Skype with Vasilisa 28 June 2017:

discussed the table with frequency differences

“weirdness” coefficient is a common corpus-linguistic coefficient, it’s simply the frequency of morpheme in one corpus divided by that in the other; normally one takes one as a reference corpus, and V. here took the Sebjan corpus as the reference, because it’s a bit bigger; this coefficient doesn’t take into account the size of the corpora

the log likelihood is a bit like a Chi-square value, it takes into account the expected frequency and the size of the corpora; an online site where one can calculate it and with a bit of explanation: <http://ucrel.lancs.ac.uk/llwizard.html>

So the log likelihood is the ‘better’ value; and values >15 are considered notable – but that’s for the normal huge corpora of standard langauges

Several things stand out in the table:

1) some differences in suffix frequencies are probably spurious and simply linked to other features (e.g. it’s clear that the higher frequency in SEB of the 1pl.in suffixes is due to the fact that this is the only 1pl suffix used)

one of these might be the higher frequency of poss.1sg as verbal subject agreement marker – it’s hard to imagine any ‘real’ difference in this between the dialects; one possible explanation might be the higher frequency of –RI as past tense in KAM vs. –čE in SEB

To investigate this further, need to look at collocations of suffixes

2) SEB seems to make a notably higher use of converbs than KAM, esp. the simultaneous (incl. multiplicative and ‘conditional’) and anterior converbs (for purposive, KAM leads) 🡪 might be due to a syntactic difference between the dialects, e.g. KAM might have more monoclausal sentences, or make more use of conjunctions

3) the proprietive –lken is more frequent in SEB, while the designative/destinative –GE is more frequent in KAM – might be worth exploring further on

Next tasks:

1) Vasilisa will compile a distribution of collocations of suffixes and their following suffix to help us understand why we find more poss.1sg with verbs in SEB, for example (and to discover further potential differences between the 2 dialects, such use of locative-marked participles to mark predicates of temporal adverbial clauses)

2) Vasilisa will investigate the difference between SEB and KAM with respect to converbs by looking at the glosses, will find a way to automate this

If needed, Vasilisa will contact Brigitte in the field by sending emails to both the cnrs and the mpg address, and Brigitte will try to answer as soon as she can.

Brigitte will come to Moscow on Wednesday, 25th October (arriving in evening) so that we can work together on the project and the presentation for ConLab meeting on 26th October, before the “Contact in Circumpolar Regions” 🡪 will need to find place to work efficiently