

Němčina je známá svou velmi produktivní slootovorbou, zejména v oblasti kompozice a derivace. V této práci se zaměřujeme na německé nominální složeniny a jejich zastoupení ve výstupech strojového překladu. Navzdory jejich důležitosti v německých textech, běžně používané metriky pro hodnocení kvality překladu, jako je BLEU, nedokážou použití složenin dostatečně zachytit. Cílem této práce bylo zkoumat generování německých složenin v modelech typu Transformer a prozkoumat faktory, které vedou k jejich tvorbě. Zjistili jsme, že strojové překladové systémy produkují méně složenin než lidé. Také se ukázalo, že kvůli velmi produktivní povaze německých složenin není možné je identifikovat na základě fixního seznamu. I po ručním vyhledání nových kompozit jich lidské překlady obsahovaly více než strojové.

Natrénovali jsme vlastní model typu Transformer pro překlad z angličtiny do němčiny, abychom to mohli zkoumat různé faktory, které ovlivňují produkci složenin, včetně segmentace slov a frekvence složenin v trénovacích datech. Dále jsme experimentovali s vynuceným dekódováním (forced decoding) a zjišťovali, jak se změní výstup systému po poskytnutí prvních slov překládané věty. Naše výsledky zdůrazňují důležitost dalšího výzkumu v oblasti strojového překladu, aby se byly překladové systémy schopny lépe přiblížit lidskému překladu a generovat více složenin.