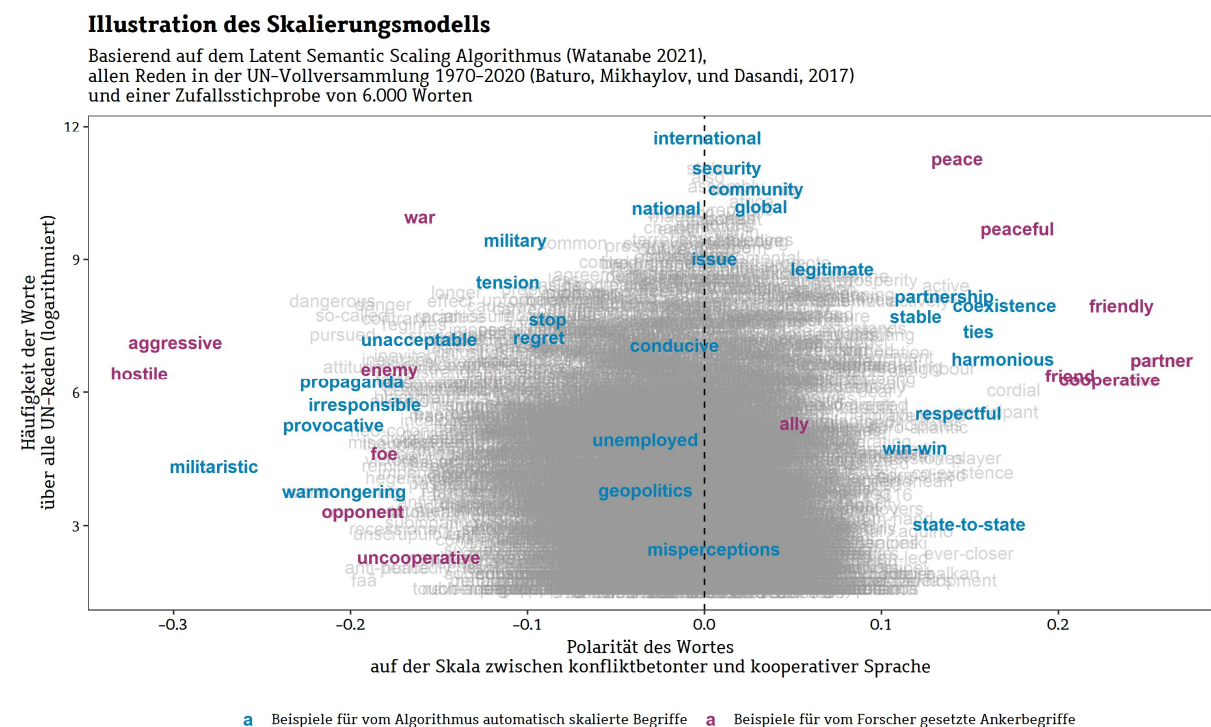


The idea was to identify conflictual vs. cooperative language in international diplomatic speech. To this end I trained a latent semantic scaling (LSS) algorithm (Watanabe 2021) on all UN General Assembly speeches 1970-2020 (from Mikhaylov et al 2017). The algorithm represents the corpus as a word vector retrieved from word-to-word cooccurrences, to then scale each word along its semantic similarity between two pre-specified poles (using Cosine similarity of word vectors as the metric). I used the following simple poles:

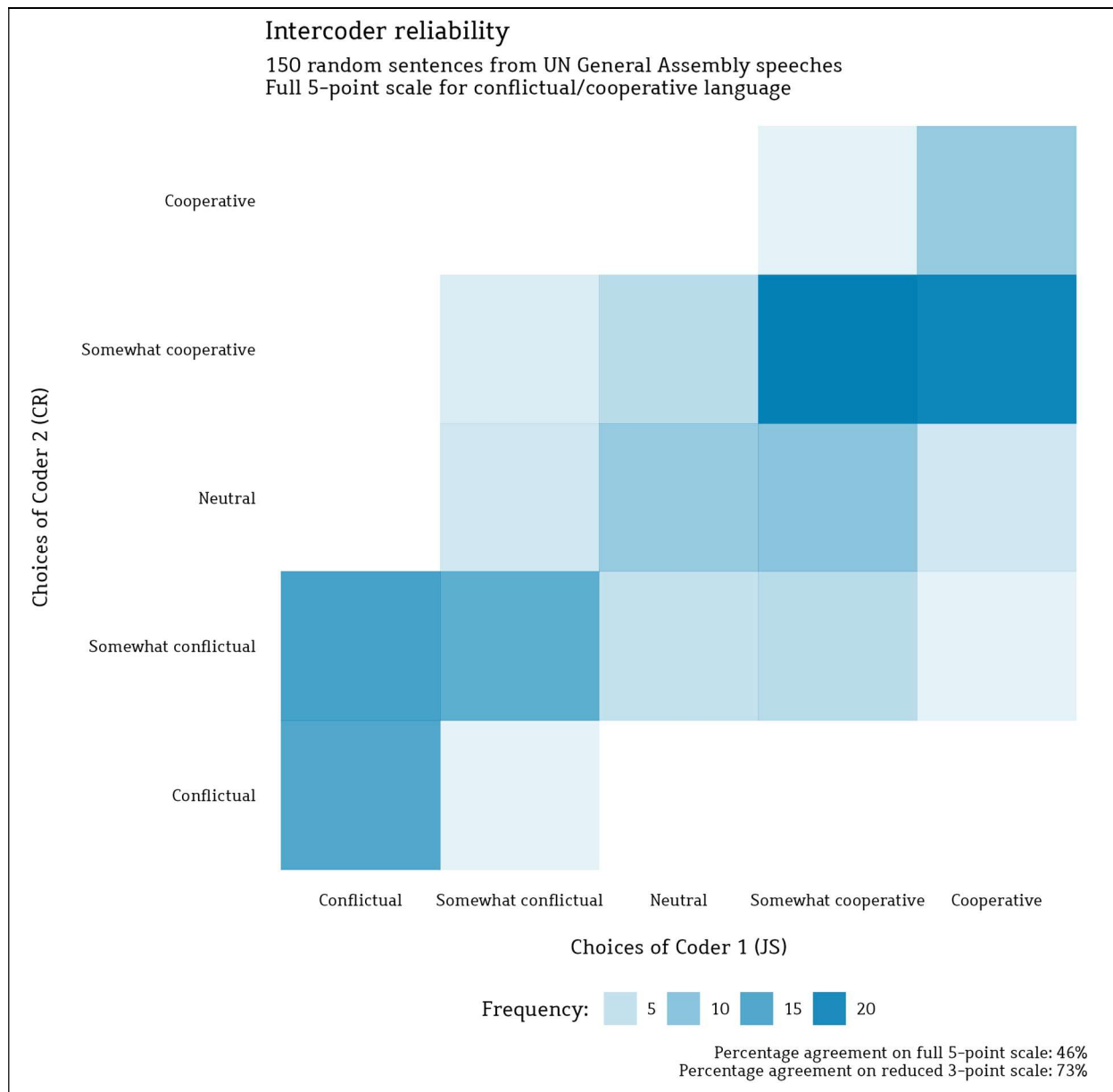
cooperative = c("friend", "partner", "ally", "peace", "peaceful", "friendly",
"cooperative"),
conflictual = c("foe", "opponent", "enemy", "war", "aggressive", "hostile",
"uncooperative")

Here is an illustration of what the algorithm has learned



Individual word weights on this scale can then in principle be used to scale a given piece of text...

How well does this work? Two coders were asked to rate 150 sentences (randomly sample from the almost 100.000 sentences spoken in the UNGA) for conflictual vs cooperative language on a 5-point scale.



Without additional context, humans are not very good at this (as usual), but this gives us some benchmark to assess how well the LSS estimates replicate human assessment of very short pieces of text.

