

# SLOVENŠTINA A STROJOVÝ PŘEKLAD

**Datum konání:** 22. 10. 2021

**Místo konání:** Filozofická fakulta Masarykovy univerzity

**Název přednášky:** Slovenština a strojový překlad

**Přednášející:** Mgr. Lucia Benková, Mgr. Ľubomír Benko, PhD.

**Počet účastníků:** 9

**Zpracovali:** Pavlína Machová

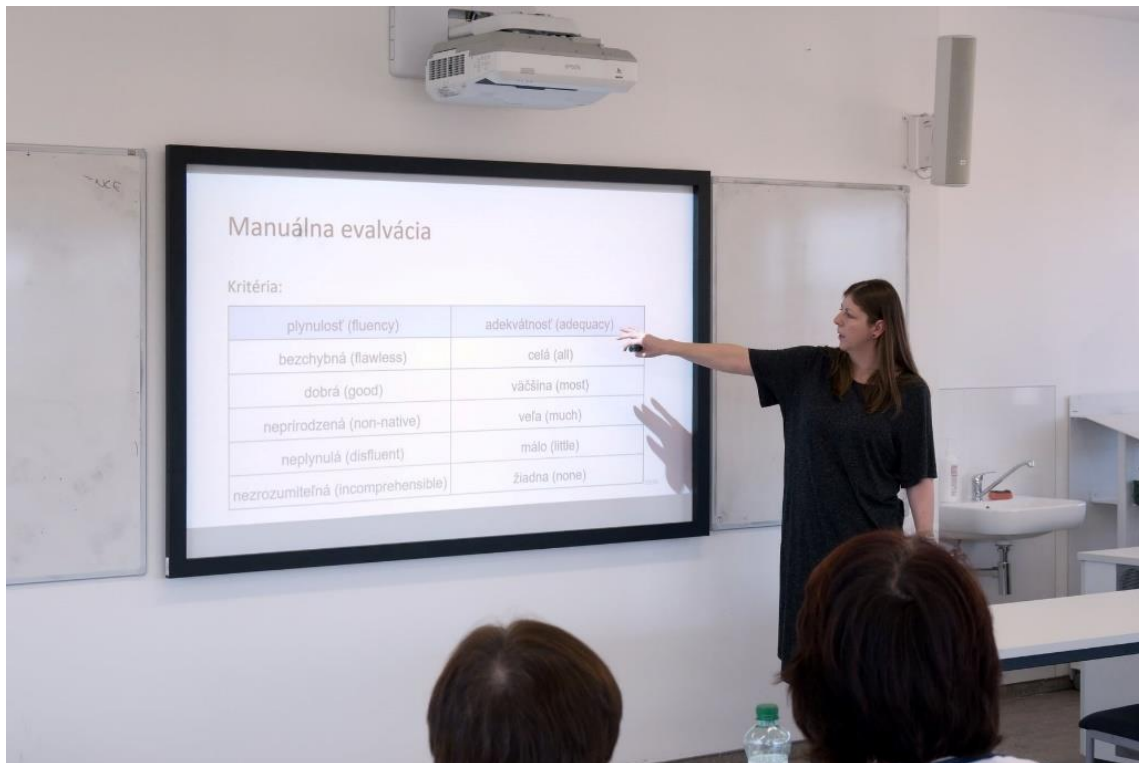
V rámci cyklu přednášek externích odborníků přijali pozvání na naši fakultu dva charismatičtí přednášející, kteří se věnují strojovému překladu i metodám jeho zefektivňování, a představili nám současný stav v oblasti zkoumání, jeho vývojovou linii, ale také svou specializaci. Nejen o těchto tématech přednášeli manželé Benkovi, kteří působí na Katedře informatiky Univerzity Konštantína Filozofa v Nitře.



Paní magistra Lucia Benková se ujala slova jako první a přehledně posluchače seznámila s využitím NLP v běžném životě, s historií strojového překladu i se základními pojmy potřebnými k pochopení dané problematiky. Tato přednášející je doktorandkou na Fakultě přírodních vied UKF a jelikož se ve své dizertační práci zaměřuje na evaluaci strojového překladu, spolupracuje s Katedrou translologie Filozofické fakulty UKF. Doktor Ľubomír Benko působí na Katedře informatiky UKF jako odborný asistent a věnuje se

zpracování přirozeného jazyka či chování uživatele na webu. Výklad své kolegyně příhodně doplňoval a společnými silami tak představili problémy i metodiku při zkoumání strojového překladu.

První pokusy o strojový překlad byly zaznamenány krátce po 2. světové válce, přičemž se od těchto dob jeho metody výrazně zdokonalily. Stále však i dnes zůstávají jisté skulinky, kvůli kterým není možné získat ve všech případech odpovídající překlad mezi zdrojovým a cílovým jazykem. Po úvodním představení základních pojmů i historie strojového překladu byli posluchači seznámeni s postupem následujícím po získání jazykových dat. Tato data prochází předzpracováním, a tedy segmentací, tokenizací, tagováním, lemmatizací, stemováním, parsováním a dalšími úpravami tak, aby bylo možné text snadněji zpracovávat a bylo dosaženo lepších výsledků.



Za dôležitou súčasť procesu je však považovaná také evaluácie strojového prekladu, ktorá sa provádí dvoma spôsobmi. Prvým z nich je manuálna evaluácia, pri níž je lingvisty vyhodnocovaná plynulosť (fluency) a adekvátnosť (adequacy) pomocí päťstupňového spektra. Za druhý spôsob je považovaná automatická evaluácia strojového prekladu, pri níž je vyhodnocovaná presnosť (precision) a pokrytí (recall).

Prednášející následně představili posluchačům editační systém OESTEPRE, jenž vzniknul na Katedře informatiky UKF. Jedná se o nástroj sloužící k evaluaci prekladu, v němž při kontrolování správnosti prekladu pracují studenti lingvistiky, kteří porovnávají automatický a lidský preklad věty ze zdrojového do cílového jazyka. Případné chyby prekladu zapisují dle jednotlivých kritérií do tabulek. Díky těmto výsledkům je možné se zaměřit na opakující se chyby, které nastávají při prekladu, a následně zvolit metodu, která přispěje k minimalizaci jejich výskytu. Metoda manuální kontroly je náročná nejen časově, a proto je snahou tuto práci omezit a nahradit ji alespoň částečně automatickou kvalifikací chyb.

Za všechny posluchače děkuji oběma přednášejícím za informačně nabitou přednášku, která byla výborná nejen díky podrobnému seznámení se zaměřením hostů, ale také díky poutavému vyprávění. I přes to, že se ke zkoumání váže mnoho překážek, mohli jsme na sebe kromě nových poznatků nechat působit také motivující nadšení pro obor, který je nám nejbližší.



### Súčasný stav, výsledky výskumu

- zameranie na hľadanie chýb strojového prekladu,
- porovnanie rôznych systémov strojového prekladu,
- evalvácia pomocou automatických metrik (PER, WER, TER, HTER),
- nutnosť ľudského prekladu a post-edítovaného prekladu,
- manuálna klasifikácia chýb pomocou kategorálneho rámca.