One popular approach to machine translation is to break sentences into small groups of contiguous words (phrases) and then to translate these phrases independently. Translations of these phrases are extracted beforehand from a large amount of bilingual data. The goal of this thesis is to detect semantical incorrectness in the extracted translations of phrases. One source of potential problems is poor quality of training data (high quality parallel data are very hard to obtain), more severe are possible problems introduced by double negative in Czech: the translated sentences might have a completely opposite meaning to the original one. We first tried to modify our prior work to penalize such erroneous translations. Then we designed and trained our own neural model to produce a semantical score for a given phrase translation. We evaluated the improvements on a small manually annotated set of translations and also in an end-to-end machine translation task. Using our model in an end-to-end machine translation system yields a significant improvement of 0.5 BLEU over the baseline. Our model also beats an existing state-of-the-art method not only in the end-to-end translation (by 0.2 BLEU), but also on the manually annotated data by a factor of more than 2 in recognition of incorrect translations.