It has often been noted that word order type might correlate with type of morphological marking of core arguments cross-linguistically. Word order type here means the dominant arrangement of the main clausal elements (S, O, and V) in a language. Core argument marking is about discriminating the arguments of a prototypical two-place transitive predicate, one more agent-like (S) and the other more patient-like (O) (Comrie 2005). Two types of morphological marking, head marking (or cross-reference) and dependent marking (or case), interact in this domain. While earlier studies (e.g. Nichols 1992) have investigated the correlation of word order and type of overt morphological marking, correlation of word order and absence of overt morphological marking, here called zero marking, has not been fully studied. Zero-marking languages [1-2] are generally expected to have dominant SVO order, which is purportedly the most economic word order available for argument discrimination. Regardless of SVO being motivated in zero-marking languages, several counterexamples exist around the globe [3-5]. West Africa and Southeast Asia are further well-known hotbeds of zero-marking SVO languages, but it is unclear whether zero marking correlates with SVO outside these areas. If not, the proposed correlation could be a mere consequence of areal tendencies, not of universal principles.

In my presentation, I discuss the results of modeling the distribution of zero marking over word order, areas and language families by multiple logistic regression in the spirit of Bickel (2008). Logistic regression tests to what extent the independent variables (here word order and areas) can predict the outcome of the dependent variable (here zero vs. overt marking). Logistic regression enables modeling e.g. the effect of areas and test whether a hypothesized association between two structural variables (zero marking and word order) is independent of them. The effect of each variable of interest was tested with maximum likelihood estimation, comparing a model with the variable of interest to one without it; p-values were deduced by randomization. Different areal resolutions were used to test the effect of areas. The effect of genealogical affiliation was tested by assessing diachronic intra-family tendencies with the Skewed Family Method (Bickel 2008), which estimates whether languages in genealogical units have maintained/developed the same linguistic property to a statistically significant degree. Data comes from a sample of ca. 800 languages, which aims at being genealogically and areally maximally representative. Based on the tests, zero marking is argued to correlate with SVO independently of areas and genealogical affiliation despite the distribution of zero marking being areally and genealogically skewed. The method allows universals, such as the proposed correlation, to be interpreted dynamically as universal structural pressure on how languages change over time.
Examples:

(1) Thai (Kam-Tai; Thailand; Iwasaki and Ingkaphirom 2005: 110)
Lék té nɔ́ɔy
Lek kick Noy
S V O
'Leek kicks Noy.'

(2) Iquito (Zaparoan; Peru; Brown 2004: 13)
Ja Marcos jihuǐta-qui nu-canisi
Already Marcos find-COMPL 3SG-bag
S V O
'Marcos has already found his bag.'

(3) Kolokuma Ijo (Ijoid; Nigeria; Williamson 1965: 58)
ɪɡɛɛrɛ erɛ-ma kɔɔrɛ-mi
crocodile woman-DEF catch-PST
S O V
'A crocodile caught the woman.'

(4) Arará Karó (Tupian; Brazil; Gabas 1999: 153)
Iyõm wat awe cape-t
father 1SG.POSS brother beat-IND
S O V
'Father beat my brother.'

(5) Quiegolani Zapotec (Zapotecan; Mexico; Black 2000: 45)
W-ɛɛy Benit mɛɛl
COMPL-take Benito fish
V S O
'Benito took a fish.'

Abbreviations: 1 first person, 3 third person, COMPL completive, DEF definite, IND indicative, O object, POSS possessive, PST past tense, S subject, SG singular, V verb.

References