Detecting word boundaries in an undeciphered script: The Byblos syllabary

The Byblos writing system (ca. 1500 BC) comprises 15 inscriptions, with the largest one containing 461 clearly discernible signs. Some further 28 unassigned fragments found in the Levante and in Italy are tentatively assigned to the Byblos corpus too. We have created a Unicode letter for each sign variant found in these inscriptions and transcribed the corpus. Feeding the corpus into our RegEx- based deciphering tool, we established a preliminary (dynamic) syllabary laying a foundation for computer-assisted deciphering efforts of the Byblos script. A first sequence analysis allows to detect morphemes and word boundaries, so that even without knowing the sound values, we gain some insight into the syntax of the language depicted by these beautiful graphemes.