

A computerized investigation of Albanian diachronic phonology

Clayton Marr

The Ohio State University

Computerized forward reconstruction, or CFR (Sims-Williams, 2018), offers an automatic and systematic means of testing hypotheses about the chronology of sound change in a language. While computing the effects of historical sound changes over millennia for thousands of etyma is laborious and extremely time-consuming, this task is accomplished within seconds by a CFR system such as DiaSim, which was created for not only evaluating hypothesized relative chronologies of sound changes, or “diachronic cascades”, but also “debugging them” by reporting statistics on how errors pattern (Marr and Mortensen, 2020). As a test case, past work applied this system to the phonological evolution of Latin into French, and a CFR-enabled “debugging” procedure improved accuracy from a 3.2% baseline for a cascade based on the 1934 received view to 84.9%. In the process, various proposals in the post-1934 literature on French were supported by the fact that they were independently produced as part of a systematic debugging process using DiaSim that was undertaken without reference to them (Marr and Mortensen, 2022), while the endeavor also may have revealed a new regular sound change in Old French, which was ultimately robustly supported by additional data (Marr, 2023b). However, as French boasts both a large corpus since medieval times and extensive past research, the experiment with French was more of a “laboratory run” to test the validity of the approach of debugging a language’s historical phonology via CFR, a prelude to bringing it into the field as an investigative technique.

This paper will bring in CFR to tackle Albanian diachronic phonology, starting with the Latin stratum of the its lexicon. Given the lack or loss of attestation of Albanian before the 15th century and its status as the only surviving member of its branch of Indo-European (Rusakov, 2018), reconstruction of Albanian diachronic phonology, and thus of Proto-Albanian, has always leaned heavily on the outcomes of strata of loanwords in Albanian from better-attested sources (Orel, 2000). Of these, the Latin layer (Çabej, 1962; Bonnet, 1998) is by far the most significant. Latin loanwords are more numerous than inheritance from Proto-Indo-European, Proto-Albanian is dated in relation to the time of contact with Latin, and Albanian diachronic phonology is in a large part an exercise in generalization from analyses of the outcomes of ancient Latin loans (Orel, 2000; Demiraj, 2006; Rusakov, 2017; De Vaan, 2018), though with significant contributions from Albanian historical dialectology (Curtis, 2018) and the other “layers”. Nevertheless, issues do remain that concern the Latin layer of Albanian, such as rival etymologies between imperial-era Latin loans and later Romance loans (Bonnet, 1998), and these have potential implications for the reconstruction of Proto-Albanian, and the greater mysteries of the language’s history within the Balkans (Friedman and Joseph, 2022). Thus, an evaluation and debugging of the received view on Albanian diachronic phonology as applied to its largest single pillar, the Latin stratum, offers both a new approach to an old but still vexing problem, and a step for CFR as an empirical method, between the curated “lab” case of French, and the “field” of understudied languages and language families.

This endeavor will apply DiaSim to CLEA, a dataset compiled in 2020–2022 and to be released with this paper, of 1007 Albanian etyma of ancient Latin origin as asserted by at least one of a set of reputed references (Bonnet, 1998; Orel, 1998, 2000; De Vaan, 2018; Topalli, 2017; Çabej, 1986), and will work from a base cascade representing the views of Orel (2000) and De Vaan (2018). The same debugging process as Marr and Mortensen (2022) will be applied, with accuracy reported for modern Albanian outcomes, and discussion of any systematic patterning of errors and possible solutions proposed.

Keywords: computerized forward reconstruction, diachronic phonology, Albanian, Latin

References

- Anamali, S. and Prifti, K. (2002). *Historia e popullit shqiptar*, volume 1. Botimet Toena. AShSh and IGJL (2006). *Fjalor i gjuhës shqipe*. Online at: <https://fjalorthi.com/Fjalorthi.com>. Bonnet, G. (1998). *Les mots latins de l'albanais*. L'Harmattan.
- Brüch, J. (1922). *Lateinische etymologien*. *Indogermanische Forschungen*, 40(1):196–247.
- Buflì, G. and Rocchi, L. (2021). *A historical-etymological dictionary of turkisms in albanian (1555-1954)*. Çabej, E. (1953). *Grupet nd, ng në gjuhën shqipe*. *Buletin për Shkencat Shoqërore*, 4.
- Çabej, E. (1962). *Zur Charakteristik der lateinischen Lehnwörter im Albanischen*. *Revue de Linguistique (București)*, 7:161–199.
- Camarda, D. (1864). *Saggio di grammatologia comparata sulla lingua albanese*. Livorno: Succesore di Egisto Vignozzi.
- Çabej, E. (1986). *Studime gjuhësore*, volume I–VIII. Prishtinë: Rilindja.
- Curtis, M. C. (2012). *Slavic-Albanian language contact, convergence, and coexistence*. The Ohio State University. Ph.D. dissertation.
- Curtis, M. C. (2018). 98. *The dialectology of Albanian*. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1800–1811. De Gruyter Mouton.
- Dashi, B. (2013). *Italianismi nella lingua albanese*. Edizioni Nuova Cultura.
- De Vaan, M. (2004). *PIE *e in Albanian*. *Die Sprache. Zeitschrift für Sprachwissenschaft*, 44(1):70–85.
- De Vaan, M. (2018). 95. *The phonology of Albanian*. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1732–1749. De Gruyter Mouton.
- Demiraj, B. (2018). 100. *The evolution of Albanian*. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1812–1815. De Gruyter Mouton.
- Demiraj, B. and Dayan, P. (1997). *Albanische Etymologien: Untersuchungen zum albanischen Erbwortschatz*, volume 134. Rodopi.
- Demiraj, S. (2004). *Gjuhësi Ballkanike*. Akademia e Shkencave e Republikës së Shqipërisë, Instituti i Gjuhësisë dhe i Letërsisë.
- Demiraj, S. (2006). *Albanian*. In *The Indo-European Languages*. Routledge.
- Ducellier, A. (1981). *La façade maritime de l'Albanie au moyen age, Durazzo et Valona du XIe au XVe siècle*. Greece, Thessaloniki: Institute for Balkan Studies.
- Fine, J. V. and Fine, J. V. A. (1994). *The late medieval Balkans: A critical survey from the late twelfth century to the Ottoman conquest*. University of Michigan Press.
- Friedman, V. A. and Joseph, B. D. (2022). *The Balkan Languages*. Cambridge University Press. Gjinari, J. (1989). *Dialektet e gjuhës shqipe*. Tiranë: Akademia e Shkencave e RPS të Shqipërisë. Haarmann, H. (1972). *Der lateinische Lehnwortschatz im Albanischen*. Hamburg: Helmut Buske.
- Haspelmath, M. (2009). *Lexical borrowing: Concepts and issues*. *Loanwords in the world's languages: A comparative handbook*, pages 35–54.
- Helbig, R. (1903). *Die italienischen Elemente im Albanesischen*. JA Barth.
- Huld, M. E. (1979). *An etymological glossary of selected Albanian items*. University of California, Los Angeles.
- Huld, M. E. (1986). *Accentual stratification of Ancient Greek loanwords in Albanian*. *Zeitschrift für vergleichende Sprachforschung*, 99(2. H):245–253.
- Hyllested, A. and Joseph, B. D. (2022). 13. *Albanian*. In Olander, T., editor, *The Indo-European Language Family*, pages 227–248. Cambridge University Press.
- Janda, R. D. and Joseph, B. D. (2003). *Reconsidering the canons of sound-change*. In *Historical Linguistics 2001: Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*, volume 237, page 205. John Benjamins Publishing.
- Jokl, N. (1984). *Sprachliche Beiträge zur Paläo-Ethnologie der Balkanhalbinsel:(zur Frage der ältesten griechisch-albanischen Beziehungen)*, volume 29. Austrian Academy of Sciences Press.

- Jorgaqi, K. (2001). Ndikimi i italishtes në letërsinë e vjetër shqipe: shek. XVI-XVII, Botimet Toena, Tiranë.
- Joseph, B. D. (2020). Language contact in the Balkans. *The handbook of language contact*, pages 537–549.
- Kore, M. K. (2013). Similarities between Albanian and Romanian in the entire language sub- systems. *Mediterranean Journal of Social Sciences*, 4(2):175–175.
- Lafe, G. (2000). Der italienische Einfluß auf das Albanische. Zweiter Teil. Wörterbuch der Italianismen im Albanischen, in *Ponto-Baltica*, 10:31–120.
- Landi, A. (1989). *Gli elementi latini nella lingua albanese*. Napoli: Edizioni Scientifiche Italiane.
- Marr, C. (2023a). The angevin–albanian element in the albanian lexicon.
- Marr, C. (2023b). A regular velar onset voicing rule for 12th century French. Submitted for publication.
- Marr, C. and Mortensen, D. R. (2020). Computerized forward reconstruction for analysis in diachronic phonology, and Latin to French reflex prediction. In *Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 28–36.
- Marr, C. and Mortensen, D. R. (2022). Large-scale computerized forward reconstruction yields new perspectives in French diachronic phonology. *Diachronica*.
- Matzinger, J. (2018). 98. The lexicon of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1788–1800. De Gruyter Mouton.
- Mazaudon, M. and Lowe, J. B. (1993). Regularity and exceptions in sound change. In *Annual Conference of the Linguistic Society of Belgium*.
- Meyer, G. (1891). *Etymologisches Wörterbuch der albanesischen Sprache*. Strasbourg: Trübner.
- Meyer-Lübke, W. (1935). *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.
- Miklosich, F. (1870). *Albanische Forschungen: I. Die slavischen Elemente im Albanischen*. Wien: Karl Gerold's Sohn; Denkschr. Akad. Wien XIX.
- Miklosich, F. (1871). *Albanische Forschungen: II. Die romanischen Elemente im Albanischen*. Wien: Karl Gerold's Sohn; Denkschr. Akad. Wien XIX.
- Newmark, L. (1998). *Albanian-English Dictionary*. Oxford University Press, USA. Online at: [http://www.seelrc.org:8080/albdict/Online Albanian Dictionary](http://www.seelrc.org:8080/albdict/Online%20Albanian%20Dictionary).
- Niedermann, M. (1905). *Contributions à la critique et à l'explication des gloses latines, volume 1*. Attinger.
- Ölberg, H. M. (1972). *Griechisch-albanische Sprachbeziehungen*. Institut für vergleichende Sprachwissenschaft der Universität.
- Orel, V. (1998). *Albanian etymological dictionary*. Brill.
- Orel, V. È. (2000). *A concise historical grammar of the Albanian language: reconstruction of Proto-Albanian*. Brill.
- Pokorny, J. (1959). *Indogermanisches etymologisches Wörterbuch, volume 1*. Bern, München: Francke Verlag.
- Prifti, E. (2012). *Aromanian elements of an Albanian argot*.
- Rusakov, A. (2017). Albanian. *The Indo-European languages*, pages 552–608.
- Rusakov, A. (2018). 94. The documentation of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1716–1731. De Gruyter Mouton.
- Schuhardt, H. E. M. (1866–8). *Der Vokalismus des Vulgärlateins I-III*. Leipzig: B. G. Teubner.
- Seliščev, A. M. and Olesch, R. (1931). *Slavjanskoe naselenie v Albanii*. Böhlau.
- Sims-Williams, P. (2018). Mechanising historical phonology. *Transactions of the Philological Society*, 116(3):555–573.
- Svane, G. (1992). *Slavische Lehnwörter im Albanischen, volume 67*. Aarhus Universitetsforlag.
- Thumb, A. (1909). *Altgriechische elemente des albanischen*. *Indogermanische Forschungen*, 26:1–20.
- Topalli, K. (2017). *Fjalor etimologjik i gjuhës shqipe*. Qendra e Studimeve Albanologjike, Instituti i Gjuhësisë dhe i Letërsisë.

Uhlisch, G. (1964). Neugriechische Lehnwörter im Albanischen. Typescript: Berlin.

Vătășescu, C. I. (1997). Vocabularul de origine latină din limba albaneză în comparație cu româna. PhD thesis, Editura Universității din București.

Vicario, F. (1993). L'influsso, lessicale veneto in albanese. *Balkan-Archiv (NF)*, 17(18):1992. Ylli, X. (2000). Das slavische Lehngut im Albanischen: Teil 2: Ortsnamen. Peter Lang International Academic Publishers.