Prehistoric loanwords in Armenian: Hurro-Urartian, Kartvelian, and the unclassified substrate

Nielsen, R.T.

Introduction 1

The prehistory of the Armenian language is a topic with a long research history. Its modern phase began with the discovery, made by Hübschmann (1875–1877), that Armenian constitutes a separate branch of the Indo-European languages and not, as previously assumed, an Iranian language. This breakthrough led to another recognition about the Armenian language, namely that its inherited, Indo-European core had undergone dramatic changes over the course of history, in terms of both phonology, morphology, and lexicon. In this process of language change, the contact between Armenian and various neighbouring languages undeniably played an important role. With regard to lexical replacement, the introduction of Middle Iranian loanwords, during the height of Persian political and cultural influence on Armenian, is probably the most striking and well-known event. It was the main reason for erroneously grouping Armenian with the Iranian languages in the first place. After the first sound laws of Armenian had been discovered, and the inherited part of the lexicon had been identified, two fundamental tasks presented themselves. First, explaining the development of the Indo-European inherited elements and establishing the place of Armenian within the family tree. Second, categorizing and stratifying the various layers of foreign influence upon the language,

pointing to possible geographic contact zones, where Armenian would have been spoken at earlier points in time.

All in all, the Armenian preservation of inherited vocabulary, compared to most other Indo-European languages, is scarce. A considerable share of the non-inherited vocabulary is made up of loans from attested languages. The three main, identifiable sources of loanwords are considered to be the Middle Iranian languages (especially Parthian), Greek, and Syriac (Meillet 1936b: 143, Clackson 2017). Yet, more than half of the Armenian lexicon remains etymologically unexplained. Ačaŕyan's *Hayerēn armatakan baŕaran* (HAB, 1926–35), a monumental etymological dictionary, which retains much of its relevance to this day, is still the only near-complete historical treatment of the Armenian lexicon. As shown below, its 10,722 lemmata can be divided into four categories, following Solta (1990: 13), on the basis of Ačaŕyan's conclusions. Although some of these conclusions have since been revised, the picture broadly remains the same. The category "Loanwords" covers Iranian, Greek and Syriac words, but also occasionally words considered to be from Phrygian, "Caucasian" languages and unattested Mediterranean languages. The category "Other" covers what Ačaŕyan considers onomatopoeic words, as well as ghost words.

Inherited	713 (≈ 7 %)
Other	421 (≈ 4 %)
Loanwords	4016 (≈ 38 %)
Unknown	5572 (≈ 52 %)

While the field of Armenian etymology has seen substantial progress since Ačaŕyan, there has been no considerable diminishment of words with completely unknown origin. Within the Indo-European family, the amount of identified, inherited lexicon is thus comparable to that of Albanian; yet Armenian is attested from fifth century CE, a full millennium before the appearance of the first Albanian text in 1555. This naturally leads to challenges for researchers who wish to reconstruct earlier stages in the development of Armenian (conventionally termed 'Proto-Armenian'), and to reconstruct the prehistory of Armenian speakers as a whole. As emphasized by Hübschmann's pioneering work, the separation of loanwords from potentially inherited lexemes is a crucial prerequisite to this task. It is hardly a coincidence that the first

four chapters of his *Armenische Grammatik* (1897) are dedicated to loanwords from Persian, Syriac, Greek, and words of uncertain origin, before dealing with the inherited vocabulary only in the fifth and final chapter.

Since Hübschmann's time, Indo-Europeanist research on the Armenian lexicon has been preoccupied mainly with the inherited element. Important early contributions include those of Bugge (1893), Meillet (1894), and Lidén (1906). Nevertheless, several smaller, prehistoric loanword layers have also received interest. The historical location of the Armenian language, the highlands of the southern Caucasus and eastern Anatolia, invites a number of potential candidates for known languages that could have influenced it during prehistory. To the north, Armenian is bordered by the Kartvelian and Nakh-Daghestanian languages; to the south, the Semitic languages; and to the east, the Indo-European Iranian languages. Before the Phrygian and later Greek incursions into Anatolia, Armenian was bordered to the south and west mainly by Anatolian IE languages like Hittite and Luwian, by Hattic, as well as several unattested languages. After the collapse of the Hittite kingdom at the end of the second and beginning of the first millenium BCE, part of this territory became Phrygian-speaking. According to ancient Greek historians, the Phrygians are considered to have entered Anatolia from the Balkans, but this narrative cannot be confirmed by other sources (Wittke 2004).

An important event in the prehistory of Armenian is the emergence of the Urartian Kingdom, centered around Lake Van. Urartu established its power at the beginning of the first millennium BCE, roughly at the same time as Anatolia was settled by Phrygians, and lasted until the sixth century BCE. As such, Urartian is presumably the latest known non-Indo-European language spoken in the Armenian Highlands before the Armenians (or at least their exonym) enter written history. This occurs with the attestation of the exonyms *Armina-* and *Arminiya-* in the Old Persian Behistun-inscription of the late sixth century BCE (see Schmitt 2008).[1] The Armenian language itself, and the Armenian self-appellation *Hay-*,

[1] In many accounts, the Behistun-inscription is considered the earliest documentation of the Armenians as such, but it cannot be theoretically excluded that the satrapy *Armina* at this time designated an area west of Persia with a different population than historical Armenia. In any case, the inscription in itself does not tell us much, if anything, about the linguistic situation at this point in time.

would remain unwritten for another millenium (see EDA 382–5 and Petrosyan 2018 with literature for a discussion of the name *Hay*). The Urartian language, whose first inscriptions were discovered in the early nineteenth century, is still relatively poorly understood (cf. Zimansky 2011: 550–53). It is not demonstrably related to any other language, apart from Hurrian, which is documented in the Mitanni kingdom of mid-second millennium BCE, and was an important minority language in the Hittite empire. Urartian is the most poorly attested of the two languages, but more important to the study of Armenian due to its attestation in historical Armenia. Together, these two languages make up the otherwise isolated and extinct Hurro-Urartian family, which is usually conjectured to have split up only around 2000 BCE (Wilhelm 2008: 105). Attempts to connect Hurro-Urartian with other families, most prominently East Caucasian (Trombetti 1923, Diakonoff 1967, Diakonoff & Starostin 1986), have not led to widespread acceptance (see Smeets 1989). It thus remains unclear where Proto-Hurro-Urartian originates. The area can reasonably be limited to eastern Anatolia, northern Mesopotamia and the Caucasus region.

Given the vicinity of the Kartvelian languages to the north, it is perhaps surprising that loanwords from Georgian or Laz are usually considered to be relatively few. On the other hand, several scholars have noted the striking typological similarity between Armenian and Kartvelian languages, especially in terms of phonology (see especially Gippert 2005). This may indicate a language shift towards Armenian in the past. That is, a number of originally Kartvelian speakers had adopted the Armenian language, and in the process carried over traits from their own language. Deeters (1926, 1927) was among the first advocates of this theory. He drew attention to the virtually identical consonant inventories of Georgian and Armenian, the fixed accent in Armenian and especially Laz, and the Armenian simplification of the PIE diphthongs, which are absent on the phonemic level in Kartvelian. As for morphological and syntactic innovations, Deeters considered the lack of grammatical gender (also absent in Hurro-Urartian) and the emergence of constructions with past participle and a genitive subject, the *genetivus auctoris* (cf. Jensen 1959: 134–35). For each of these features it is difficult to exclude that they are not the result centuries of bilingualism, i.e. a *Sprachbund* effect, rather than the result of language shift. It may be noted, however, that a strong diffusion of phonological and

syntactic features without the transfer of many lexical items can be considered a reliable predictor of a relatively sudden language shift, without a long period of bilingualism in the two language communities (Thomason 2010: 35–39).

Finally, some researchers have been preoccupied with lexical traces of language contact between Armenian and one or more unknown languages, conventionally termed 'substrate languages', which also left traces in other Indo-European languages. An important reason for studying these loanwords is that they, by virtue of their particular geographical spread, can be assumed to have been adopted relatively soon after the dissolution of the Indo-European languages, while Proto-Armenian was still spoken in the relative vicinity of other Indo-European daughter languages. Another reason is that the line of research concerned with these unclassified contact languages is still relatively young, in particular with respect to its impact on Armenian. The study of substrate words can thus contribute not only to our knowledge about the prehistory of the Armenian language, but also to developing the methodology of identifying and analyzing substrate words.

1.1 Aim and method of this work

The present dissertation has two fundamental aims. The first aim is to critically evaluate and delimit three prehistoric layers of loanwords in Armenian: those from Hurro-Urartian, Kartvelian, and the shared substrate of Indo-European. Each of these layers represent distinct linguistic contact events in the prehistory of the language. This task is carried out by critically reviewing proposals for loanwords within these strata and adducing new material where possible. The other aim of this study is to determine to what extent these three, delimited loanword corpora can inform our knowledge about the prehistory of Armenian speaking populations, in particular their movements and the timing of those.

As mentioned above, Hurro-Urartian and Kartvelian represent two of the important contact languages that predate the Iranization of the Armenian-speaking area and could, for that reason, be termed 'prehistoric'. The second and third chapters of the dissertation proceeds to delimit the evidence for Armenian contact with these languages, and to estimate the timing and dating of

these prehistoric contact events. Followingly, the fourth and largest chapter deals with the lexical traces of language contact between Armenian and one or more unknown languages.

Other layers of loanwords remain outside the scope of this work. Loanwords from Anatolian languages are discussed in detail by Martirosyan (2017) and Simon (2013, 2021a). This work would not be able to make significant contributions to this research problem at the time being. Another complex question, likewise outside the scope of this work, concerns the potential loanwords from Semitic languages other than Syriac, including Akkadian. On this topic, see J̌ahowkyan 1982, Diakonoff 1982, and Greppin 1989 with references; note also the rather idiosyncratic monograph by Mkrtčʿyan (2005). Finally, the three major loanword strata – Iranian, Syriac and Greek – are of a size that calls for separate, dedicated works.[2]

The principal sources of lexical material in this work include the two-volume *Nor baġirkʿ haykazean lezowi* (NBHL, 1836) and Ačaṙyan's HAB (second edition, 1971–79). Where necessary, these sources are supplemented by Greppin 1978, a philological study of bird names, and Łazaryan 1981, a dictionary of plant names. English glosses are generally taken from Petrosean 1875 or based on my own translations of the glosses in NBHL and HAB. For dialectal forms, the primary source is *Hayocʿ lezvi barbaṙayin baṙaran* (HLBB) as well as Ačaṙyan 1913. More recent etymological works, especially that of Martirosyan (EDA), are primary sources of existing etymolgical proposals and have also been consulted for additional information about dialectal forms and corrections of forms and meanings.

[2] The fundamental work on these strata remains, of course, Hübschmann 1897, which has still not been fully superceded in its scope (notwithstanding etymological dictionaries proper). For the discussion of Iranian loanwords, cf. Bolognesi 1960, with important additions by Schmitt 1983. On Greek loanwords, see Thumb 1900 and Kölligan 2020c. On Syriac loanwords, see Morani 2011 and Kitazumi & Rudolf 2021. Helpful overviews of all three strata, as well as many new contributions, are offered by Olsen (1999), including a section on loanwords from other languages, as well as words with no etymology.