

Zamyšlení na komunikování statistiky v médiích

Vzhledem k zvolenému tématu práce jsem se pozastavil nad článkem, jehož odkaz jeden z mých přátel lingvistů umístil na Facebook. Jeho název „Jazyk má vliv na to, jak jsme bohatí, říká nová studie. A čeština na tom není dobře.“ navádí na pokus upoutat čtenáře na úkor reálných dat. Po letmém pročtení i další údaje jako „U rodin, které mluví jazyky se slabým odkazováním na budoucnost, je o 24 procent nižší pravděpodobnost, že budou silnými kuřáky.“ (Blažek, 2012) mi přišlo, že je článek podezřelý. Kupodivu obsahuje i odkaz na původní studii (Chen, 2012), která je k dispozici ke stažení a umožňuje tak porovnat celkové vyznění s původním textem.

Český text je postaven kolem úvodní myšlenky: „Jazyky, které přímo nevyjadřují budoucí čas, nenutí své mluvčí přemýšlet rozdílně o budoucnosti a přítomnosti. Lidé v zemích jako Německo, Japonsko či Čína tak odpovědněji přistupují ke svému zdraví a osobním financím.“ (Blažek, 2012) Co se týče argumentů, kterými toto tvrzení podepírá, jsou spíše volně formulované, než, že by obsahovala nějaká pevná čísla. Formulace jsou spíše shrnující než definičně přesné jako například: „Ve své studii zjistil, že lidé v zemích, které se umístily vysoko v celkové odpovědnosti ke zdraví a financím, téměř vždy mluvili jazykem se „slabým“ odkazováním na budoucí čas.“

V druhé části článku jsou naopak „citována“ přesná čísla. „U rodin, které mluví jazyky se slabým odkazováním na budoucnost, je o 24 procent nižší pravděpodobnost, že budou silnými kuřáky. Existuje u nich také o 13 procent menší riziko obezity. Jsou v lepší fyzické kondici při odchodu do penze. V mnoha ohledech mají dlouhodobě lepší zdraví,“ říká Chen.

Článek kromě studie samé zmiňuje i odmítavé reakce jiných jazykovědců a také zmiňuje, že studie ještě nebyla publikována a prochází recenzním řízením.

Podíváme-li se na citovaný text s přesnými čísly a na část původního článku, která tato čísla uvádí, dostaneme se k následujícímu zjištění. Český text se pokusil obrátit význam věty, aniž by se zabýval převodem čísel. Pro autora je tedy stejné tvrzení, že jazyky se slabým odkazováním na budoucnost (SLOB) o 24% snižují pravděpodobnost silného kuřáctví jako tvrzení opačné. Původní text totiž uvádí: „...strongly grammaticalized FTR leads to 24% higher probability of having ever smoked (daily for a year or more)“ (Chen, 2012) Je zjevné, že se nejedná o stejný údaj. Podle Blažka by tedy byl poměr silných kuřáků u jazyků se silným odkazováním vs. u slabého odkazování byl $100/76 = 1.32$, zatímco podle Chena $124/100 = 1.24$. Podobné chyby se logicky Blažek dopouští i u obezity.

Oprostíme-li se od této poměrně významné faktické chyby, můžeme se zabývat tím, jak moc se změnilo vyznění celého textu. Blažek například vůbec neřeší, co to znamená, že lidé budou silnými kuřáky. Je silným kuřákem člověk, který vykouří krabičku denně, nebo stačí jedna cigareta? Druhým prvkem, který má značný dopad na vyznění je samotné číslo 24%. Je to vysoká hodnota a navádí k tomu, že celá čtvrtina populace může být ovlivněna, co se týče kouření. Nikde

není zřejmé, jaká je vlastně ona výchozí hodnota pravděpodobnosti, k níž se toto zvýšení vztahuje. Dlužno podotknout, že ani Chen ve svém článku neuvádí, jaké pravděpodobnosti jsou ty výchozí. Těsně následující číslo 13% týkající se obezity také navozuje dojem, že se zabýváme stejným výchozím stavem. Výchozí pravděpodobnost je téměř určitě jiná a výsledný dojem je zavádějící.

Chen se ve svém článku věnuje několika okruhům, které Blažek všechny zmiňuje. K podepření své teorie využívá logistické regresní analýzy a její výsledky Blažek ve svém textu využívá. Nedá se říci, že by nějakou významnou skutečnost opomněl.

Celkově se dá říct, že Blažkův článek relativně dobře reflektuje Chenovy závěry. Problémem je chyba s obráceným poměrem, která zřejmě plyne spíše z nepozornosti, než z úmyslu výsledky měnit. Udávání tvrdých čísel subjektivně zkresluje dojem a vyvolává pocit, že efekt je opravdu velký. Ve skutečnosti nejsme vůbec schopni učinit jiný závěr než, že nějaký efekt zde pravděpodobně je. Krátce bych mohl říci, že tento článek je spíše z těch lepších a nedeformuje původní zdroj příliš.

Poměrně fajn myšlenky, jen některé by se daly ještě více rozvést, chtěla bych poprosit o Vaše reakce na komentáře... potom práci ráda přijmu...

AR

Literatura:

▲ Blažek, O. (2012). *Jazyk má vliv na to, jak jsme bohatí, říká nová studie. A čeština na tom není dobře*. Staženo 14. 4. 2012 z <http://zpravy.ihned.cz/svet-usa/cl-55195160-jazyk-ma-vliv-na-to-jak-jsme-bohati-rika-nova-studie-a-cestina-na-tom-neni-dobre>

▲ Chen, M., K. (2012). *The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health, Behaviors and Retirement Assets*. Staženo 14. 4. 2012 z <http://faculty.som.yale.edu/keithchen/papers/LanguageWorkingPaper.pdf>