

Segmentační metody jsou důležitou součástí při simultánním strojovém překladu, jelikož by v ideálním případě měly být zodpovědné za dělení vstupních vět na takové úseky, k jejichž překladům není potřeba znát kontext, který se dostaví až po nich. Optimální dělení je pak takové, pro něhož mají úseky zachovávat výše zmíněnou vlastnost minimální délky. Dodnes ovšem není stále jasné, jakým způsobem lze takovou optimální segmentaci vyprodukovat. Proto jsme nejprve popsali instrukce, jak získat optimální segmentaci simulováním dokonalého lidského tlumočnicka s neomezeným časem, a pak tímto způsobem oantovali již existující dataset ESIC. Následně jsme navrhli několik segmentačních metod, které jsme charakterizovali z hlediska frekvence a délek segmentů, a typů slov, které byly segmentačními metodami nejčastěji rozdělovány. Kromě segmentačních metod jsme pak také implementovali a analyzovali dva neuronové strojové překladače – první standardně naučený na dokončených paralelních větách a druhý dotrénovaný na částečných překladech. Nakonec jsme vyhodnotili vyprodukované úseky všech segmentačních metod pomocí evaluačního nástroje SLTev a zhodnotili, jaký vliv měla na jejich výsledky volba použitého překladače.