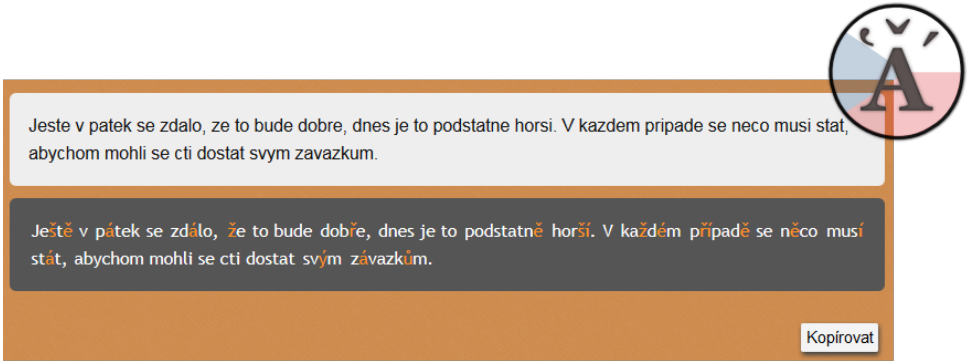
Někdo rozumí počítačům. Počítače nerozumí nikomu. Naučíme je to?

Diacritics Restoration

Otázka: **Nevadi, kdyz pisu bez hacku a carek?**

Odpověď: **Ne.**

Lidé často píšou bez diakritiky. Pokud chceme takovým textům rozumět a třeba v nich vyhledávat, musíme si diakritiku domyslet. Někdy je to snadné, protože slovo bez diakritiky ve slovníku nenajdeme (např. `pocitac`). Jindy je třeba využít znalostí o frekvencích výskytu a spoluvýskytu slov, které získáme z korpusu. Z nich program spočítá pravděpodobnost oháčkováného slova (např. `měj se` místo `mej se`).



Problémy s chybějící diakritikou jsou běžné i v mnoha dalších jazycích, např. ve slovenštině, galštině, španělštině, rumunštině či haitské kreolštině.

Centrum zpracování přirozeného jazyka
Fakulta informatiky
Masarykova univerzita