

---

---

# EMPIRICKÝ VÝZKUM VÁGNÍHO VÝZNAMU: VÁGNÍ SKALÁRNÍ MODIFIKÁTORY V ČEŠTINĚ

RADEK OCELÁK

---

---

V následujícím příspěvku předkládám výsledky empirického sémantického výzkumu, který jsem provedl v rámci své bakalářské práce, zabývající se fenoménem významové vágnosti.<sup>1</sup> Na jeho úvod nastiňuji – jen v nezbytné míře – teoretický kontext výzkumu.

Sémantickou vágnost je nejprve třeba vymezit vůči jiným fenoménům, které se často shrnují pod obecnější pojem významové neurčitosti. Termínem *vágní* označuji význam takového výrazu, pro nějž není zřejmé, zda určité objekty spadají do jeho extenze. Jinak řečeno, o sémantickou vágnost jde tam, kde kvůli neostře ohraničeným referenčním možnostem výrazu (nejasným hraničním pojmu) nelze s jistotou určit pravdivostní hodnotu výpovědi. Emblematickými a v literatuře často užívanými příklady jsou anglické výrazy s neurčitou extenzí *tall* (od jaké výšky lze o člověku pravdivě tvrdit, že je *vysoký?*), *heap* (při kolikátém odebraném zrnku přestane být hromada *hromadou?*), za příklad by však mohlo sloužit bezpočet jiných výrazů přirozeného jazyka. V tomto příspěvku se zabývám konkrétní kategorií výrazů – vágními skalárními modifikátory v češtině, tj. výrazy jako *docela*, *velmi*, *celkem*. U těchto synsémantik je problematické hovořit o extenzi či referenci, jejich funkcí však je (spolu)působit vágnost jmenné nebo slovesné fráze, na níž se podílejí – označují je proto rovněž jako vágní.<sup>2</sup>

Sémantice tohoto typu výrazů se v poslední době věnoval V. Veselý.<sup>3</sup> S užitím binárních sémantických modelů Veselý předkládá jednoduchý, explicitní pohled na systém výrazů (sám užívá termínu „neurčité kvantifikátory“) jako *dost*, *velmi*, *skoro*, *přínejmenším*, *příliš* apod., který podněcuje k dalšímu rozvíjení, ale i ke kritice z hlediska empirické adekvátnosti. Vůči jeho pojetí daného systému mám dvě zásadní výhrady:

Veselý vychází z pojmu *kvantifikace*. Chápe jej sice neobvykle širokým způsobem (řadí sem kromě udávání množství a rozsahu – „*hodně lidí*“, „*hodně vody*“ – i udávání intenzity statického či dynamického příznaku – „*hodně mrzlo*“), i přesto je však – výhrada první – jeho výklad da-

---

1 Ocelák, R.: *Vágní význam a jeho empirické modelování – případ vágních skalárních modifikátorů v češtině*. Nepublikovaná bakalářská diplomová práce, FF UK, Praha 2010. – Mnohé problémy, jichž se v příspěvku dále dotýkám, zvláště otázky teoretické, jsou v této práci pojednány mnohem důkladněji.

2 Jinými, od vágnosti odlišnými případy tzv. významové neurčitosti jsou např. kontextová závislost, víceznačnost či informační podspecifikovanost. Vymezení vágnosti vzhledem k různým odlišným typům významové neurčitosti podává např. J. V. Neustupný (Neustupný, J. V.: On the Analysis of Linguistic Vagueness. In: Aarts, B. a kol. (eds.): *Fuzzy Grammar: A Reader*. Oxford University Press, New York 2004, s. 341–350) a R. Keefe (Keefe, R.: *Theories of Vagueness*. Cambridge University Press, Cambridge 2000, s. 10). ESČ (Karlík, P. – Nekula, M. – Pleskalová, J. (eds.): *Encyklopedický slovník češtiny*. NLN, Praha 2002, s. 517) definuje pojem *vágnost* šíře, zahrnuje pod něj i zde odlišené typy sémantické neurčitosti.

3 Veselý, V.: Lexikální sémantika neurčitých kvantifikátorů v současné češtině. *Slovo a slovesnost*, 70, 2009, s. 175–192.

ných výrazů jakožto udávajících kvantum v některých typech užití velmi násilný.<sup>4</sup> Domnívám se, že příslušný systém výrazů je třeba charakterizovat pomocí pojmu *skalárnosti* – ten zahrne jak užití v úzkém smyslu kvantifikační či intenzifikační (a pojem kvantifikace není pak nutno neúnosně rozvolňovat), tak užití jiná. Synsémantika, která Veselý pojednává, proto označuji termínem *skalární modifikátory*. Funkcí skalárních modifikátorů je udávat interval z určité jednoznačně řazené škály možností – např. ve výpovědi „*Velmi* mě to láká“ modifikátor *velmi* udává popisovaný stav jako některý z jistého intervalu stavů, který je úsekem na škále hypotetických stavů seřazených od minimálního po maximální „lákání“. O jiný interval na téže škále by pak šlo v třeba v případě údaje „*Celkem* mě to láká“. Udávaný interval může být ve výše vymezeném smyslu vágní (v případě modifikátorů *velmi, celkem, docela, poměrně, skoro, téměř, přibližně, sotva*), nebo přesně vymezený (*příliš, přinejmenším*) – v obou případech jde o neurčitost ve smyslu informační podspecifikovanosti (reálný stav je udán prostřednictvím intervalu stavů, mezi něž spadá, a tedy nepřesně), v druhém se však nejedná o vágnost. Zatímco sémantiku nevágních skalárních modifikátorů Veselý v rámci svého pojetí vykládá přesvědčivě,<sup>5</sup> jeho analýzu modifikátorů vágních – výhrada druhá – mám za velmi nedostatečnou a budu se jim dále věnovat. Konkrétněji se zaměřím na subsystém vágních skalárních modifikátorů spojujících se s výrazy, které samy již vágně udávají úsek škály, a to úsek přiléhající k jednomu jejímu pólu (tj. např. s adjektivy *vysoký, malý*, adverbium *úspěšně*, predikativem *tepló*, slovesem *sněžit* – jde tedy o modifikátory *docela, celkem, velmi, dost, poněkud, velice, nepřítis* atd.).

Sémantiku těchto modifikátorů V. Veselý modeluje binárními sémantickými strukturami. Z těch plyne, že některé z modifikátorů (*velmi, hodně, dost*) vybírají na škále interval hodnot vyšších, než je hodnota v daném kontextu obvyklá, standardní; jiné (*ne- příliš, málo*) interval hodnot nižších. To lze však exaktně formalizovat jen za předpokladu, že vybíraný interval hodnot obsahuje buď *všechny* hodnoty větší než standardní hodnota, nebo *všechny* hodnoty menší – tedy že udávaný interval je přesně vymezený. Takový model by ignoroval *vágnost* významu těchto modifikátorů. Veselý správně připouští, že kvantifikátory „často implikují také informaci, jak velký je rozdíl mezi kvantem bázevým a kvantem udávaným“<sup>6</sup>. Konkrétněji např. uvádí, že pro kvantifikátor/modifikátor *dost* je rozdíl mezi standardní hodnotou a vybíraným intervalem ve srovnání s *hodně* hodnocen jako menší („*dost* tiché místo“ × „*hodně* tiché místo“)<sup>7</sup>. Takové hodnocení je však již mimo hranice užitého formalismu; autor může rozdíly mezi standardními a vybíranými hodnotami pro jednotlivé modifikátory pouze odhadem popsat jako velké či malé. Zde se ukazují meze Veselého introspektivní metody – můžeme se třeba ptát, v jakém vztahu jsou rozdíly standardních a vybíraných hodnot pro *hodně* a *velmi* či zda je skutečně příslušný rozdíl menší pro *dost* než pro *hodně*. Začneme-li uvažovat o množství dalších, autorem nezmiňovaných vágních skalárních modifikátorů, jasně se ukáže, že sémantická introspekce je pro postižení detailních sémantických rozdílů mezi prostředky daného subsystému zcela nedostatečná. Bez empirických dat o užívání těchto výrazů stěží dokážeme posoudit sémantickou rozdílnost (popř. formulovat věrohodný závěr o sémantické ekvivalenci)

4 Srov. užití typu „Tvrdit něco takového je *skoro* šíření poplašné zprávy“ či „Za pět let bude *přinejmenším* nadporučíkem“, v nichž rozhodně nelze spatřovat udávání množství, rozsahu ani intenzity.

5 Rozdíl vágních a nevágních kvantifikátorů/modifikátorů ovšem autor výslovně nezmiňuje. – K přesnosti intervalů udávaných tzv. modálními skalárními modifikátory typu *přinejmenším, nanejvýš* srov. také studii Geurts, B. – Nouwen, R.: „At least“ et al.: the Semantics of Scalar Modifiers. *Language*, 83, 2007, s. 553–559.

6 Veselý, V.: Lexikální sémantika neurčitých kvantifikátorů v současné češtině. *Slovo a slovesnost*, 70, 2009, s. 178.

7 Tamtéž, s. 181.

výrazů *docela, celkem, vcelku a poměrně, velmi, nesmírně, hodně a moc, ne- moc, ne- příliš a málo* atd. Empirický výzkum provedený v rámci mé bakalářské práce se zaměřil právě na užívání nejběžnějších výrazů z tohoto subsystému a přinesl oproti neempirickému přístupu V. Veselého některé nové detailní informace o jejich vzájemných sémantických vztazích.

K získání empirických dat o užívání výrazů s vágním významem se obecně nabízela nejprve metoda experimentu s kontrolovaným referentem či metoda metajazykového dotazování. V daném typu experimentu jde o podněcování referenčního či predikativního užívání zkoumaného typu výrazů o objektech kontrolovaných experimentátorem.<sup>8</sup> Cíl získat dostatečné množství skutečně reprezentativních dat by si však vyžádal experiment v rozsahu, jaký nebylo v mých možnostech realizovat. Získávání dat pomocí metajazykového dotazníku je snadnější, takto získaná data by ovšem nevypovídala o užívání výrazů samém, nýbrž o tom, jak mluvčí toto užívání reflektují – ani tuto metodu jsem proto nevyužil.

Byla dále nasnadě otázka, zda nelze vytěžit potřebná data z dostupných rozsáhlých souborů elektronicky prohledavatelného, holého či lingvisticky anotovaného textu. Zde je zásadní obtíž – pro zkoumání extenze, tj. referenčních možností užívaného výrazu, je nezbytné mít informace o extralingvistických objektech, k nimž výraz v jednotlivých užitích referuje; pouhý výrazový záznam jazykové komunikace však referenty jednotlivých výrazů nepřístupňuje. Tuto nesnáž lze nicméně určitým způsobem obejít. V obrovském rozsahu elektronicky přístupných textů se totiž sporadicky objevují taková užití výrazů s vágním významem, jejichž referent je v dostupném kontextu popsán co do klíčové vlastnosti také přesněji. V případě vágních skalárních modifikátorů může jít např. o doklady typu „Dnes je tam *docela teplo*, 24 stupňů“ či „Měřím 186 cm. – Tak to jsi *dost vysoko*, já mám jen 174 cm.“ Svůj výzkum jsem založil právě na vyhledávání velmi specificky vymezených dokladů takového typu – souvyskytů vágního a přesného vyjádření tělesné výšky téhož referenta – dospělé ženy, kde vágní vyjádření má podobu „*vágní skalární modifikátor* + vysoká/malá“ a přesné vyjádření je údajem o výšce v centimetrech.<sup>9</sup>

Stanovený kontext udávání lidské tělesné výšky spojuje několik výhod. Spojení „*vágní skalární modifikátor* + vysoká/malá“ se v něm užívá relativně často, což je jedním z předpokladů pro získání takového množství vyhovujících dokladů, které by umožnilo případné zjištění rozdílů v užívání modifikátorů shledat statisticky významnými. Stupně na škále tělesné výšky se zároveň běžně udávají přesným způsobem, počtem fixních jednotek vyjadřujícím exaktně ověřitelnou míru vlastnosti – to umožní statisticky ověřovat významnost korelace mezi vágními vyjádřeními a spoluvyskytujícími se exaktními hodnotami. Konečně lze v takto úzce vymezeném kontextu předpokládat nevelký rozptyl hodnot pokládaných různými mluvčími za hodnotu standardní.<sup>10</sup> Specifikujeme-li ovšem množinu vyhovujících dokladů takto striktně,

8 Za prototyp takového výzkumu lze považovat experiment publikovaný ve známé studii W. Labova (Labov, W.: *The Boundaries of Words and their Meanings*. In: Aarts, B. a kol. (eds.): *Fuzzy Grammar: A Reader*. Oxford University Press, New York 2004, s. 67–90), v němž byly zkoumány referenční možnosti výrazů *cup, mug, bowl, vase* a prostupnost jejich hranic. V našem případě by bylo ovšem nutno experiment navrhnout s ohledem na sysnématický charakter vágních skalárních modifikátorů.

9 Na detailní popis podmínek, jež musel každý nalezený doklad pro zařazení do výzkumu splnit, mohu z rozsahových důvodů jen odkázat do své bakalářské práce (Ocelák, R.: *Vágní význam a jeho empirické modelování – případ vágních skalárních modifikátorů v češtině*. Nепublikovaná bakalářská diplomová práce, FF UK, Praha 2010, s. 39n.).

10 Souhlasím tedy s předpokladem V. Veselého (Veselý, V.: Lexikální sémantika neurčitých kvantifikátorů v současné češtině. *Slovo a slovesnost*, 70, 2009, s. 175–192), že kontextově doplňovaná standardní hodnota je složkou sémantiky zkoumaných vágních skalárních modifikátorů, a specifikací kontextu se snažím její možnou variaci v počítování autorů získaných dokladů co nejvíce omezit.

znamená to, že hledáme doklady velice řídké, a není tak možno se vyhnout jejich vyhledávání v celé šíři českojazyčného internetu, jakkoli to s sebou oproti lingvistickým korpusům nese jistě nevýhody.<sup>11</sup> Provedl jsem tedy výzkum na internetovém materiálu – dokladech daného typu pro 6 vágních skalárních modifikátorů (*celkem, docela, hodně, dost, poměrně, velmi*).<sup>12</sup>

V první části výzkumu, zabývající se spojeními modifikátorů s adj. *vysoká*, bylo získáno celkem 189 dokladů. Z toho připadalo na každý z modifikátorů *celkem, docela, poměrně, dost, hodně, srovnatelné* množství 30–47 dokladů, na zbylý modifikátor *velmi* pouze 10 dokladů (připisují to jistě jeho knižnosti).

Tabulka č. 1 uvádí základní statistické míry souborů přesných hodnot v centimetrech, jimiž byla ve získaných dokladech zároveň s vágním určením pomocí jednotlivých modifikátorů přesně udána výška osoby – počet hodnot, aritmetický průměr, medián a směrodatnou odchylku.<sup>13</sup>

	<b>celkem</b>	<b>docela</b>	<b>poměrně</b>	<b>dost</b>	<b>hodně</b>	<b>velmi</b>
<b>počet dokl.</b>	34	47	30	32	36	10
<b>aritm. průměr</b>	174,059 cm	175,947 cm	177,15 cm	180,609 cm	183,194 cm	186,7 cm
<b>medián</b>	174 cm	176 cm	177,25 cm	179 cm	181,5 cm	185,5 cm
<b>směr. odch.</b>	4,341 cm	4,337 cm	4,509 cm	5,871 cm	5,835 cm	6,183 cm

Tabulka 1 – Statistické míry souborů získaných hodnot pro jednotlivé vágní skalární modifikátory ve spojení s adj. *vysoká*. Zaokrouhleno na 3 desetinná místa.

Jak tabulka ukazuje, v zachyceném **vzorku** užívání platí, že průměr přesně udaných hodnot výšky v centimetrech popisujících současně s vágním údajem (*modif. + vysoká*) téhož referenta se pro jednotlivé modifikátory liší – stoupá v pořadí *celkem, docela, poměrně, dost, hodně, velmi*.

11 Omezující specifika získávání jazykového materiálu z internetu podrobněji rozebírám ve své bakalářské práci (Ocelák, R.: *Vágní význam a jeho empirické modelování – případ vágních skalárních modifikátorů v češtině*. Nepublikovaná bakalářská diplomová práce, FF UK, Praha 2010, s. 40n.) – hlavním z nich je, že jde o zdroj nereferenční a prohledatelný vždy jen neúplně. Při vyhledávání byly proto průběžně zaznamenávány pozitivní doklady (uvádím je v krátkém kontextu a s adresou nálezu v příloze své bakalářské práce), bez nároku na nalezení všech existujících dokladů v úplnosti. Výzkum je proto prověřitelný po stránce akceptovaných dokladů, vzhledem k nestálosti internetového obsahu a průběžně se měnícím výsledkům téhož dotazu ve vyhledávací však není možno jej replikovat beze zbytku a prověřit také všechny výsledky dotazů, které nebyly mezi doklady akceptovány.

V mé bakalářské práci (Ocelák, R.: *Vágní význam a jeho empirické modelování – případ vágních skalárních modifikátorů v češtině*. Nepublikovaná bakalářská diplomová práce, FF UK, Praha 2010, s. 41) viz rovněž detailní popis vyhledávacího postupu a využitých dotazů. Je podstatné, že byl pečlivě zachován princip náhodnosti a reprezentativnosti, který je podmínkou pro statistické hodnocení významnosti zjištěných rozdílů. Srov. Volín (Volín, J.: *Statistické metody ve fonetickém výzkumu*. Epoque, Praha 2007, s. 19): „Oba principy jsou spolu úzce spjaty. Pokud má totiž výběrový soubor reprezentovat soubor základní, pak z něj musí být vybrán zcela náhodně. Každý prvek základního souboru musí mít při výběru stejnou šanci stát se součástí souboru výběrového.“ Ve výzkumu byly zadávány takové dotazy, které vyhledávaly příslušné vágní vyjádření a nijak nepředurčovaly, jaké hodnoty vyjádření přesného se případně objeví v souvškytu s ním. (Např. byly vyhledávány takové stránky, které obsahují frázi „hodně vysoká“ a současně zkratku „cm“.)

12 Tyto modifikátory byly vybrány pro svou relativně vysokou frekvenci a pro obtížnost introspektivního posuzování sémantických vztahů např. ve skupinách *celkem, docela a poměrně* či *dost, hodně a velmi*.

13 Ve statistickém zpracování dat se opírám o monografii J. Volína (Volín, J.: *Statistické metody ve fonetickém výzkumu*. Epoque, Praha 2007), podávající na příkladech problémů fonetiky obecné statistické poučení. K výpočtu měř užívám funkce PRŮMĚR, MEDIAN a SMODCH.VÝBĚR programu MS Excel.

Ve vzorku tedy např. platí, že spojení *celkem vysoká* bylo užito o referentech s průměrně nižší tělesnou výškou než spojení *docela vysoká*. To by mohlo naznačovat, že pomocí modifikátoru *celkem se* obecně udávají na škále hodnoty, jejichž průměr je bližší hodnotě kontextového standardu a vzdálenější od příslušného pólu škály než průměr hodnot udávaných modifikátorem *docela*. Nejprve je však nutno statisticky prověřit možnost, že rozdíly nalezené mezi soubory hodnot jsou pouze náhodnými vlastnostmi vzorku, z nichž nelze usuzovat na charakter **populace**, tj. všech existujících dokladů daného charakteru, z nichž byl vzorek 189 dokladů s dodržáním principu náhodnosti a reprezentativnosti získán.

Jelikož hodnotíme vzájemně nezávislé soubory s rozdělením hodnot nepříliš vzdáleným od tzv. normálního rozdělení a s vzájemně srovnatelným hodnotovým rozptylem, můžeme využít příslušný **oboustranný t-test**.<sup>14</sup> Ten srovnává vždy dva soubory hodnot a udává **hodnotu  $p$** , vyjadřující pravděpodobnost chybného zamítnutí **nulové hypotézy**. Ta v našem případě říká, že celá populace existujících dokladů stanoveného charakteru pro spojení *modifikátor A + vysoká* se svým aritmetickým průměrem příslušných přesných hodnot nijak neliší od populace takových dokladů pro spojení *modifikátor B + vysoká*, a že tedy rozdíl mezi získanými vzorky obou populací je pouze náhodný. Zamítnout nulovou hypotézu pak znamená tvrdit, že průměr příslušných hodnot *celé* populace A, jejíž vzorek vykázal oproti vzorku populace B nižší průměr těchto hodnot, je odlišný od průměru příslušných hodnot v *celé* populaci B. (Tato **alternativní hypotéza** přitom tvrdí pouze odlišnost průměrů obou populací, nikoli to, který z nich je vyšší.) Pokud je odlišný jejich průměr, nutně je pak odlišné také rozložení těchto hodnot – v našem případě to znamená, že se daných dvou spojení (*modif. A + vysoká; modif. B + vysoká*) neuzívá stejným způsobem, tj. jejich význam se jistým způsobem liší.<sup>15</sup>

Hodnota  $p$  vypočtená t-testem udává, že pokud by platila nulová hypotéza (průměry populací byly stejné), byla by pravděpodobnost  $p$ , že absolutní hodnota rozdílu průměrů náhodně vybraných vzorků o velikosti vzorků, jaké jsme vybrali my, by byla nejméně tak vysoká jako absolutní hodnota rozdílu nalezeného na vzorcích námi vybraných. Pro nyní prezentovaný výzkum byla zvolena v humanitních vědách obvyklá hladina významnosti na úrovni  $p = 0,05$  (tj. 5 %) – za **statisticky významný** tedy považují takový rozdíl, pro nějž dá t-test hodnotu  $p < 5$  %. Následuje J. Volína<sup>16</sup> budu dále za rozdíl **statisticky okrajově významný** považovat rozdíl, pro nějž se bude  $p$  pohybovat v rozmezí 5–8 %.

*Tabulka č. 2* obsahuje výsledky t-testu pro každou dvojici modifikátorů ve spojení s adj. *vysoká* (údaj „ $x^{E-y}$ “ znamená „ $x * 10^{-y}$ “ – takto jsou zaznamenávány hodnoty menší než 0,0001 %):

14 Viz Volín, J.: *Statistické metody ve fonetickém výzkumu*. Epoque, Praha 2007, kap. 5 – *Studentovy t-testy*. V programu MS Excel jde o funkci TTEST, dvoustranný, typ 2.

15 Zůstávám zatím na úrovni sémantiky celého spojení, např. *celkem vysoká* – k závěrům o sémantice vágních skalárních modifikátorů samých přikročím až na základě údajů z obou částí výzkumu.

16 Volín, J.: *Statistické metody ve fonetickém výzkumu*. Epoque, Praha 2007, s. 37.



<b>docela</b>	5,68357 %				
<b>poměrně</b>	0,69605 %	24,60739 %			
<b>dost</b>	0,00025 %	0,0117 %	1,20422 %		
<b>hodně</b>	2,71804 <sup>E-08</sup> %	6,27435 <sup>E-07</sup> %	0,00181 %	7,35591 %	
<b>velmi</b>	4,96569 <sup>E-07</sup> %	1,79925 <sup>E-06</sup> %	0,00056 %	0,72622 %	10,40347 %
	<b>celkem</b>	<b>docela</b>	<b>poměrně</b>	<b>dost</b>	<b>hodně</b>

Tabulka 2 – T-test dvojic souborů hodnot (viz Tabulka 1) získaných pro jednotlivá spojení *modifikátor* + adj. *vysoká*. Zaokrouhлено na 5 desetinných míst.

Některé z údajů v této tabulce jen potvrzují, co bylo možno s jistotou předpokládat již na základě introspektivního posouzení sémantiky výrazů – výrazný rozdíl v umístění intervalu udávaného dvojicí skalárních modifikátorů na škále se projevuje velmi nízkou hodnotou *p*. Takový nepřekvapivý rozdíl je vykázan např. dvojicemi *celkem vysoká* × *hodně vysoká*; *docela vysoká* × *velmi vysoká* aj. Další údaje však již přinášejí netriviální poznatky o sémantice daných spojení. Z dvojic, pro které jsme kvůli introspektivní nepřesvědčivosti především žádali empirické ověření sémantického rozdílu, byl rozdíl v užití s pravděpodobností přesahující 95 % shledán u těchto: *dost vysoká* × *velmi vysoká*; *poměrně vysoká* × *dost vysoká*; *docela vysoká* × *dost vysoká* a *celkem vysoká* × *poměrně vysoká*. Statisticky okrajově významný rozdíl byl nalezen u dvojic *dost vysoká* × *hodně vysoká* a *celkem vysoká* × *docela vysoká*. Rozdíl pod úrovní okrajové statistické významnosti byl shledán u dvojic *hodně vysoká* × *velmi vysoká*; *docela vysoká* × *poměrně vysoká*.

Pro dvojice, u nichž jsme našli statisticky významný rozdíl, tedy můžeme (s *p*-procentní pravděpodobností omylu) odmítnout nulovou hypotézu a zaujmout k ní komplementární hypotézu alternativní: populace, z nichž jsme vybrali statisticky významně rozdílné vzorky, se svými průměry liší. Tato alternativní hypotéza sama o sobě pouze tvrdí, že sémantika daných dvou spojení není totožná, nespécifikuje však, *pro které* z nich je průměrná udávaná hodnota vyšší. Pokud je však mezi vzorky populací A a B statisticky významný rozdíl, jsme oprávněni usoudit (byť již bez přesně vyčíslené pravděpodobnosti), že vzorek s vyšším průměrem hodnot odpovídá populaci s vyšším jejich průměrem.<sup>17</sup>

Druhá část výzkumu se zaměřila na spojení týchž modifikátorů s adj. *malá*. Bylo získáno celkem 138 dokladů. Z nich připadlo na jednotlivé modifikátory *celkem*, *docela*, *poměrně*, *dost*, *hodně* srovnatelné množství 20–36 dokladů, pro spojení s modifikátorem *velmi* bylo pak opět nalezeno podstatně méně, pouze 7 dokladů. Tabulka č. 3 opět uvádí základní statistické míry souborů získaných hodnot pro jednotlivé modifikátory:

17 Náležitost tohoto závěru mi v osobní komunikaci potvrdil J. Volín: „[V]ýsledek se interpretuje opravdu tak, že skupina B reprezentuje populaci s vyšším průměrem a že riziko náhody je velmi malé. Nulová hypotéza se týká stejnosti a je jakousi linií mezi vaším výsledkem a stavy opačnými. Vyčíslování pravděpodobnosti vůči kontinuu opačných stavů se nevyžaduje – bylo by to výpočetně složité a nic zásadního by to neznamenalo.“

	<b>celkem</b>	<b>docela</b>	<b>poměrně</b>	<b>dost</b>	<b>hodně</b>	<b>velmi</b>
<b>počet dokl.</b>	25	36	22	28	20	7
<b>arit. průměr</b>	162,4 cm	160,083 cm	159,136 cm	158,946 cm	155,65 cm	151,214 cm
<b>medián</b>	163 cm	160 cm	160 cm	159,5 cm	156,5 cm	150 cm
<b>směr. odch.</b>	5,148 cm	4,959 cm	4,004 cm	5,175 cm	4,945 cm	4,339 cm

Tabulka 3 – Statistické míry souborů získaných hodnot pro jednotlivé vágní skalární modifikátory ve spojení s adj. *malá*.

Jednotlivé modifikátory jsou tedy na základě průměru příslušných přesných hodnot v daném vzorku seřazeny přesně opačně oproti první části výzkumu. Opačný je ovšem také pól dané škály, k němuž přiléhá interval stupňů udávaných na této škále modifikovaným adjektivem. Proto pro tento vzorek platí, že průměr hodnot udávaných jednotlivými modifikátory na dané škále se vzdaluje standardní hodnotě a blíží příslušnému pólu v témže pořadí jako v prvním případě: *celkem*, *docela*, *poměrně*, *dost*, *hodně*, *velmi*. Tato shoda pořadí by již mohla ukazovat na obecnou sémantiku těchto výrazů, nikoli pouze sémantiku konkrétního spojení *modif. + vysokál/malá*. Opět je ovšem třeba statisticky ověřit, s jakou jistotou můžeme vztah průměrů hodnot vykázaný vzorkem považovat i za vlastnost populací existujících dokladů, z nichž byl vzorek získán.

Tabulka č. 4 uvádí hodnotu  $p$ , jež je výsledkem t-testu pro každou dvojici vágních skalárních modifikátorů ve spojení s adj. *malá*:

<b>docela</b>	8,24451 %				
<b>poměrně</b>	2,05219 %	45,23538 %			
<b>dost</b>	1,86005 %	37,54827 %	88,77671 %		
<b>hodně</b>	0,00604 %	0,2244 %	1,57736 %	3,169 %	
<b>velmi</b>	0,0012 %	0,0074 %	0,01251 %	0,09353 %	4,58326 %
	<b>celkem</b>	<b>docela</b>	<b>poměrně</b>	<b>dost</b>	<b>hodně</b>

Tabulka 4 – T-test dvojic souborů hodnot získaných pro jednotlivá spojení *modifikátor + adj. malá*. Zaokrouhleno na 5 desetinných míst.

Mnohé rozdíly byly tedy opět shledány statisticky významnými. Co se týče dvojic modifikátorů, pro něž jsme zvláště žádali empirické ověření introspektivně málo zřetelné sémantické odlišnosti, byl výsledek  $p < 5\%$  z první části výzkumu zopakován u těchto párů spojení: *dost malá* × *velmi malá*; *celkem malá* × *poměrně malá*. V případě dvojic *poměrně malá* × *dost malá*; *docela malá* × *dost malá* nebyl nyní oproti původnímu  $p < 5\%$  shledán ani okrajově významný rozdíl. Navzdory nižšímu počtu dokladů byl v užití dvojice *dost malá* × *hodně malá* shledán statisticky významný rozdíl ( $p < 5\%$ ) oproti původnímu výsledku statisticky okrajově významnému. Pro dvojici *celkem malá* × *docela malá* byl oproti původní okrajové statistické významnosti nalezen výsledek hranici okrajové významnosti těsně nesplňující. Oproti původnímu statisticky nevýznamnému výsledku byl u dvojice *hodně malá* × *velmi malá* nalezen statisticky významný rozdíl ( $p < 5\%$ ). Podobně jako při prvním testování nebyl nalezen ani okrajově

významný rozdíl pro dvojici *docela malá* × *poměrně malá*. U dvojic, kde byl nalezen statisticky významný rozdíl, můžeme opět usoudit, že průměry populací se liší v témže směru, v němž se liší průměry získaných vzorků.

Přistupuji nyní k interpretaci výsledků obou částí výzkumu. Zásadní je otázka, zda můžeme na základě dat o užívání spojení *modifikátor + vysoká/malá* v kontextu udávání lidské tělesné výšky činit závěry nejprve o významu těchto spojení obecně, bez ohledu na kontext, dále pak o významu vágních skalárních modifikátorů samých. Považuji obojí zobecnění za přijatelné. Lze se těžko domnívat, že by synsémantika, jakými jsou vágní skalární modifikátory, měla více různých významů obvyklých v různých kontextech. K usouzení na význam samotných modifikátorů nás podle mého soudu opravňuje to, že vztahy ve dvojicích spojení *modif. A + adj.* a *modif. B + adj.* se ukázaly být značně stabilní: U 9 takových dvojic (např. *dost + adj.* × *velmi + adj.*) se v obou částech výzkumu statisticky průkazně ukázalo, že spojení adjektiva s jedním modifikátorem (např. *velmi*) udává průměrně hodnotu bližší příslušnému pólu škály než spojení s modifikátorem druhým (např. *dost*); u dalších 6 dvojic takový vztah nebyl shledán s plnou statistickou průkazností v obou částech výzkumu. Pro žádnou dvojici nebyl však ani s okrajovou statistickou významností shledán v jedné části výzkumu vztah opačný než v druhé části. To svědčí ve prospěch teze o nezávislosti významu skalárních modifikátorů na tom, jaké autosémantikum modifikují.

S vědomím, že nejpodstatnější sémantické informace, jež výzkum přinesl, byly již podány samým uvedením výsledků statistického testu získaných dat a že větší jistotu v daném zobecnění může přinést jen další, širě založený empirický výzkum, navrhuji následující **interpretaci výsledků**: za prokázanou považuji vzájemnou sémantickou odlišnost (včetně orientace rozdílu) u těch dvojic vágních skalárních modifikátorů, pro něž byl jak ve spojení s *adj. vysoká*, tak ve spojení s *adj. malá* shledán statisticky alespoň okrajově významný rozdíl. *Tabulka č. 5* uvádí statistickou významnost rozdílů shledaných v obou částech výzkumu pro jednotlivé dvojice modifikátorů, kombinace splňující podmínku pro konstatování sémantické odlišnosti jsou vyznačeny tučně.

<b>docela</b>	ON				
<b>poměrně</b>	VV	NN			
<b>dost</b>	VV	VN	VN		
<b>hodně</b>	VV	VV	VV	OV	
<b>velmi</b>	VV	VV	VV	VV	NV
	<b>celkem</b>	<b>docela</b>	<b>poměrně</b>	<b>dost</b>	<b>hodně</b>

**Tabulka 5** – Statistická významnost rozdílů nalezených pro jednotlivé dvojice vágních skalárních modifikátorů v obou částech výzkumu. (V – statisticky významný rozdíl; O – statisticky okrajově významný rozdíl; N – statisticky nevýznamný rozdíl).

Výsledky tudíž interpretuji takto: empirický výzkum ukázal, že pro následující dvojice vágních skalárních modifikátorů platí, že první uvedený modifikátor udává na škále průměrně stupeň bližší hodnotě kontextového standardu (S) a vzdálenější příslušnému pólu škály (P), než modifikátor druhý. Jde o dvojice: *celkem* × *poměrně*; *celkem* × *dost*; *celkem* × *hodně*; *celkem* × *velmi*; *docela* × *hodně*; *docela* × *velmi*; *poměrně* × *hodně*; *poměrně* × *velmi*; *dost* × *hodně*; *dost*



× *velmi*. U dalších dvou dvojic nebyla stanovená podmínka pro konstatování sémantického rozdílu splněna jen těsně: *celkem* × *docela* ( $p_1 < 8\%$ ;  $p_2 = 8,24\%$ ); *hodně* × *velmi* ( $p_1 = 10,4\%$ ;  $p_2 < 5\%$ ) – je pravděpodobné, že by u nich nový, rozsáhlejší výzkum taktéž shledal sémantický rozdíl. U zbylých tří dvojic zůstala podmínka pro konstatování sémantického rozdílu výrazně nesplněna: *docela* × *poměrně* ( $p_1 = 45,24\%$ ,  $p_2 = 24,61\%$ ), *docela* × *dost* ( $p_1 < 5\%$ ,  $p_2 = 37,55\%$ ), *poměrně* × *dost* ( $p_1 < 5\%$ ;  $p_2 = 88,78\%$ ) – provedený empirický výzkum v nich neprokázal existenci sémantického rozdílu.

Získaná data tedy neumožnila beze zbytku určit pořadí daných 6 modifikátorů podle blízkosti jimi udávané průměrné hodnoty příslušnému pólu škály – s dostatečnou jistotou ukázala pouze „fragmenty“ tohoto pořadí. Pokud bychom předpokládali sémantickou odlišnost i ve dvojicích *celkem* × *docela*, *hodně* × *velmi*, u nichž zjištěný rozdíl jen těsně nedosáhl hranice statistické významnosti, bylo by možno pořadí rekonstruovat do následující podoby:

$$S < \textit{celkem} < \overset{\textit{poměrně}}{?} < \overset{?}{\textit{docela}} < \textit{dost} < \textit{hodně} < \textit{velmi} < P$$

Předložené výsledky výzkumu se týkají *průměru* hodnot udávaných jednotlivými modifikátory. Postihují tedy pouze část jejich sémantiky: chybí především informace o rozložení udávaných hodnot. Pro přesnější údaje v tomto směru by byl nezbytný širší založený empirický výzkum se zapojením složitějších statistických metod. Prezentovaný výzkum však přesto umožnil některé netriviální, empiricky podložené závěry o sémantických vztazích mezi vágními skalárními modifikátory v češtině. Bylo tak demonstrováno, že navrženou metodou je minimálně u některých výrazů možno empiricky ověřit či korigovat naše introspektivní poznání vágního významu.

*Vděčím Janu Volínovi a Janu Hajičovi ml. za poskytnuté konzultace ke statistickému zpracování získaných dat. Za případné chyby pochopitelně zodpovídám sám. Děkuji rovněž Evě Lebeckové, vedoucí mé bakalářské práce, jejíž empirické výsledky jsem v této stati shrnul.*

## LITERATURA

- Geurts, B. – Nouwen, R.: „At least“ et al.: the Semantics of Scalar Modifiers. *Language*, 83, 2007, s. 553–559.
- Karlík, P.; Nekula, M.; Pleskalová, J. (eds.): *Encyklopedický slovník češtiny*. NLN, Praha 2002.
- Keefe, R.: *Theories of Vagueness*. Cambridge University Press, Cambridge 2000.
- Labov, W.: The Boundaries of Words and their Meanings. In: Aarts, B. a kol. (eds.): *Fuzzy Grammar: A Reader*. Oxford University Press, New York 2004, s. 67–90.
- Neustupný, J. V.: On the Analysis of Linguistic Vagueness. In: Aarts, B. a kol. (eds.): *Fuzzy Grammar: A Reader*. Oxford University Press, New York 2004, s. 341–350.
- Ocelák, R.: *Vágní význam a jeho empirické modelování – případ vágních skalárních modifikátorů v češtině*. Nepublikovaná bakalářská diplomová práce, FF UK, Praha 2010.
- Veselý, V.: Lexikální sémantika neurčitých kvantifikátorů v současné češtině. *Slovo a slovesnost*, 70, 2009, s. 175–192.
- Volín, J.: *Statistické metody ve fonetickém výzkumu*. Epoque, Praha 2007.

---

*Oddělení gramatiky  
Ústav pro jazyk český  
Akademie věd ČR, v.v.i.  
<radioc@seznam.cz>*