2013)¹. Modern research has expanded from classical etymology to sociolinguistics, computational linguistics, and corpus development (Koteet al.2019)¹.

“Language is the armory of the human mind, and at once contains the trophies of its past and the weapons of its future conquests”. This remarkable citation by Samuel

Taylor Coleridge² reminds me how important it is to reveal the unidentified elements of language itself, and particularly of a complex and ancient one. In this context, Albanian language has always fascinated my thoughts, not only as an Albanian born, but also in terms of its historical background and mystery. Albanian language is one of the oldest in the Balkan region, a very complex one with rich content, and with an interesting phonetic system. Many historians and authors in their studies have articulated their standpoints about the Albanian language in the past, revealing thus valuable data about the ancient origin of this language. The main elements of the Albanian language have been for a long time disputed among scholars, such is the case of the two distinguished authors in historic studies, (J. Wilkes and A. Stipcevic)³⁴, the significant contribution of whom is of a great importance to this matter. The intention of this paper is to discuss the different prospects that these two authors have regarding the Albanian language issue, putting emphasis on the origin, which branch this language belongs to and foreign language influences, reflecting thus the British and Balkan scholar’s theories.

Literature review

An important issue that (J. Wilkes) approaches in his recent book “The Illyrians” is the origin of the Albanian language as a distinct one in the Balkans. According to his work, the Albanian language is very old in terms of its formation, but it derives from many languages spoken by the ancient people living in the Balkans during the Bronze Age. Wilkes thinks that many important linguistic

elements, discovered from historical and archeological data, conclude that the Albanian language is inherited from the Illyrian, Messapian and Thracian languages. Therefore, supporting also a few British historians, he suggests that Albanian language overall, is a combination of the above languages. His crucial point in supporting this theory, is the vocabulary similarities that these languages share together as shown in these examples; Illyrian word (rhinos) – Albanian word (re) cloud, Messapian word (bilial) – Albanian word (bija) daughter and Thracian word (sica) – Albanian word (thika) knife.

In contrast to Wilkes, the Croatian historian Stipcevic, similarly to his Balkan colleagues, considers the origin of the Albanian language to be much older than the Bronze Age, emphasizing a Pre- Bronze Age formation of this language. Unlike Wilkes, Stipcevic stands for the theory that Albanian language is the direct descendant of the Illyrian language, representing thus a survival of an indigenous Illyrian language spoken in what is now Albania. In the same way, Stipcevic opposes Wilkes theory, emphasizing the grammatical, vocabulary and phonetic similarities between the two languages. In accordance with his theory, many Illyrian toponyms, hydronyms, and anthroponyms have been linked to Albanian, and so are translated only through Albanian language. Some of these similarities are shown by the following examples; Illyrian name (Bardhyllus), Illyrian town (Ulcin) and Illyrian tribe (Dardania) are respectively translated into Albanian language, (White Star), (Wolf’s town) and (Pear’s people). In other words, Stipcevic’s conclusion reveals an Albanian language inheritance only from the Illyrian language.

Another issue reflecting the differences between Wilkes and Stipcevic is the concern of the Albanian language’s existence as an Indo European language or not. Wilkes stresses that; Albanian language might not be in the Indo European group of languages due to its distinct and complex

¹ Çoçka, P. (2013). The Greek element in the first etymological dictionary of Albanian language. *Anglisticum Journal (IJLLIS)*, 2(4). Retrieved from <https://www.anglisticum.org.mk/index.php/IJLLIS/article/download/1312/1842/4783>

² *Biographia Literaria, or Biographical Sketches of My Literary Life and Opinion* 1817.

³ J. J. Wilkes – “The Illyrians” Penguin Books, 1992. (Autor: John J. Wilkes)

⁴ Aleksandar Stipčević – “The Illyrians: History and Culture” Noyes Press, 1977. “Iliri – povijest, život, kultura”, 1974)

Review

linguistic system, which differs from the other languages belonging to this group. In addition, he suggests that, since Albanian language stands alone in the Language Family Trees, it means that this language has evolved its own rules without Indo European influence. In brief, presenting this language as a Satem, which is another diverse language branch, Wilkes drew the conclusion that the Albanian language might not be an Indo European language.

Dr. Stipcevic, on the other hand, strongly insists that the Albanian language belongs to the Indo European group of languages. Unlike Wilkes, he defines Albanian as a Centum, a subgroup of the Indo European language, sustaining thus other Balkan historians such as F. Papazoglou, one of the notable supporters of this theory. Whereas, Wilkes stresses that due to the unique linguistic system Albanian is not an Indo European language, Stipcevic identifies the Albanian language to have similarities in terms of vocabulary and sounds with other Indo European languages such as Italian and Romanian.

In other words, (Stipcevic) as he has already described in his book "The Illyrians: History and Culture" supports the theory of the Albanian language as an Indo European language.

The next important aspect to consider is the influence of foreign languages toward Albanian language, in terms of vocabulary. The differences between Wilkes and Stipcevic about this subject are considerable. Wilkes, akin to some other British historians, believes that Albanian vocabulary was subject to many changes from foreign language influence in the Balkans. A further point in his studies is the classification of the foreign languages that influenced the Albanian language during the Dark Ages. He thinks that Greek, Latin and Slavic were the main languages that influenced Albanian. However, according to him, Greek and Roman civilization and culture played the most important role in this language impact.

In contrast to Wilkes, Stipcevic, affected also by other Balkan and German historians, considers the foreign language influence toward Albanian language in a low level. His fundamental point to this issue is that, due to the formidable

mountainous region and to the resilient tribal society, Albanians have preserved pretty well their language, avoiding thus assimilation and enormous foreign language influence. Unlike Wilkes, Stipcevic's prospect supports the theory of a greater Latin language influence rather than Greek, emphasizing that Roman rule for almost five centuries in Illyria (today Albania) was a determinant factor in this matter. In a few words, Stipcevic's theory suggests that foreign languages have influenced indeed Albanian language, but not in a great proportion as Wilkes assumes, stressing also that the Latin language has the greater role in this.

Overall, the arguments shown in this paper indicate the differences and maybe some minor similarities that scholars such as; Wilkes and Stipcevic have toward the Albanian language issue. Although, they did not share the same prospects about this subject, their work definitely is very valuable with the purpose of building the necessary framework to define all the important elements of the Albanian language. At the same time, I am pleased to know that, the general opinion of most European scholars shown by the recent studies is that the Albanian language is the inheritance of the Illyrian language, and that the Albanians are the direct descendants of the Illyrians.

Methodology

This paper follows a qualitative review methodology, combining document analysis, comparative linguistic analysis, and content synthesis. It reviews key historical grammars, etymological dictionaries, and recent computational models of Albanian. Sources include peer-reviewed journal articles, open-access conference papers, and validated linguistic corpora (Paçarizi, 2008; Çoçka, 2013; Kote et al., 2019; Abdurahmani, 2025), J. Wilkes and A. Stipcevic

This study employs a qualitative review methodology to examine the development, structure, and description of the Albanian language across historical and contemporary linguistic resources. The approach integrates document analysis, comparative linguistic

Review

analysis, and content synthesis, allowing the researcher to trace how grammatical rules, lexical evolution, and computational descriptions of Albanian have been treated in scholarly literature.

Results

The literature analysis produced four major findings:

1. Historical continuity and Indo-European roots: Evidence supports Albanian's position as the sole survivor of the ancient Illyrian branch (Paçarizi, 2008).
2. Dialectal dynamics: Geg and Tosk dialects remain the two dominant variants, with the 1972 Orthography Congress establishing Tosk-based standard Albanian (Paçarizi, 2008).
3. Sociolinguistic adaptation: Migration and globalization have created hybrid linguistic identities, where bilingualism influences vocabulary and syntax (Abdurahmani, 2025).
4. Technological integration: Recent NLP initiatives—such as the manually annotated Albanian corpus and neural lemmatizers (Kote et al., 2019)—illustrate growing efforts to digitize Albanian linguistic data.

Discussion

These findings show that the Albanian language is not only historically rich but also adaptable to modern challenges. The standardization process unified education and administration but also marginalized northern dialectal features. A balanced policy recognizing both variants would preserve linguistic diversity while ensuring national cohesion. The digital transformation of Albanian remains in early stages; limited annotated corpora restrict the development of translation and speech-recognition tools. Investment in NLP resources would expand the visibility of Albanian in global computational research (Piton & Lagji, 2010). Sociolinguistically, Albanian serves as a marker of identity in diaspora contexts, functioning as both a communicative and symbolic tool of heritage preservation (Abdurahmani, 2025).

Conclusions

Research demonstrates that Albanian linguistics is transitioning from traditional comparative methods toward interdisciplinary and digital paradigms. Future research should develop large-

scale dialectal corpora, integrate Albanian into international NLP frameworks, and encourage cross-border collaboration among Albanologists. Albanian, seen through diverse studies, represents both an ancient linguistic heritage and a modern evolving system ready to engage with contemporary science and technology.

The qualitative review demonstrates that integrating traditional linguistic scholarship with computational resources provides a richer, more precise understanding of the Albanian language. Although verification of some cited works remains uncertain, the combination of historical analysis and modern corpus-based tools highlights the evolution of Albanian studies and underscores the need for further digitally supported linguistic research.

References

1. Abdurahmani, E. T. (2025). *New Perspectives for Albanian Linguistic Field. International Journal for Research and Development (IJRD)*. Retrieved from <https://www.journal-uamd.org/index.php/IJRD/article/view/505>
2. Çoçka, P. (2013). *The Greek element in the first etymological dictionary of Albanian language. Anglisticum Journal (IJLLIS), 2(4)*. Retrieved from <https://www.anglisticum.org.mk/index.php/IJLLIS/article/download/1312/1842/4783>
3. Kote, N., Biba, M., Kanerva, J., Rönqvist, S., & Ginter, F. (2019). *Morphological tagging and lemmatization of Albanian: A manually annotated corpus and neural models*. arXiv preprint arXiv:1912.00991.
4. Paçarizi, R. (2008). *Albanian Language*. AAB University, Prishtina. Retrieved from <https://aab-edu.net/wp-content/uploads/2020/02/08-Rrahman-Pacarizi-english.pdf>
5. Piton, O., & Lagji, K. (2010). *Morphological study of Albanian words and processing with NooJ*. arXiv preprint arXiv:1002.0485.
6. J. J. Wilkes – “The Illyrians” Penguin Books, 1992. (Autor: John J. Wilkes)
7. Aleksandar Stipčević – “The Illyrians: History and Culture” Noyes Press, 1977. “Iliri – povijest, život, kultura”, 1974)
8. *Biographia Literaria, or Biographical Sketches of My Literary Life and Opinion 1817*.