Typology I: Solution to Homework for Lecture 2

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

1. **A Hungarian friend has seen an internet video proving that Hungarian is not related to Finnish at all, but descended from Sumerian. The following list of word equations is given in the video to prove that mainstream Uralistics is nothing but a conspiracy. Explain to him/her why the data are problematic, and why a linguist will not accept this sort of argument.**

Gloss	Sumerian	Hungarian	Finnish
"to hear"	hal	hall	kuulla
"bridge"	id	híd	silta
"moon"	húl	hold	kuu
"long"	uš	hosszú	pitkä
"cool"	sid	hűs	kylmä
"jealous"	erim	irigy	mustasukkainen
"horror"	ušum	iszony	kauhu
"palm"	tibit	tenyer	kämmen
"lake"	túl	tó	järvi
"wild"	bad	vad	villi
"to cut"	ag	vág	leikata

   Apart from the issue that in such a short word list, the words are likely to be specifically selected for the ideological purpose behind the claim, the main problem is that the word list does not demonstrate any regular sound correspondences.
   Consider the first letters of all the words in the list. While in the two examples given, Hungarian t- does seem to correspond to Sumerian t-, v- is apparently either b- or -, h- is either h-, s-, or -, and i- is either e- or u-. The words with t- remain as the only candidate cognates. But consider the pair *tibit* and *tenyer*. Apparently, Sumerian -b- is here claimed to correspond to Hungarian -ny-. But in the pair *ušum* and *isszony*, -m seems to be the counterpart to -ny. By and large, the given data do not allow us to establish regular sound correspondences. Given the high probability of finding words that are similar due to chance, the data therefore do not demonstrate anything about the relationships between any of the three languages involved.

2. **Here is some further data about parts of the High German Consonant shift. Based on these data, determine regular sound correspondences between the languages, and build a hypothesis about the historical development. Finally, using your knowledge of phonology, try to describe the observed phenomenon as concisely as possible.**

Gloss	Dutch	High German	Swiss German
“apple”	[ˈʔapəl]	[ˈʔap͡fl]	[ˈʔœp͡fu]
“cat”	[kaːt]	[ˈkat͡sə]	[k͡xɑt͡s]
“path”	[pɑt]	[p͡faːt]	[p͡fɑːt]
“plough”	[pluːχ]	[p͡fluːk]	[p͡flyːk]
“two”	[tveː]	[t͡svaɪ]	[t͡svɛɪ]
“to come”	[ˈkoːmə]	[ˈkɔmn]	[k͡xo]

There are three obvious regular correspondences which can be taken to suggest regular sound changes.

- Dutch and (and English) [p] = High German and Swiss German [p͡f]
- Dutch and (and English) [t] corresponds to High German and Swiss German [t͡s], with the exception of the final -t in the words for “path”
- Dutch and High German (and English) [k] = Swiss German [k͡x]

Note that the final sound in the words for “plough” shows a different pattern. (Underlyingly, the last segment is a [g] phoneme, which regularly becomes [χ] in Dutch, and undergoes final devoicing in German. Final devoicing also explains the exceptional final -t in the words for “path”).

For the three observed sound correspondences, we can easily find a common description. In all cases, a voiceless stop at a given place of articulation corresponds to an affricate, i.e. the stop gets a fricative release at the same place of articulation.

The question now is in which direction the historical development went. Does Swiss German represent the original situation, and High German and Dutch underwent a process of deaffricatization (incomplete in the High German case)? Or is Dutch the original, and there is affrication in High German and Swiss German? To decide this question, we would need to consider different languages, or take historical information into account. Given our knowledge of other closely related languages like English, which also have simple plosives in the relevant positions, the affrication hypotheses seems more plausible, since the affricates seem to be a local phenomenon of Southern Germany and Switzerland. We conclude that Standard High German represents a state in the middle between the original situation (only plosives) and complete affrication (only affricates, as in Swiss German), as the sound shift has only applied at two places of articulation.

(The true story is, of course, much more complex. The examples were chosen carefully to hide the fact that the shifts actually only occurred under more complex conditions. Again simplifying slightly, there are three contexts in which affrication occurred: in onsets (*path* vs. *Pfad*), as part of geminate stops (*copper* vs. *Kupfer*), and after sonorants (*heart* vs. *Herz*). Moreover, the systematic correspondences were blurred later by various loans between German dialects at different stages of the shift, as well as barely recognizable loanwords from Romance languages which did not undergo the shift.)

3. **Imagine that you have reconstructed the proto-languages of all language families. Treat these reconstructions as languages, and reapply the comparative methods to reconstruct their common ancestors. Moving back into pre-history by applying this method again and again, it should be possible to move up in the language tree until you arrive at the common proto-language of humankind. Why do most linguists believe that it is impossible to reconstruct Proto-World in this way?**

   The primary problem is of course that the few similarities we will still find at such a time depth can just as easily be explained by chance. But if we can securely establish proto-languages based on reconstructed sound laws, wouldn't this in fact reduce the time depth we need to bridge? The reason why most linguists do not accept this argument is that in the common view, the antecedent of this argument does not hold: we cannot securely establish proto-languages!

   In a sense, a proto-language is just an abstraction over available data about a language family. If e.g. Ancient Greek had disappeared without a trace, the reconstruction of PIE would certainly look different today. Even if no additional data becomes available, the most accepted reconstruction tends to shift over the decades as the data are reanalyzed. In many cases, different researchers will come up with very different competing reconstructions. If we try to find a compromise between all plausible variants of a proto-language, this consensus will be small and far from a complete description of a language. The language might contain only a few dozens of words, the sound system might be underspecified (e.g. unknown vowels), or the meaning of many words might remain unclear. Such a reconstruction still fulfills the purpose of establishing the language family, but it cannot be treated as a full language which could be used as a basis for further reconstruction.

4. **Explain in your own words what a Swadesh list is, and what it is used for. Why are these lists typically so short? Name examples of possible problems you might encounter if you want to build a large Swadesh-type list that can be applied across cultures and climate zones.**

   A Swadesh list is a list of universal basic concepts that are taken to be stable against borrowing and semantic shifts. The classical Swadesh list (1952) lists 215 meanings, mostly containing kinship terms, body parts, natural phenomena, and basic actions. Parallel Swadesh lists for many languages are used in many branches of computational historical linguistics such as lexicostatistics (the automated detection of genealogical relationships based on lexical similarities) and glottochronology (the estimation of dates of language divergence).

   The list is so short because it is surprisingly difficult to find many concepts which are valid across all cultures, and can therefore be expected to be lexicalized in every language. Reasons include:

- varying levels of technology (the rope is the only universal tool!)
- very impoverished flora in some climate zones
  (virtually none in deserts, only moss and lichen in the tundra)
- the fauna vastly differs between continents
  (e.g. no higher mammals in Australia, no carnivores in Madagascar except some which only occur there)
- no higher numbers in many languages
  (at least "two" and "three" exist in virtually all of them)
- not many color terms (only "dark" and "bright" are universal)
- huge differences in kinship systems (highly culture-dependent; not even the concept "father" is universal!)

This reduces the possible sources for universal concepts to body parts (except those heavily influenced by taboo), very basic grammatical elements (pronouns, inflectional endings, negation), as well as globally occurring natural phenomena and very basic actions (for which no tools are need). Not surprisingly, these semantic fields are exactly the ones which are dominating Swadesh-type lists.

5. **For each of the following concepts, decide whether they are good candidates for a Swadesh list. Explain your decisions using the criteria discussed in the lecture.**

   - **bread**: **OK**; a basic food item in many parts of the world since the beginning of agriculture; but what about hunter-gatherer cultures? what if there is no generic word for bread, but many more specific names? what if grain is consumed in other forms?
   - **louse**: **great**; universal (virtually the only animal that occurs everywhere in the world)
   - **interesting**: **good**; a universal mental state, technology-independent, but often expressed by a compound (not basic)
   - **to snore**: **OK**; universal and culture-independent, but prone to onomatopoesia
   - **snow**: **OK**; words for natural phenomena tend to be very stable, but there are climates without snow
   - **steel**: **bad**; late Iron Age technology in most parts of the world, and an important trade good; likely to be a Wanderwort; most of the world's cultures did not develop this level of technology

6. **(*) Pick ten concepts from the Swadesh list given in the lecture, and choose three languages (e.g. your native language, a closely related language, and an unrelated language from another continent). Write down (or look up) the translations of your ten concepts in your three languages.**

Gloss	German	Swedish	Turkish
“two”	zwei	två	iki
“three”	drei	tre	üç
“big”	groß	stor	büyük
“small”	klein	liten	küçük
“to come”	kommen	komma	gelmek
“to drink”	trinken	dricka	içmek
“woman”	Frau	kvinna	kadın
“stone”	Stein	sten	tas
“fire”	Feuer	eld	ateş
“tree”	Baum	träd	ağaç

- **Count the number of similar words for each language pair.** Let us pretend we are a computer. Assuming that the first position in a word is the most stable (and which sounds are similar to each other), we could arrive at the following result:

		Count	Similar words
German	Swedish	5	zwei/två, drei/tre, kommen/komma, trinken/dricka, Stein/sten
German	Turkish	1	klein/küçük
Swedish	Turkish	2	kvinna/kadın, ateş/eld

- **Does the number of similar words reflect the established genealogical relationships?**
  The relationship between German and Swedish is slightly less strong than one would probably expect, and the number of similar words between Swedish and Turkish is surprisingly large. By and large, however, the method correctly determined German and Swedish to be related, and Turkish to be unrelated.
- **In case you know enough about the history of your languages: are the similar words true cognates?**
  All the German-Swedish cognate pairs are true cognates, whereas klein/küçük and kvinna/kadın are not. The most interesting case is ateş/eld. The Turkish word is a loanword from Persian which goes back to PIE *at(e)r- “fire”, while the Swedish word ultimately comes from PIE *aidh- “to burn”. The two PIE roots, although semantically similar, do not appear to have a common origin. Still, the fact that the word was borrowed into Turkish already makes this an incorrect cognate pair.
- **Are there any similar words between unrelated languages? Is there any plausible explanation?**
  The explanation for klein/küçük and kvinna/kadın is pure chance, and perhaps a slight bias in my selection of the Swadesh concepts caused by the necessity to find a nice example. For ateş/eld, the question boils down to asking why the two PIE roots involved are similar, and it is well possible that they are deeply related at a timedepth greater than what we can attain using the comparative method. It is therefore a good example to see why some people believe that lexicostatistical methods can detect long-distance relationships.