

Jiří Dan: Psychologická analýza textu.

Učební text k předmětu Pedagogicko-psychologická diagnostika. PřF MU, Brno 2008/2009

Analýza textu má v sociálních vědách a v psychologii dlouhou tradici. V psychologii ovšem dlouhou dobu stála ve stínu jiných metod, jako např. introspekce, pozorování nebo fyziologických měření.

Od počátků psychologie jako empirické vědy do poloviny 20. století byly v základních učebnicích psychologie věnovány kapitoly analýze řeči a analýze psaného slova v souvislosti s chápáním pojmů a vytvářením představ. Tak např. Rostohar (1950) se zabývá funkčním významem pádů, v rámci obecného psychologického rozboru tedy analyzuje co znamená, když je pojem užit v určitém pádu. Například nominativ znamená pojem, který je ve větě ve funkci podmětu nebo přísudku. Tyto úvahy byly východiskem pro analýzu projevů emocí v textu psaném a v řeči. Problém byl definován jako identifikace způsobů vyjadřování citového postoje mluvčího. K často citovaným statím reprezentující mezioborový přístup patří např. příspěvek Miroslava Grepla, Citově motivované aktualizace v rovině modální stavby výpovědi (1966).

Analýza textu je systematická redukce textu (nebo řetězce symbolů) na standardizovaný soubor statisticky manipulovatelných symbolů, který reprezentuje přítomnost, intenzitu nebo četnost sociálně-vědeckých relevantních znaků (Mehl, 2006). Tato definice výslovně zahrnuje kvalitativní i kvantitativní metody analýzy.

V oblasti kvalitativní psychologické analýzy textu převládá doposud „manuální“ přístup. Můžeme přitom vysledovat dva směry. V rámci jednoho psycholog sleduje styl psané či mluvené řeči a pokouší se identifikovat slova nebo myšlenky vymykající se s užitého stylu a tyto interpretovat ve smyslu možného vnitřního konfliktu, skryté nebo skrývané emoce, případně trvalejších osobnostních zvláštností. Při analýze rukou psaného textu je možno toto hledisko doplnit o sledování odchylek v grafickém projevu. Určité slovo může být napsáno velkým tlakem, větším písmem atp. Při individuální diagnostice je však třeba závěry formulovat velmi obezřetně, vždy se znalostí dalších psychologických dat, protože jeden znak, odchylka od „běžného“ projevu může mít více významů, stejně tak jako u projevů neverbální komunikace.

Druhý směr chápe psaný text jako materiál, který je možno analyzovat, nakolik se do něho promítají dojmy člověka z jeho osobního světa a jeho osobnosti. Jedinec formuluje text nebo odpovídá na otázky podle smyslu, který pro něj tato situace má, a podle toho, co cítí během této odpovědi. Při následné analýze je možno vyslovovat hypotézy o nevědomých sklonech, přáních, tendencích. Zvláštním případem je analýza tzv. chybných výkonů v psané nebo mluvené řeči. Jsou popsány chybné výkony při tichém čtení nebo v psaných projevech. Ne vždy je lze interpretovat jako projev nevědomých mechanismů obrany proti úzkosti. Člověk bez psychologického vzdělání může zvyšovat svoji vnímavost na chybné výkony své a druhých osob a pokusit se je začlenit mezi metody sebepoznání a poznávání druhých.

Klasifikaci současných kvantitativních přístupů k analýze textu uvádí přehledně Mehl (2006).

Manuální tématická obsahová analýza v psychologii platila dlouhou dobu za zlatý standard analýzy textů a je dnes s úspěchem užívána, například při kódování odpovědí v projektivních testech. Častá kritika nedostatečné spolehlivosti tohoto přístupu většinou není oprávněná. Manuální metody přece jen jednu praktickou nevýhodu mají: je nutné zaškolení a kalibrace posuzovatelů a dodatečně ruční kódování textového materiálu, což představuje velké nároky na čas a zdroje.

Klasifikace přístupů k analýze textu

	dimenze	popis	příklad užití
metoda	manuální	kódování provádějí zaškolené osoby	obsahová analýza odpovědí na otevřené otázky
	s podporou PC	kódování provádí počítač	automatická identifikace referencí o sobě v emailu (např. „já“, „mne“)
záměr	instrumentální	cílem je stanovení výskytu lingvistických znaků	stanovení četnosti regresivních fantazií v psychodynamickém interview
	zobrazující	cílem je zobrazení struktury textu	stanovení síly asociací v autobiografickém eseji (např. rodina, hněv, agrese)
přístup	tématický	identifikace témat nezávislá na kontextu	stanovení četnosti slov indikujících strach (např. „nervózní) při spontánním užívání řeči osobou
	sémantický	identifikace sémantických vztahů mezi tématy	stanovení, jak často jsou v deníku obsažená agresivní slova asociována s pisatelem nebo třetími osobami
šíře výběru	specifický	selektivní výběr úzce definovaných lingvistických znaků	analýza protokolů z terapie z hlediska stupně abstrakce sebereflexí
	široký	vytvoření kompletního lingvistického profilu	analýza řeči pacientů při terapeutickém prvním rozhovoru (např. z hlediska slovní zásoby a výběru slov)
zaměření	obsah řeči	analýza textu z hlediska znaků jeho obsahu	analýza konverzací na chatu z hlediska dominantního tématu (např. povolání, vztahy, hobby)
	styl řeči	analýza textu z hlediska stylistických znaků	analýza komplexnosti větné stavby ve školní slohové práci jako indikátor vývojových procesů

Po roce 1970 začala nová éra metod analýzy textů s nástupem využití počítačů. Pomocí automatizace kódování je možno zpracovávat velká množství textu. Textové zdroje se digitalizují. V současnosti většina metod kvantitativní analýzy textů využívá počítačů, jsou instrumentální v intenci a tématické v přístupu. Technicky jsou založeny na jednoduchém principu, že každé slovo analyzovaného textu je porovnáno s interním slovníkem a výskyt určitých témat je určen jednoduchým sečítáním. Jde o instrumentálně sémantickou analýzu textů. Na tomto principu pracují vyhledávače – hledají zadaná slova, klíčová slova, řetězce slov. Nejeefektivnější se ukazuje vyhledávání termínů, nejčastěji dvouslovných (adjektivum + substantivum). Časově efektivní je například vyhodnocování studentských prací, esejů podle obsažených klíčových slov, jak o tom referují např. Shermis a Burnstein (2003). I v nejnovějších programech tohoto typu však nelze vyloučit chybné vyhodnocení.

Poslední generace metod k analýze textů je převážně zobrazující (representational) a sémantická v přístupu. Technicky jsou tyto programy založeny na principu kovariance.

Hledají se kontexty, ve který se slova vyskytují. Struktura textu je určena tím, že je spočítána síla asociací mezi klíčovými pojmy (tj. předem danými slovy). Nejdříve se tyto postupy využívaly pro analýzu korpusů, tedy digitalizovaných databází té které řeči (např. slovenštiny, viz <http://korpus.juls.savba.sk/>).

S využitím speciálního software je možno například zjistit s jakým jiným plnovýznamovým slovem se v jazyce nejčastěji vyskytuje konkrétní zadané slovo. Například v češtině je slovo „sen“ nejčastěji spojeno s adjektivem „zlý“. V psychologii jsou dosud užívány zobrazující sémantické programy pro analýzu textů převážně pro explorativní účely. K. A. Neuendorf (2008) nabízí webové stránky, kde najdeme obsáhlou příručku k obsahové analýze textu.

Přednosti a slabiny přístupů k analýze textů.

Většina dnes užívaných metod k analýze textů je založena na jednoduchém principu frekvence, výskytu slov. Není potom divu, že studie často vyvolávají kritiku. Kritici poukazují na to, že analýza jednotlivých slov, tím že nemá na zřeteli gramatické informace, nerozlišuje různé výroky. Mehl uvádí příklad: der Hund biss den Mann (pes kouše muže). Tento výrok obsahuje stejná slova jako der Mann biss den Hund (muž kouše psa). Chybně klasifikují významy slova, které jsou specifické pro kontext: „Jsem do tebe zblázněný“, „ty ses zbláznil“. Programy založené na analýze jednotlivých slov neumí přiměřeně zohlednit ironické výroky, „to jsi tomu zase dal“, nebo metaforickou řeč, „cítí se jak v sedmém nebi“. Také komplexní algoritmy vyhodnocení novějších metod žel nezabrání tomu, aby v jednotlivých případech program neudělal chybu.

Přednosti před jinými metodami spočívají v ekonomičnosti získávání dat, v objektivnosti, vyhovující spolehlivosti a validitě (zjišťujeme to, co uvádíme, že zjišťujeme). Enormní rychlost zpracování dat umožňuje zkoumat extrémně rozsáhlé vzorky. Tím se odstraní vliv rušivých faktorů. Objektivnost (nezávislost výsledku na osobě examinátora) vyhodnocení pozitivně působí na reliabilitu (stabilitu výsledku, přesnost „měření“) a umožňuje porovnání výsledků mezi studii.

Jako zdroje textů pro analýzu mohou sloužit digitalizované varianty původně „papírových“ předloh, e-mailové zprávy a další komunikace na internetu. Výhodou je rovněž neomezená analýza archivovaných dat. Zatímco dotazník umožňuje jen dodatečnou analýzu položek, umožňuje analýza textů po rekódování hlubší vyhodnocení aspektů, které se dodatečně ukázaly jako důležité, i když původně nebyly sledovány.

Literatura:

Grepl, M.: Citově motivované aktualizace v rovině modální stavby výpovědi. SPFFBU, Brno, 1966, A 14.

<http://korpus.juls.savba.sk/>

<http://academic.csuohio.edu/kneuendorf/content>: Neuendorf, K.A.: The Content Analysis Guidebook 1.4.2008

Mehl, M.R.: Textanalyse. In: Franz Petermann, Michael Eid (Hrsg.): Handbuch der Psychologischen Diagnostik. Hogrefe, Göttingen 2006, s. 196-202.

Rostohar, M.: Psychologie jako věda o subjektivní skutečnosti. Společnost pedagogického musea v Brně, Brno 1950.

Shermis, M. D., Burstein, J. C.: Automated Essay Scoring. A Cross-Disciplinary Perspective. Lawrence Erlbaum Associates, Mahwah, New Jersey 2003