

Alena Vodáková; František Vodák
Příspěvek k problému validizace

Kybernetika, Vol. 5 (1969), No. 3, (212)--222

Persistent URL: <http://dml.cz/dmlcz/125847>

Terms of use:

© Institute of Information Theory and Automation AS CR, 1969

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*
<http://project.dml.cz>

Příspěvek k problému validizace

ALENA VODÁKOVÁ, FRANTIŠEK VODÁK

V práci je řešen problém zvyšování validity výsledků společenskovešedního výzkumu redukcí chyby „neporozumění informačním položkám“, která je vázána na verbální interakci v komunikačních situacích pozorovacích technik (interview, dotazník, přímé pozorování.) Je ukázáno, že tento proces dílí validizace lze ztotožnit s procesem výběru vhodného kódu pro definovaný kanál.

1. ÚVOD

Stále silnější tendence společenských věd k co možná největší objektivnosti a exaktnosti si vynucuje jejich matematizaci, či přesněji fyzikalizaci. Jak dalece je tento trend účelný, prověří v konkrétních případech výzkumná praxe. Zůstává však faktem, že některé postupy běžné v oboru kybernetiky a teorie informace se stávají neodmyslitelnou součástí metodiky společenských věd.

V rámci této práce hovoříme o sociologii, ale problém chybovosti zpráv získaných při interakci dvou lidských subjektů tento rámec značně přerůstá a je vlastní každému společenskovešednímu výzkumu, při kterém k této interakci dochází.

Značná část sociologických výzkumů realizuje svou funkci pomocí pozorovacích technik, jako je dotazník, anketa, interview a přímé pozorování. Bezprostředním cílem těchto technik je získání empirických informací o zkoumaných jevech. Oblast použití pozorovací techniky proto můžeme v rámci celého výzkumu vymezit pomocí popisu procesů vedoucích k získávání těchto informací. Volba konkrétních procesů je vždy determinována výzkumným záměrem. Výzkumný záměr jako determinant techniky má dva aspekty: meritorní a realizační. Jinými slovy výzkumný záměr můžeme vyjádřit operacionistickou definicí sociologických pojmů a kategorií, které chceme určitou technikou ověřovat a zkoumat, nebo jej můžeme definovat jako snahu o získání reálného obrazu zkoumané skutečnosti. Tyto dva aspekty se promítají do volby procesů pozorovací techniky jako požadavky adekvátnosti a validity. Oba požadavky se vzájemně podmiňují, to znamená, že je stejně bezcenné

přesné dodržení merita záměru při postupech, které nezaručují validitu informací, jako volba validních postupů pro zkoumání jevů, které se nekryjí s kategoriemi stanovenými výzkumným záměrem. Proto dále vymezíme „validizační sféru pozorovacích technik“, která je širším, obecným polem zájmu této práce, pomocí popisu procesů získávání informací s přihlédnutím k oběma uvedeným aspektům:

Obecně lze souhrn procesů vztahujících se ke každé pozorovací technice znázornit chronologickým řetězcem, jehož jednotlivými články jsou účelově interpretované pracovní postupy.

Nultým článkem našeho řetězu nechť je apriorní volba syntetických celků ve smyslu sociologických pojmů, kategorií či typů, jejichž zkoumání je záměrem výzkumu. Každý z těchto celků je definován strukturálními složkami, které jsou pro něj z hlediska výzkumného záměru určující. (Např. sociologická kategorie „neformální pracovní role“ může být vyjádřena třemi složkami: plnění pracovního úkolu, pracovní autorita a vztah ke spolupracovníkům.) Strukturální složky je při dalším zpracování pracovního modelu nutno vyjádřit (podle logiky terénu a technických možností sběru informací) jako logicko-informační položky, jejichž zjišťování se pak stává bezprostředním cílem jednotlivých akcí pozorovací techniky.

Prvním článkem nechť je tedy přiřazení jedné nebo více logicko-informačních položek (dále jen položek) jednotlivým strukturálním složkám a stanovení konfigurací položek tak, aby odpovídaly syntetickým celkům. (Např. ke strukturální složce „pracovní autorita“ lze přiřadit položky 1. Kdo se s ním chodí radit o pracovních úkolech, 2. Kdo by s ním rád dlouhodobě spolupracoval, 3. Kdo by ho zvolil za svého vedoucího. Tyto položky spolu s položkami přiřazenými k ostatním dvěma složkám tvoří pak konfiguraci příslušející ke kategorii „neformální pracovní role“.) Jedna a táž informační položka může být součástí více konfigurací, a tím může sloužit k vyjádření více syntetických celků. (Např. položka „Kdo by s ním rád spolupracoval“ může být přiřazena i ke složce „vztah ke spolupracovníkům“ a může být zároveň i součástí konfigurace, která náleží ke kategorii „pracovní pozice“.) Je nutno poznamenat, že teprve takto vymezený první článek může být počátkem validizační sféry pozorovacích technik. Nultý článek je nutno chápat šířeji, než pouze jako východisko pozorovacích technik. Je však vymezen takovým způsobem, že jej validizační postupy ve sféře pozorovacích technik mohou, jak se ukáže níže, zasáhnout.

Druhým článkem našeho řetězu nechť je získání informací při průzkumu v terénu. Tento článek lze většinou rozložit do tří bodů:

- a) Konstrukce otázek nebo pozorovacích kategorií tak, aby implicitně nebo explicitně obsahovaly informační položky.
- b) Klázení otázek nebo pozorování chování a interakcí.
- c) Záznam odpovědí ve smyslu přiřazení informačních položek zkoumané skupině lidí.

Za třetí článek řetězu a také za konec naší validizační sféry považujeme inverzní přiřazení konfigurací položek syntetickým celkům.

V kterémkoli článku takto vymezené validizační sféry může vzniknout odchylka od výzkumného záměru, kterým zde rozumíme získání reálného obrazu zkoumané skutečnosti tím, že dojde k narušení validity informací, tj. k chybnému přiřazení informačních položek skupině lidí. Příčinou této nežádoucí odchylky může být systematická chyba, tj. chyba vyplývající z přímého či zprostředkovaného, vědomého či nevědomého zásahu lidského individua jako článku konkrétního komunikačního aktu do procesu transformace přenášené informace, nebo chyba nástrojů a pomůcek měření spojená s jejich neadekvátní volbou ve smyslu zajištění měřitelnosti a zajištění spolehlivosti měření. V procesu jednotlivých komunikací dochází ke kumulaci a vzájemnému podmiňování jednotlivých chyb, čímž vzrůstá celková chyba pozorování (nepředpokládáme-li validizační zákrok) a tudíž i odchylka od výzkumného záměru. Principiálně můžeme tedy hovořit o chybě podmiňující a chybě podmíněné v souvislosti se vzájemnou vazbou jakýchkoli dvou chyb výše vymazané validizační sféry pozorovacích technik.

Záměrem tohoto příspěvku je zkoumání jednoho typu takové vazby dvou na sobě závislých chyb. V intencích tohoto záměru si vymezíme chybu podmiňující jako jednu ze systematických chyb způsobených přímým či nepřímým zapojením lidského jedince do komunikačního aktu. Za chybu podmíněnou budeme považovat kauzálně navazující chybu získaných syntetických celků, která už přímo zvyšuje celkovou odchylku výsledků výzkumu od záměru výzkumu ve smyslu snížení validity těchto výsledků.

Z možných chyb podmiňujících zvolme tu, která je vázána na verbální porozumění jednotlivým informačním položkám, implicitně či explicitně obsažených v otázkách (při technice dotazníku, ankety a interview), či v pozorovacích kategoriích (při technice přímého pozorování). Verbálním neporozuměním rozumíme jakkoli subjektivně motivované nepochopení smyslu informační položky, je-li vyjádřena pomocí nejednoznačných slovních elementů. Lidským článkem, u kterého může k neporozumění dojít, může zásadně být respondent, tazatel, pozorovatel i interpret (eventuálně další zprostředkující jedinec). Proto celý validizační postup, který zde bude popsán, je zásadně aplikovatelný na kteroukoli z uvedených pozorovacích technik, přičemž jej lze použít u chyby neporozumění kteréhokoli lidského článku příslušných komunikačních aktů. Pro větší terminologickou konkrétnost budeme validizační postup vysvětlovat v aplikaci na techniku individuálního interview a pro metodologickou konkrétnost vybereme z možných lidských článků, u kterých může dojít k neporozumění, pouze jeden, a to v případě interview nejevidentnější článek – respondenta. Chyba neporozumění, vázaná na respondenta, objeví se v následujícím chronologickém řetězu, kterým lze znázornit podle obecného verbálního schématu na počátku tohoto úvodu umístění chyb v postupech při technice interview:

- I. chyba způsobená verbálním nepochopením smyslu výchozího systému kategorií při jejich logickém strukturování do informačních položek a jejich konfiguraci (což vede k apriorní chybě položek);

- IIa. chyba způsobená verbálním neporozuměním informačním položkám při jejich převádění do otázek pro interview (což vede k apriorní chybě otázek);
- IIb₁. chyba způsobená verbálním nepochopením otázek záznamního listu tazatelem (což vede k chybě jeho vyjádření ve styku s respondentem);
- IIb₂. chyba způsobená verbálním nepozorováním respondenta otázkám (což vede k chybě jeho odpovědi);
- IIc. chyba způsobená verbálním nepochopením odpovědi respondenta tazatelem (což vede k chybě jejich záznamu);
- III. chyby způsobené verbálním nepochopením záznamu odpovědi interprety (což vede k chybě přiřazení odpovědi informačním položkám a později kategoriím).

Z uvedených chyb verbálního neporozumění v procesu interview budeme se tedy zabývat pouze chybou IIb₂. To znamená, že budeme předpokládat, že otázka kladená tazatelem respondentovi není dosud zatížena žádnou chybou neporozumění, že tuto chybu vyvolá zásah respondenta, a že v následném procesu už nedojde k jinému dalšímu verbálnímu neporozumění.

Z takto vymezené podmiňující chyby verbálního neporozumění informačním položkám vyplyne podmíněná chyba výsledných syntetických kategorií, které pomoci interview získáme.

Obecně můžeme zařadit postup identifikace a odstraňování obou jmenovaných chyb mezi procesy validizace. Pro účely této práce můžeme procesy validizace obecně charakterizovat dvěma základními kroky, které mohou být realizovány v různých alternativách a chronologických řazeních:

Za první validizační krok (i ve smyslu historickém) lze považovat aposteriorní vyjádření odchylky od skutečnosti pomocí analýzy vzniklých chyb. Přitom lze směřovat k logickověcným důkazům validity nebo k exaktnímu vyjádření chyb pomocí matematického aparátu.

Na poznatky získané v prvním kroku lze navázat v druhém validizačním kroku, který směřuje k apriorní validizaci, tj. k minimalizaci předpokládané odchylky od skutečnosti pomocí apriorního stanovení, zmenšení či vyloučení chyb. Přitom jde vždy – u obou kroků – o dva cíle: Přesnou časovou a místní lokalizaci chyby a o vyjádření velikosti chyby. Ideální současné plnění obou cílů lze na úrovni dosavadních poznatků těžko zajistit. Domníváme se, že jednou z přístupných cest aproximace k tomuto dualitnímu cíli je kombinace empirických šetření jako prostředku pro pravděpodobnostní vyjádření lokalizovaných chyb na jedné straně a matematických postupů k vyjádření souhrnných, komplexních či výsledných chyb v určitém časovém kroku výzkumu na straně druhé.

2. SCHÉMA KOMUNIKAČNÍHO KANÁLU

Vycházejíce z úvodu, budeme obecně charakterizovat výzkumný záměr počtem přesně definovaných syntetických celků (kategorií, typů apod.), které mají sloužit

k modelování zkoumaného systému, situace atd. Označme tyto syntetické celky x_1, x_2, \dots, x_N . Každý z x_q ($q = 1, \dots, N$) nahradíme přesně určenou konfigurací položek, kterou budeme zkoumat v interakci respondent – tazatel. Označme tyto položky $1, \dots, a - 2$ (volba počtu $a - 2$ vyplývá z dalšího) a jejich konfigurace u_q ($q = 1, \dots, N$). Chyby, které zatěžují získané kategorie, jsou podmíněny chybami položek $1, \dots, a - 2$.

V každé otázce, kterou tazatel klade, je explicitně nebo implicitně obsaženo několik položek. Obecně můžeme předpokládat, že na tytéž položky se můžeme ptát explicitně otázkou uzavřenou (např.: „Jste ženat nebo svobodný, rozvedený či vdovec?“), nebo otevřenou otázkou („Jaký je Váš rodinný stav?“), která obsahuje implicitně tytéž položky (ženatý, svobodný, rozvedený, vdovec). Kromě toho můžeme předpokládat, že každou otázku lze hypoteticky nahradit jednou nebo několika dichotomickými formulacemi. (V uvedeném příkladu by šlo o nahrazení čtyřmi dichotomickými otázkami s možnými odpověďmi ano-ne: „Jste ženat?“, „Jste svobodný?“, „Jste ...“) Každá z dichotomických otázek obsahuje dvě položky, jednu explicitně a druhou implicitně (v našem případě 1. dichotomická otázka obsahuje explicitně „jsem ženat“ a implicitně „nejsem ženat“).

Předpokládáme pro jednoduchost pouze dichotomické formulace otázek. Vše, co zde bude napsáno, lze však rozšířit jednoduchým způsobem na otázky obsahující libovolný počet položek. Upozorňujeme ještě, že položky, o kterých hovoříme, nemusí být představovány pouze indikativními slovy, ale mohou to být také výroky, věty atd. (např. položkami mohou být motivace ke vstupu do zaměstnání, jako: blízkost bydliště, vyšší příjem, uplatnění odborné specializace apod.)

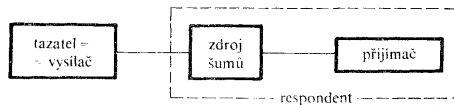
Z úvodu také vyplývá, že z chyb, které zatěžují položky $1, \dots, a - 2$, se budeme zabývat těmi, které jsou způsobeny označením některých položek (popřípadě všech) slovem nebo skupinou slov, která mají význam egocentrických částic ve smyslu Russellové [1]. Jsou to slova, jejichž význam se mění se změnou mluvčího a jeho pozice v čase a prostoru. Pokládáme-li za rozhodující pozici tazatele, u níž zde předpokládáme přímou determinovanost záměrem výzkumu, pak můžeme hovořit o chybě, která vzniká tím, že respondent zamění význam, který danému slovu přiřkládá tazatel, za význam, který přiřkládá tazatel jiné egocentrické částici, která popřípadě i nemusí být obsažena v $1, \dots, a - 2$. Chyba vznikající touto záměnou, bude naší chybou podmiňující. Při dalším zkoumání předpokládáme, že zaměnitelný charakter egocentrických částic determinuje do značné míry formulaci otázek, a že je proto možno aproximativně uvažovat o chybě záměny otázek.

Za předpokladu, že přiřazení u_q k x_q je adekvátní a vyskytuje se pouze podmiňující chyba, kterou jsme nyní definovali, můžeme upřesnit cíle této práce: Tážeme se, jak velkou podmíněnou chybou jsou zatíženy získané konfigurace u_q , a tedy také syntetické celky x_q , nebo kolik u_q (a tedy x_q) můžeme získat s chybou tak malou, jak si určíme, známe-li číselné vyjádření podmiňující chyby.

Jak ukázal např. V Lamser [2], každou situaci, kterou navozuje interview i jiné pozorovací techniky, je možno chápat jako komunikační akt. Aproximativně lze

tyto situace znázornit jako komunikační kanály. Domníváme se, že charakter zá-
měrného rozhovoru jako nástroje výzkumné techniky tuto aproximaci umožňuje.
(To ovšem neznamená, že je stejně vhodný i pro jiné případy společenské komuni-
kace.) Námí zkoumanou situací můžeme tedy znázornit ve dvou krocích jako kanály
na obr. 1 a 2.

V komunikaci na obr. 1 tazatel vysílá otázky, které označíme A_1, \dots, A_{b-1} a které
jsou pro nás abecedou vysílače. Libovolnou otázku A_m může respondent ve důsledku



Obr. 1.

toho, co bylo řečeno, zaměnit za některou další A_k z abecedy vysílače, popřípadě
za A_b , kde $k > b - 1$. Tvoří-li tedy abecedu vysílače pouze otázky, které nás mají
přivést k žádaným položkám, pak abecedu přijímače tvoří kromě těchto také další,
které jsou přípustné podle charakteru egocentrických částic. Označme souhrn
možných otázek pro příjemce jako množinu M a abecedu vysílače jako množinu
 $A \subset M$. Rozdíl množin $M \setminus A$ necht' je pak poslední symbol abecedy přijímače A_b .



Obr. 2.

Toto označení je zavedeno jen pro větší stručnost a nikterak neovlivní obecnost
našich úvah. Stejně tak abychom mohli vybrat z $M \setminus A$ c údajů, označit je A_b, \dots
 \dots, A_{c-1} a zbytek jako A_c , abecedu příjemce by pak tvořily symboly $A_1, \dots, A_{b-1}, \dots$
 \dots, A_c , resp. bychom za abecedu příjemce hypoteticky mohli brát všechny otázky,
které se mohou vyskytnout, jestliže jejich počet je konečné číslo. Tedy abecedu
přijímače tvoří A_1, \dots, A_b . Záměna otázky respondentem je v podstatě mylným
přijetím symbolu přijímačem. Označme pravděpodobnost, že bude přijato A_k ,
bylo-li vysláno A_m jako $g(A_k | A_m)$, $m = 1, \dots, b - 1$, $k = 1, \dots, b$, pak platí
 $\sum_k g(A_k | A_m) = 1$ pro všechna m .

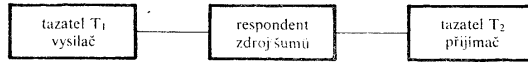
Dále ve shodě s našimi předpoklady v úvodu, je-li vysláno A_m a přijato A_k , pak
tato skutečnost nijak neovlivňuje vyslání a přijetí dalších symbolů. Z toho všeho
vyplývá, že kanál na obr. 1 je diskrétní kanál bez paměti s šumy.

V rámci komunikace na obr. 2 otázky A_k vyvolávají odpovědi ano, resp. ne, které
označíme indexy γ resp. \circ . Tazatel tedy získává položky k_1 nebo k_0 , $k = 1, \dots, b$.
Tento přenos se děje podle předpokladů v úvodu bez poruch. To znamená, že je-li
vysláno respondentem ano, tedy v podstatě k_1 , je také jako ano (a tedy k_1) přijato

tazatelem. Na této skutečnosti nemění nic fakt, že se může jednat o odpověď na otázku, která byla respondentem mylně přijata. To znamená, přijme-li respondent v rámci kanálu na obr. 1 místo otázky A_m otázku A_k , dojde-li tedy k záměně ego-centrických částic, tak v následné odpovědi bude vysláno a objektivně přijato k_1 , nebo k_0 . Subjektivní příjem bude pochopitelně jiný.

Jde tedy o diskretní bezpaměťový kanál bez šumů.

Dále nás zajímá pravděpodobnost, s jakou po vyslání otázky A_m tazatelem obdržíme v respondentově odpovědi skrytou položku k_1 nebo k_0 . Tuto pravděpodobnost



Obr. 3.

označme $w(k_1 | A_m)$ nebo $w(k_0 | A_m)$. Pro jednoduchost předpokládejme, že z hlediska tazatele u každé otázky existuje stejná pravděpodobnost odpovědi ano a ne. Pak

$$w(k_1 | A_m) = w(k_0 | A_m) = \frac{1}{2}g(A_k | A_m),$$

kde $g(A_k | A_m)$ bylo definováno výše.

Z toho, co bylo řečeno o charakteru otázek a našich komunikačních kanálů vyplývá, že můžeme provést následující substituci. Označme položky $1_0, 1_1, \dots, b_0, b_1$ jako $1, \dots, a$, kde $a = 2b$. Komunikační kanály na obr. 1 a 2 můžeme nahradit kanálem na obr. 3, kde vysílač T_1 (tj. tazatel v čase t_1) vysílá symboly z abecedy $1, \dots, a - 2$ a přijímač T_2 (tj. tazatel v čase t_2) přijímá symboly z abecedy $1, \dots, a - 1, a$.

Pravděpodobnost přijetí j , bylo-li vysláno i , je

$$(1) \quad w(j | i) = \frac{1}{2}g(A_k | A_m)$$

$k = j/2$ nebo $j/2 + \frac{1}{2}$ je-li j liché, $m = \frac{1}{2}$, nebo $i/2 + \frac{1}{2}$ je-li i liché, a platí $\sum_j w(j | i) = 1$, kde $i = 1, \dots, a - 2$. Také v dalším textu vždy $j = 1, \dots, a$, $i = 1, \dots, a - 2$.

Jde tedy o diskretní bezpaměťový kanál s šumy a s abecedami $1, \dots, a - 2; 1, \dots, a$ u vysílače a přijímače.

Zavedme předpoklad, že počet položek, ze kterých konstruujeme jednotlivé konfigurace u_q , je vždy stejný a roven n pro všechna $q = 1, \dots, N$. Tento požadavek je v konkrétním případě lehce splnitelný, zařadíme-li do konfigurace jednu nebo více redundantních položek, tj. položek indiferentních vůči tvořeným kategoriím. Je-li např. $a - 2 = 5$ a $n = 4$, pak z hlediska žádané konfigurace $(1 \ 2 \ 3 \ 3) \equiv (1 \ 2)$, je-li položka 3 redundantní.

Máme-li počet položek v abecedě příjemce a vysílače $a, a - 2$, pak jsou-li předávány konfigurace délky n , existuje celkem a^n způsobů, kterými tyto konfigurace mohou být přijaty. Tážeme-li se, kolik konfigurací získáme s chybou nejvýše λ (což

jsme vymezili jako úkol práce na začátku kapitoly), je to totéž, jako bychom se ptali, kolik N konfigurací z a^n můžeme vybrat tak, aby po vyslání u_q byla přijata u_0 blíže u_q , než všem ostatním $N - 1$ konfiguracím s pravděpodobností nejméně $1 - \lambda$. Necht' je D_q nějaká podmnožina a^n taková, že je-li vysláno u_q a přijato u_0 , pak $P\{u_0 \in D_q\} \geq 1 - \lambda_q$. Pak výše položená otázka je totožná s otázkou, na kolik N nepřekrývajících se podmnožin D_1, \dots, D_N můžeme rozložit a^n tak, aby po vyslání libovolné u_q ležela přijatá konfigurace v podmnožině D_q s pravděpodobností ne menší než $1 - \lambda$, kde $\lambda = \max_q \lambda_q$. Můžeme shrnout:

Problém zvyšování validity syntetických celků, zatížených chybou, o které hovoříme, je totožný s výběrem vhodného kódu pro náš kanál, neboť postup, který jsme naznačili, je totožný s tvorbou kódu (n, N, λ) , kde počet konfigurací N udává objem kódu a λ je chybou kódu.

Otázky, které jsme si kladli na začátku této kapitoly, je nyní možno vyjádřit takto:

- a) máme dáno n a λ , jaké největší N je přípustné?
- b) máme dáno n a N , jaké je λ ?

Tyto typické otázky teorie kódování řešili Shannon [3], Wolfowitz [4] a jiní. Použijeme zde některé závěry jejich prací. Označme objem kódu, který je největší, $N(n, \lambda)$. Z teoremů dokázaných v [4] vyplývá, že

$$(2) \quad \text{ld } N(n, \lambda) = nC + o(n),$$

kde ld je dvojkový logaritmus a term $o(n)$ je určitým způsobem závislý na λ a počtu symbolů abecedy. Nadto je v [4] dokázáno, že

$$(3) \quad \text{ld } N(n, \lambda) = nC + O(\sqrt{n}),$$

kde term $O(\sqrt{n})$ je pro $\lambda > \frac{1}{2}$ pozitivní.

Strassen [5] krom toho dokázal, že

$$(4) \quad \text{ld } N(n, \lambda) = nC - \sqrt{(n)} T(k) + O(\text{ld } n),$$

kde funkce $T(k)$ je v [5] nalezena.

Symbol C v rovnicích (2), (3) a (4) označuje kapacitu kanálu danou rovnicí

$$(5) \quad C = \max_{\pi} [H(\pi) - \sum_i \pi_i H(w(\cdot | i))],$$

kde $w(\cdot | i)$ je vektor se složkami $(w(1 | i), \dots, w(a | i))$. $H(w(\cdot | i))$ je „entropií“ vektoru $w(\cdot | i)$ a platí

$$H(w(\cdot | i)) = -\sum_j w(\cdot | i) \text{ld } w(\cdot | i).$$

π je tzv. pravděpodobnostní vektor, který má $a - 2$ složek π_i , jejichž hodnota je

určena rovnicí (5). π' je vektor, jehož složky jsou určeny jako $\pi'_j = \sum_i \pi_i w(j/i) \cdot H(\pi')$ je opět „entropií“ vektoru π' , která je určena jako $H(\pi') = -\sum_j \pi'_j \text{ld } \pi'_j$.

Kotz [6] dokázal následující výsledek. Nechť $a(>1)$, $b(>1)$ je počet elementů v abecedě vysílače a přijímače. Nechť $c = \min(a, b)$ a ε libovolné číslo $\in (0, \frac{1}{2})$. Pak vždy existuje kód $(n, 2^{n(C-\varepsilon)}, \lambda)$, kde

$$(6) \quad \lambda \leq 2 \exp \{ -n\varepsilon^2 / [g(c) (\text{ld } c)^{2-\delta}] \}.$$

Zde $g(c)$ je pozitivně monotonně klesající funkce c , $\delta = \varepsilon [D h(c) (\text{ld } c)^2]$. Tvary funkcí $g(c)$, D , $h(c)$ jsou v [6] nalezeny a pro c od 2 do 25 spočteny. C je opět kapacita kanálu daná rovnicí (5).

Aplikovány na náš problém nám tedy rovnice (2), (3), (4) umožňují určit největší počet N , pro které bude chyba syntetických celků nejvýše rovna předem zvolenému λ . Tím jsme zodpověděli naši otázku a).

Odpověď na otázku b) nám dala práce [6] a hledanou chybu spočítáme podle rovnice (6).

3. ZÁVĚR

Na základě výsledků této práce můžeme tedy apriorně nebo aposteriorně matematicky vyjádřit vztah mezi počtem syntetických celků a chybou, kterou jsou zatíženy, jestliže komunikační interakce je zprostředkována slovy s charakterem egocentrických částic. V této práci jsme zdroj šumů lokalizovali do osoby respondenta (článek sub IIb₂). Beze změn, nebo s malými úpravami však můžeme náš postup, a tedy také výsledky, vztáhnout ke všem ostatním částem chronologického řetězce, který jsme vymezili na konci kap. I. Z tohoto hlediska mají výsledky zcela obecnou platnost a jejich aplikovatelnost je značně široká.

Pro konkrétní práci sociologa má nesporně největší význam aplikační dovození předložených závěrů, tj. určitý návod na validizační postupy pro sestavování dotazníku, otázek pro interview, atd. Protože však jde o kapitolu, která má s teorií informace a kybernetikou vazbu už silně zprostředkovanou, bude vhodnější publikovat ji na jiném místě. Zde bychom chtěli pouze poukázat na práce Belsona [7] a Nuckolse [8], jejichž postupů by bylo možno s určitými obměnami použít pro empirické zjištění $g(A_k | A_m)$, které v aplikační sféře hraje základní úlohu.

Závěrem bychom chtěli upozornit na omezení, které na úrovni této práce bylo nutné, ale které by mělo být v budoucnu odstraněno. U komunikačních kanálů na obr. 1, 2 a 3 jsme předpokládali, že vyslání a přijetí jakýchkoli symbolů abecedy neovlivňuje další komunikační přenos, předpokládali jsme tedy, že jsou to kanály bez paměti. Je však nasnadě, že v případě konkrétní interakce, ve které vystupuje lidský subjekt, máme co činit s kanály s předhistorií, kde všechny dříve předané symboly a konfigurace ovlivňují následující, nebo alespoň s kanály s konečnou pamětí, kde předání jednoho symbolu ovlivňuje předání určitého počtu dalších.

Matematická teorie těchto kanálů je již rozpracována [4] a mohli jsme tedy čistě teoreticky náš problém řešit za použití těchto předpokladů a výsledků. V případě kanálu s pamětí nebo předhistorií bychom však kromě $g(A_k | A_m)$ museli empiricky zjišťovat ještě jiné funkční a stavové veličiny a naše řešení by se tak velmi vzdálilo od současných reálných možností empirického výzkumu.

(Došlo dne 29. května 1968.)

LITERATURA

- [1] Russell B.: *Human Knowledge: Its Scope and Limits*. Podle sborníku *Logika, jazyk, věda*. Svoboda, Praha 1967.
- [2] Lamser V.: *Komunikace a společnost*. Academia, Praha (v tisku).
- [3] Shannon O. E.: *A mathematical theory of communication*, *Bell System Techn. J.*, 27 (1948).
- [4] Wolfowitz J.: *Coding Theorems of Information Theory*. Springer-Verlag, Berlin-Göttingen-Heidelberg 1961.
- [5] Strassen V.: *Asymptotic expansions in information theory*. *Colloquium on Combinatorial Methode in Probability Theory*, *Matematiks Institute, Aarhus University*, August 1–10, 1962, 74–77.
- [6] Kotz S.: *Exponential bounds on the probability of error for memoryless channel*. *Ann. Math. Stat.*, 32 (1961), 2, 577–582.
- [7] Belson W. A.: *Studies in Readership*. Institute of Practitioners in Advertising, London 1962.
- [8] Nuckols R. C.: *A note on pre-testing public opinion questions*. *Journal of Applied Psychology* 37 (1953), 2, 119–120.

The Contribution to the Problem of the Validation

ALENA VODÁKOVÁ, FRANTIŠEK VODÁK

The article concerns one of the problems of the validation of the results obtained from the field research in social science. It is problem of the reduction of the error of "misunderstanding the informational items" connected with verbal interaction in the frame of research techniques (interview, questionnaire, direct observation). The starting point of the solution is operational categorisation of the preparations and realisations procedures in those techniques. The description of the steps of validation follows. The chronology of communication situations the error arises from is then substituted by the discrete memoryless channel. It is demonstrated that it is possible to identify the process of validation of the research results with the process of the choice of the proper code for the channel defined and to use mathematical solutions of the works [4], [5], [6].

Ing. Alena Vodáková, Sociologický ústav ČSAV, Jilská 1, Praha 1.

František Vodák, prom. fyzik, katedra fyziky FS-ČVUT, Husova 5, Praha 1.