

Typology I: Homework for Lecture 11, 2016/07/06

(The questions marked with (*) are research questions you can use to deepen your understanding, the others could be exam questions.)

1. Assume only the language families and languages in the table below existed. For the languages in bold face there is a description of the typological feature (f), whereas for the others there is no information available. The value of feature f taken by each language is indicated as f1, f2, etc. Assume we engage in a typological study about feature f and need to sample from these families and languages. Name a) the *biggest convenience sample* possible given these families and languages, b) a *phylogenetically balanced convenience sample*, c) a *phylogenetically balanced variety sample*.

INDO-EUROPEAN	AUSTRONESIAN	AFRO-ASIATIC	ESKIMO-ALEUT
French _{f1}	Maori _{f3}	Hebrew _{f1}	Inuktitut
English _{f1}	Tagalog	Oromo	Yupik _{f5}
Welsh _{f1}	Rarotongan _{f4}	Amharic _{f2}	
Pashto _{f2}			
Ladino			

2. Use the Venn diagram on the lecture slides in combination with the data on frequencies of word orders (WALS chapter 81A) to fill in the actual numbers of languages that should be in the different areas/sets representing “Set of all languages”, “Set Y”, “Y and Z”, “Set Z”. Note that you have to put in the numbers of languages that are *only* in a specific set and not in the others. Tipp: start with the set that is overlapping with all other sets, i.e. “ $Y \cap Z$ ”.
3. Take the following two passages as “corpora” representing the English language:

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do. Once or twice she had peeped into the book her sister was reading [...] Alice in Wonderland

Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin [...]

Universal Declaration of Human Rights, Article 2

Calculate the *synthesis index*, *inflectional index* and *derivational index* after Greenberg for both passages. What are striking differences/similarities in the indexes, and what do we learn from those with regards to a corpus oriented typology?