In this talk, I will deal with Russian speech of older speakers of Nanai and Ulcha (Southern Tungusic, the Amur region). A great inter-speaker variation takes place: some bilingual Nanai and Ulcha speakers use a "near-pidgin" Russian variety, the speech of some others does not differ greatly from the monolingual benchmark. The data used in the study come from the Corpus of contact-influenced Russian of Northern Siberia and the Russian Far East (http://web-corpora.net/ruscontact/corpus.html). This is a small spoken corpus provided with a manual annotation of contact-induced grammatical features (non-standard agreement, non-standard argument encoding etc.). Based on this annotation, I will try to assess the inter-speaker variation attested in the corpus.
On the one hand, I will show
- which contact-induced features appear to be more stable, i.e. equally represented in texts produced by different speakers, and which ones contribute to inter-speaker variation most of all
- which features behave similarly, i.e. are equally frequent / infrequent in texts produced by the same speakers.

On the other hand, I will discuss
- how speakers group together according to contact-induced features typical of them
- whether these clusters of speakers correlate with any sociolinguistic parameters
- whether they go in line with the researcher's intuition or look surprising.

An additional motivation for this study is methodological. I will test how precisely the existing corpus annotation captures the degree of deviation from the monolingual benchmark and inter-speaker variation.