

Kniha je věnovaná využití počítačové lingvistické analýzy českého textu při studiu vztahů mezi osobností člověka a slovy, která používá. Uvádí východiska a výsledky psychologicko-lingvistického výzkumu CPACT (*Computational Psycholinguistic Analysis of Czech Text*, GA ČR 16-19087S), ale také perspektivy pro další výzkum v oblasti studia mezilidské verbální komunikace.



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice



HLEDÁME MEZI LIDMI A SLOVY  
VÝZKUM GAČR 16-19087S



GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY



9 788073 947149

Výzkum CPACT: Počítačová  
psycholinguvistická analýza českého textu

D. Kučera  
et al.

# Výzkum CPACT: Počítačová psycholinguvistická analýza českého textu

Dalibor Kučera, Jana Marie Havigerová, Jiří Haviger,  
Václav Cvrček, Zuzana Komrsková, David Lukeš,  
Tomáš Jelínek, Tomáš Urbánek, Justýna Franková

The publication focuses on documentation of a three-year research project CPACT, *Computational Psycholinguistic Analysis of Czech Text* (CSF grant nr. 16-19087S). It presents theoretical background of the project, methods, research design and selected results. The publication and the research itself show a considerable potential of the use of computational-linguistic analysis, especially in psychological disciplines.



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice



9 788073 947149

CPACT Research: Computational  
Psycholinguistic Analysis of Czech Text

D. Kučera  
et al.

# CPACT Research: Computational Psycholinguistic Analysis of Czech Text

Dalibor Kučera, Jana Marie Havigerová, Jiří Haviger,  
Václav Cvrček, Zuzana Komrsková, David Lukeš,  
Tomáš Jelínek, Tomáš Urbánek, Justýna Franková



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice

# Výzkum CPACT: Komputační psycholinguvistická analýza českého textu

Dalibor Kučera  
Jana Marie Havigerová  
Jiří Haviger  
Václav Cvrček  
Zuzana Komrsková  
David Lukeš  
Tomáš Jelínek  
Tomáš Urbánek  
Justýna Franková

Pedagogická fakulta Jihočeské univerzity v Č. Budějovicích (2018)



GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY



HLEDÁME MEZI LIDMI A SLOVY  
VÝZKUM GAČR 16-190875



Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice

## KATALOGIZACE V KNIZE - NÁRODNÍ KNIHOVNA ČR

Kučera, Dalibor

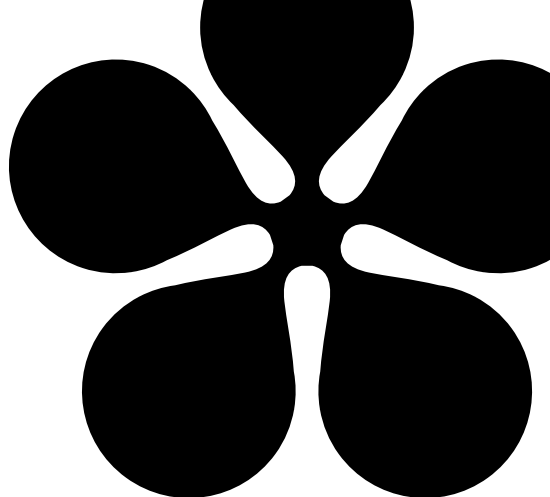
Výzkum CPACT : komputační psycholingvistická analýza českého textu / Dalibor Kučera, Jana Marie Havigerová, Jiří Haviger, Václav Cvrček, Zuzana Komrsková, David Lukeš, Tomáš Jelínek, Tomáš Urbánek, Justýna Franková. -- Vydání první. -- V Českých Budějovicích : Pedagogická fakulta Jihočeská univerzity, 2018. -- (Varia)

ISBN 978-80-7394-714-9

81'23 \* 81'322 \* 159.9.072.5 \* 001.81/.87 \* 811.162.3 \* 81'42 \*  
(048.8:082)

- CPACT (projekt)
- výzkumné projekty -- Česko -- 2011-2020
- psycholingvistika
- počítačová lingvistika
- psychodiagnostika
- čeština
- texty
- jazyková analýza
- kolektivní monografie

159.9 - Psychologie [17]



Dalibor Kučera  
Jana Marie Havigerová  
Jiří Haviger  
Václav Cvrček  
Zuzana Komrsková  
David Lukeš  
Tomáš Jelínek  
Tomáš Urbánek  
Justýna Franková

**Výzkum CPACT:  
Komputační psycholingvistická analýza českého textu**

CPACT Research:  
Computational Psycholinguistic Analysis of Czech Text

Recenzovali:  
prof. PaedDr. Iva Stuchlíková, CSc.  
Mgr. Eva Flanderková, Ph.D.

Vydavatel: Pedagogická fakulta Jihočeské univerzity v Č. Budějovicích  
Sazba a grafická úprava: Dalibor Kučera  
Tisk: Typodesign, s.r.o.  
Vydání první, edice Varia, 2018, 291 s.

Tato kniha je výsledkem badatelské činnosti podporované  
Grantovou agenturou České republiky, reg. č. GA ČR 16-19087S.

ISBN 978-80-7394-714-9

## Shrnutí

Publikace je věnovaná dokumentaci a vybraným výstupům tříletého výzkumného projektu CPACT, Komputační psycholingvistická analýza českého textu (*Computational Psycholinguistic Analysis of Czech Text*, GA ČR 16-19087S). Představuje teoretická východiska projektu, použité metody, výzkumný design a harmonogram projektu, stejně jako projektová data a způsob jejich zpracování. Seznamuje také s vybranými výsledky, kterých bylo v rámci výzkumu CPACT dosaženo.

Výzkum CPACT byl zaměřen na studium souvislostí mezi osobností člověka a slovy, která užívá; zejména na úrovni vztahů lingvistických charakteristik psaného a mluveného textu s výstupy psychologických testů zaměřených na sebeposouzení jedince a jeho posouzení druhou osobou. Pro tento účel byl sestaven třímodulový plán. První modul, zahrnující kvótní soubor participantů, dále označovaný jako P200 (n=200), sestával z dvojic blízkých osob. Tento soubor kvótně reprezentoval českou populaci dle dat ČSÚ k roku 2015 v kategoriích pohlaví, věku a vzdělání. V rámci jednodenních výzkumných sezení zde účastníci poskytli své osobní informace, vytvořili čtyři obsahově odlišné psané texty, absolvovali dva nahrávané polostrukturované ústní rozhovory a vyplnili dvě baterie psychologických testů; první zaměřenou na sebeposouzení, druhou na posouzení druhé osoby ve dvojici. Podklady byly získány v kontrolovaném prostředí, dle předem stanoveného scénáře. Druhým modulem (tzv. P20+) byl pak soubor dalších 72 osob, klinicky diagnostikovaných se specifickou duševní poruchou (anxieta a/nebo deprese), které poskytly celkem čtyři typy psaných textů a vyplnily upravenou testovou baterii.

Veškerá textová data byla následně komputačně (kvantitativně) analyzována v rámci 26 základních lingvistických kategorií pokrývajících 195 proměnných, dále 31 kombinovaných lingvistických proměnných, 9 poměrových lingvistických proměnných a 8 proměnných–indexů, tj. v celkovém počtu 243 proměnných – jedinečných textových parametrů. Zpracování psychologických dat u kvótního souboru vycházelo z výsledků 45 dotazníkových škál (baterie 12 testů) ve verzi sebeposouzení a 36 škál (baterie 9 testů) ve verzi posouzení

druhou osobou, resp. z 25 škál (6 testů) u klinického souboru P20+. Součástí výzkumu byl také třetí modul (tzv. P2), který zahrnoval subjektivní hodnocení výzkumných textů školenými posuzovateli, jehož výstupy byly srovnávány s výše uvedenými daty.

Publikace i samotný výzkum poukazují na značný potenciál ve využití počítačně-lingvistické analýzy zejména v rámci psychologických disciplín (např. v oblasti moderní psychodiagnostiky). Je rovněž komentován aspekt morfoloické, stylistické i sémantické variability různých typů textu. Studie se také věnuje problematice aplikace psychologických testů ve variantách sebeposouzení a posouzení druhou osobou a souvislostem, které s jejich administrací a interpretací výsledků souvisí.

Knihy přináší nejen souhrnné informace a podklady, ze kterých budou vycházet další projektové publikační výstupy, ale poskytuje také detailní dokumentaci, připravenou k využití pro autory dalších studií, kteří se věnují stejné či příbuzné výzkumné oblasti.

*Klíčová slova:* psychologie, lingvistika, počítační, diagnostika, osobnost

## Summary

The publication focuses on documentation of and on selected findings from a three-year research project CPACT, Computational Psycholinguistic Analysis of Czech Text (CSF grant nr. 16-19087S). It presents theoretical background of the project, used methods, research design and time schedule of the project as well as project data and the way of their processing. It also shows some results gained within the frame of the CPACT research.

The CPACT research focused on studying the relationships between one's personality and the words the person uses; especially on the level of relationships between linguistic characteristics of written and spoken texts and results from psychological tests in self and other report variants. For these ends, a three-module plan has been established. First project module includes quota sample of participants, further referred to as P200 (n=200), consisting of pairs of close people was selected. The sample was quota representative of the Czech population according to the data for 2015 of the Czech Statistical Office in the categories gender, age and education. During individual research sessions, the participants gave their personal information, produced four written texts with different contents, took part in two recorded semi-structured interviews and filled in two batteries of psychological tests: self-report and other report on the other person in the pair. The data were collected in a controlled environment, following an in advance given scenario. The second module (the so-called P20+) worked with a sample of other 72 people, clinically diagnosed with a specific mental disorder (anxiety and/or depression), who produced the total of four types of written texts and filled in an adapted test battery.

All the textual data were analysed computationally (quantitatively) in 26 basic linguistic categories (195 variables), 31 combined linguistic variables, 9 proportional linguistic variables and 8 variables–indexes, i.e. the total of 243 variables – unique text parameters. Processing of psychological data in the quota sample was based on the results from 45 questionnaire scales (a battery of 12 tests) in the self-report



variant and 36 scales (a battery of 9 tests) in the other report variant, or of 25 scales (6 tests) in the complementary sample (P20+). The reported research included a third module (the so-called P2) that was based on subjective assessment of research texts by trained assessors whose outputs were compared with the above presented data.

The publication and the research itself show a considerable potential of the use of computational-linguistic analysis, especially in psychological disciplines (e.g. in the area of modern psychodiagnostics). It also comments on aspects of morphological, stylistic and semantic variability of various types of texts. The study also focuses on the issue of application of psychological tests in self-report and other report variants and on administration and interpretation related to their administration and interpretation.

The book brings not only summary information and materials on which further project publication outputs will build but also presents a detailed documentation, ready for use by authors of other studies focusing on the same or related research areas.

*Key words:* psychology, linguistic, computational, diagnostics, personality

# Obsah

Table of contents

<b>1</b>	<b>Úvod</b> Introduction	10
<b>1.1</b>	<b>Vymezení základních pojmů</b> Rationale <i>Dalibor Kučera, Jana M. Havigerová</i>	11
<b>2</b>	<b>Komputační lingvistika a psychologický výzkum</b> Computational linguistics and psychological research	
<b>2.1</b>	<b>Současná východiska komputační lingvistiky a její aplikace</b> Theoretical background of computational linguistics and its application <i>Tomáš Jelínek</i>	16
<b>2.2</b>	<b>Komputační lingvistika v kontextu psychologického výzkumu: Současná témata a vybrané přístupy</b> Computational linguistics in the context of psychological research: Current issues and selected approaches <i>Dalibor Kučera</i>	19
<b>2.3</b>	<b>Problematika dotazníkového sebeposouzení a posouzení druhým v psychologickém výzkumu</b> Self-report and other-report questionnaires use in psychological research <i>Dalibor Kučera</i>	31
<b>3</b>	<b>Výzkum CPACT</b> CPACT research	
<b>3.1</b>	<b>Úvod do výzkumu CPACT</b> Introduction to the CPACT research <i>Dalibor Kučera</i>	44
<b>3.2</b>	<b>Výzkumný soubor a metody</b> Research sample and methods <i>Dalibor Kučera</i>	59

<b>3-3</b>	<b>Charakteristika dat a práce s daty</b> Characteristics of and work with data	
3-3.1	<b>Psychometrické charakteristiky použitých testů</b> Psychometric characteristics of the administered tests <i>Tomáš Urbánek, Dalibor Kučera, Jiří Haviger</i>	88
3-3.2	<b>Lingvistická analýza textů</b> Linguistic analysis of texts <i>Tomáš Jelínek</i>	110
<b>3-4</b>	<b>Zpracování dat a výsledky výzkumu CPACT</b> Data processing and results from the CPACT research	
3-4.1	<b>Statistická explorace vztahů mezi osobnostními charakteristikami a textovými parametry</b> Statistical exploration of relationships between personality traits and text parameters <i>Jiří Haviger</i>	125
3-4.2	<b>Zpracování výsledků dotazníkových variant sebehodnocení a posouzení druhou osobou</b> Processing of results of self-report and other report questionnaire variants <i>Dalibor Kučera, Jiří Haviger</i>	141
3-4.3	<b>Rozsah registrované variability textů</b> Scope of register variability of texts <i>Václav Cvrček, Zuzana Komrsková, David Lukeš</i>	153
3-4.4	<b>Odraz osobnosti v textu – rysy modelu Big Five</b> Reflection of a personality in texts – Big Five traits <i>Jana M. Havigerová, Jiří Haviger, Justýna Franková</i>	173
3-4.5	<b>Odraz osobnosti v textu – projevy deprese</b> Reflection of a personality in texts – depression <i>Jana M. Havigerová, Jiří Haviger, Justýna Franková</i>	201
<b>4</b>	<b>Závěr</b> Conclusions	244
<b>5</b>	<b>Literatura</b> References	247
<b>6</b>	<b>Přílohy</b> Appendices	271

# 1 Úvod

Kniha „Výzkum CPACT: Komputační psycholingvistická analýza českého textu“ (*CPACT Research: Computational Psycholinguistic Analysis of Czech Text*) představuje základní publikační výstup ze stejnojmenného výzkumného projektu, realizovaného v období let 2016–2018 na Jihočeské univerzitě v Českých Budějovicích. Přináší nejen podrobnou dokumentaci výzkumu z pohledu jeho východisek, designu a použitých metod, ale prezentuje také první vědecké výstupy, kterých bylo v rámci projektu dosaženo. Tyto studie jsou samozřejmě jen malou výsečí z oblastí, kterými jsme se ve výzkumu CPACT zabývali. Po vydání publikace proto budou následovat další odborné texty, ve kterých budeme uvedena témata dále rozpracovávat a přinášet další zajímavé výsledky.

Publikace je rozdělena do dvou základních bloků. První blok představuje trojici kapitol souhrnně označených jako „Komputační lingvistika a psychologický výzkum“, která rozpracovává lingvistická a psychologická východiska, ze kterých náš projekt vychází. Druhým blokem je pak skupina kapitol pojmenovaných „Výzkum CPACT“, ve kterých naleznete jak detailní popis výzkumu, tak výše zmíněné studie.

Na přípravě kapitol se podílelo devět autorů, avšak kompletní tým projektu zahrnoval 34 pracovníků – ať už členy projektového týmu, projektové asistenty nebo administrativní a technické pracovníky. Jen díky jejich spolupráci a nasazení bylo možné projekt úspěšně realizovat a dokončit, stejně tak jako vydat tuto knihu. Všem našim kolegům velice děkujeme.

Zvláštní dík pak patří paní Jiřině Hamerníkové, panu Františku Hlavatému, paní Renátě Mikulové a paní Andree Slabé, za jejich odhodlání a pečlivost při realizaci výzkumných sezení, stejně jako paní doktorce Jele Hrnčiarové za zprostředkování kontaktu a spolupráci při sběru dat klinické populace. Děkujeme také účastníkům našeho výzkumu, za jejich ochotu, důvěru a čas, který našemu výzkumu věnovali.

*Dalibor Kučera a Jana M. Havigerová*

## 1.1

# Vymezení základních pojmů

*Dalibor Kučera, Jana M. Havišerová*

Následující část textu je věnovaná především terminologii, resp. vymezení několika pojmů, se kterými pracujeme napříč publikací a u nichž považujeme správnou interpretaci za velmi důležitou. Je třeba zmínit, že zde uvedená definice nemusí vždy odkazovat ke konsenzuálnímu pojetí termínu, ale spíše k vymezení, které využíváme k vytižení daného jevu v rámci našeho výzkumu, resp. v rámci této knihy.

Projekt CPACT se pohybuje na pomezí dvou základních disciplín, psychologie a lingvistiky. Náš výzkum tak označujeme jako **psycho-lingvistický**. Lingvistická část výzkumu je zároveň oborově specializovaná, neboť je v ní využíváno výlučně počítačového zpracování přirozeného jazyka a spadá tak do oboru **komputační lingvistiky** (viz kap. 2.1). Charakter našeho výzkumu je tak ve výsledku popsán jako komputačně–psycholingvistický (viz např. Krahmer, 2010), resp. využívající **komputační psycholingvistické analýzy**. Podrobně se této problematice věnujeme v kapitole 2.2.

**Jazykem** označujeme systém znaků a pravidel, jehož základní funkce je komunikační, dorozumívací a kterým se řídí všichni mluvčí daného jazyka (označovaná dle de Saussurea *langue*; viz Haroche, Henry, & Pecheux, 1971). Konkrétní promluva v daném jazyce pak bývá označována dle de Saussurea jako *parole* (např. Černý, 1996). *Parole* je založena na obecném jazykovém základu (nelze realizovat *parole* bez znalosti *langue*). Jak výstižně uvádí Černý (1996, s. 136): „Pozorujeme-li jazyk přímo v lidské životní praxi a zachycujeme-li a zkoumáme aktuální jazykový materiál, zjistíme brzy, že jazyk neexistuje ani v sebemenším jazykovém společenství v nějaké zcela jednotné, „čisté“ podobě, a dojdeme nakonec k závěru, že každý jednotlivý uživatel používá své vlastní varianty jazyka, která se aspoň v detailech liší od všech individuálních variant ostatních příslušníků jazykového společenství. Taková individuální varianta jazyka, neboli **idiolekt** (z řec. *ídios*

– osobní, soukromý a *legein* – říci, mluvit), se dá charakterizovat jako soubor všech jazykových zvyklostí jednoho člověka v jednom časovém stadiu (v jedné době).“ Každý jedinec má tedy svou jedinečnou slovní zásobu a svůj osobitý způsob, kterým ji skládá k sobě (individuálně specifickou gramatiku a syntax). Idiolekt se vytváří v průběhu osvojování jazyka a celoživotně se přirozeně proměňuje. Metaforicky se dá říci, že idiolekt je v daném čase „zrcadlem“ naší životní historie, odráží se v něm celá osobní historie jedince, všechny vlivy, jimž v minulosti byl a aktuálně je jedinec vystaven, vyvíjí se pod vlivem příslušnosti k nějaké sociální skupině, stupni vzdělání, komunikačnímu partnerovi, komunikační situaci a dalších okolností (viz např. Karlík, Nekula, & Pleskalová, 2002; Müllerová, 1978; Průcha, 2011). Mezi takoveto vlivy můžeme řadit i tzv. registr (též označovaný jako rejstřík, funkční styl či dokonce norma; Uličný, 2014).

Pojem **registr** označuje varietu jazyka, jejíž podoba a užití jsou podmíněny typem komunikační situace a která je ohraničena sociohistorickými procesy, přičemž Slančová a Slančová (2012) mluví o situačně či sociálně „podmíněném konvencionalizovaném jazykovém a parajazykovém chování“. Každou promluvu (*parole*) lze charakterizovat nějakým registrem, například registr veřejného projevu, registr kázání, registr komentování sportovního utkání, registr písemného životopisu atp. (ibid.). Jazyk by v tomto kontextu dokonce mohl být pojímán jako soubor všech existujících registrů. Ve vztahu k naší studii je podstatné, že do každé promluvy (psané či mluvené *parole*) se v rámci různých situačně podmíněných registrů vždy promítá také idiolekt, který i v situačně podmíněném textu zanechává specifický lingvistický otisk (tzv. *footprint*) osobnosti mluvčího/pisatele. Jinými slovy zde předpokládáme, že specifické osobnostní rysy se promítají do psaných a mluvených textů v různých situačních kontextech vymezených lingvistickými rejstříky (registry, funkčními styly) prostřednictvím specifických projevů idiolektu v daném textu a že lidé s podobnými rysy mohou promítat idiolekt do textu podobným způsobem (např. neurotičtí jedinci budou mít tendenci v dopisech a žádostech preferovat tzv. *ich*-formu).

Pro označení specifických znaků textu, které můžeme sledovat a jejich rozsah v textu vyjádřit, využíváme termín **textové parametry**.

Jedná se především o morfosyntaktické, lexikální a sémantické jevy (případně jejich kombinace), které mohou nabývat různých hodnot, čehož je možné využít např. při následném statistickém zpracování (viz kap. 3.3.2). Mezi označení (pojmy), která zde můžeme považovat za víceméně synonymická, je možné zařadit např. textové proměnné, znaky textu, texové markery či textové charakteristiky.

**Scénářem** rozumíme definici a nastavení situace, ve které probíhá specifická forma sociální interakce (komunikace). Popisuje její aktéry, situační kontext (např. prostředí a časový rámec) a předpokládá typické komunikační projevy těchto aktérů.

V rámci psychologické části textu běžně užíváme označení **osobnostní charakteristiky**. Jedná se o pojmenování psychologicky relevantních vlastností člověka projevujících se v jeho chování a prožívání. Toto vymezení vychází víceméně z definice osobnosti, tj. z „charakteristik člověka, které jsou podkladem konzistentních vzorců chování, přičemž těmito charakteristikami se rozumí myšlenky, city a pozorovatelné (navenek vyjádřené) chování“ (Pervin, 1996). Na základě uvedeného lze pak člověka např. popsat či zhodnotit. Jedná se tedy o poněkud širší vymezení pojmu **rys osobnosti**, tj. relativně stále vlastnosti člověka, která je konzistentní, perzistentní a stabilní (viz Blatný, 2010; Hartl & Hartlová, 2000, s. 517), či **charakteru**, tedy eticky zhodnocené osobnosti (viz Hartl & Hartlová, 2000, s. 207). Do rámce osobnostních charakteristik bychom tedy mohli mimo uvedeného zahrnout i např. motivy či charakteristiky spojené s inteligencí a sebepojetím. Pro doplnění uvádíme, že osobnostní charakteristiky, které jsou nějakým způsobem kvantifikované (např. v podobě bipolární škály extraverte – introverte) bývají označovány jako osobnostní dimenze (Blatný, 2010). Uvedené osobnostní charakteristiky pak vytvářejí více či méně koherentní komplex, individualitu jedince (tzv. osobnostní profil).

Při popisu procesu psychologického testování užíváme čtyři základní termíny. Prvním je **psychologický test**. V rámci této knihy hovoříme o standardizované metodě pro vyšetření duševních funkcí, schopností a osobnosti s možností zachytit jejich stav, změny a poruchy (viz Kondáš, 1992). Typem psychologického testu je pak **dotazník**. Zde se jedná o metodu hromadného získávání údajů pomocí předkládání otázek

či tvrzení (viz Hartl & Hartlová, 2000). Dotazníky se zaměřují na popis jedné nebo více osobnostních charakteristik, které sledují pomocí příslušných **škál** (tzv. škál dotazníkových, testových či obecně psychologických). Je zřejmé, že v ideálním případě bude test obsahující dotazníkovou škálu „extraverze“ měřit osobnostní charakteristiku extraverzi (resp. tento psychologický konstrukt). V reálných podmínkách se však dotazníkové škály cílovým osobnostním charakteristikám pouze více či méně úspěšně přibližují. Každá škála sestává z určitého množství **položek**, u dotazníku z konkrétních otázek (např. „Jste veselý?“) nebo tvrzení (např. „Považuji se za někoho, kdo je vynalézavý.“), na které respondent odpovídá. Z hlediska kategorizování odpovědí na položku je třeba upozornit na jednu terminologickou víceznačnost. Pojem „škála“ je totiž využíván i pro označení konkrétního uspořádání odpovědí, tzv. kategorií odpovědí. Kupříkladu Likertova škála tak označuje míru souhlasu s tvrzením pomocí odpovědí jako např. „nesouhlasím – spíše nesouhlasím – neutrální – spíše souhlasím – souhlasím“. Obsah termínu „škála“ je proto závislý na kontextu, ve kterém je využíván.

Rádi bychom na závěr doplnili, že z důvodu přehlednosti a vyšší čitelnosti textu v celé knize preferujeme užívání generického maskulina (např. termín „účastník výzkumu“, není-li uvedeno jinak, označuje souhrnně ženy i muže).



2

Komputační  
lingvistika  
a psychologický  
výzkum

## 2.1

# Současná východiska počítačnické lingvistiky a její aplikace

*Tomáš Jelínek*

Komputační lingvistika je moderní vědecký obor, který využívá počítače k analýze přirozeného jazyka a vytváří automatické postupy, kterými s přirozeným jazykem manipuluje. Současná komputační lingvistika pracuje jednak na úkolech s praktickým využitím, jako je automatický překlad, automatický převod mluvené řeči do písemné podoby aj., jednak na aplikacích, jež využijí primárně vědci, například na automatické analýze přirozeného jazyka: morfologické, syntaktické, lexikální a podobně, přičemž rozvoj v jedné oblasti leckdy přináší užitek i v oblastech jiných.

Historii počítačové lingvistiky lze sledovat od druhé světové války, kdy Spojenci využili jeden z předchůdců dnešních počítačů, Colossus (Wells, 2009) k rozluštění zašifrovaných německých zpráv. Během studené války pak obě strany usilovaly o vytvoření systému automatického překladu (z ruštiny do angličtiny a naopak). Tato snaha byla sice neúspěšná (na první skutečně fungující systémy automatického překladu se čekalo ještě téměř padesát let), ale položila základy současné počítačové lingvistiky: exaktní metody, algoritmizaci dílčích úloh, vytváření počítačových slovníků, formalismus. Od šedesátých let minulého století se také rozvíjí korpusová lingvistika, což je obor příbuzný lingvistice počítačové, který pro zkoumání přirozeného jazyka vychází z rozsáhlých textových souborů v počítačové podobě. Jedním z prvních morfologicky označovaných korpusů v elektronické podobě byl Brown Corpus (Francis & Kučera, 1964). S rozvojem kapacit a rychlosti počítačů, zvláště pak od začátku jednadvacátého století, dochází k velkému rozkvětu počítačnické lingvistiky a mnoha jejích podoborů jako je výše jmenovaný automatický překlad, záznam mluvené řeči a její generování z textu, oprava gramatiky v psaném textu, automa-

tické vytváření textů z počítačových databází (např. vytváření jednoduchých článků o sportovních událostech), shrnutí (sumarizace) textu a podobně, zároveň se výrazně zlepšují i lingvistické aplikace jako morfosyntaktické značkování, automatické rozlišování významu slov aj. Počítačová lingvistika se prosazuje i jako pomocný nástroj pro jiné humanitní vědy, nabízí například nástroje pro automatické zjišťování informací z rozsáhlých textů, a spoluvytváří širší obor *digital humanities* (počítačové humanitní vědy, viz např. Piotrowski, 2016). V nedávné době bylo dosaženo výrazného zlepšení některých aplikací díky použití tzv. sítí umožňujících hluboké učení (*deep neural networks*, viz např. Manning, 2015).

Komputační lingvistika při automatickém zpracování přirozeného jazyka často volí mezi dvěma postupy nebo je kombinuje: zpracování na základě pravidel vytvořených počítačovými lingvisty, tedy dílčích algoritmů pro řešení jednotlivých jevů, a strojového učení, kdy se počítačový program na předzpracovaných „trénovacích“ datech „naučí“, jak sledované jevy zpracovávat, a tato zjištění potom aplikuje na data nová, nezpracovaná. Tyto metody se také nezřídka kombinují. Například automatický překlad se komputační lingvisté nejprve snažili řešit vývojem pravidel v kombinaci se slovníkem, kdy jsou určitá slova a slovní kombinace překládány předem definovaným způsobem. Nepodařilo se tak ale dosáhnout spolehlivého překladu především proto, že nebylo možné takto postihnout všechnu potenciální jazykovou variabilitu. Současné systémy automatického překladu jsou ponejvíce založeny na strojovém učení, využívají trénovací data složená z miliónů vět paralelně ve výchozím a cílovém jazyce, k tomu případně ještě rozsáhlejší texty z obou jazyků samostatně. Na těchto datech se počítačové programy (sítě umožňující hluboké učení) natrénují a výsledkem je celkem spolehlivý překlad, který lze využívat dnes (např. Bojar & al., 2016).

Současná počítačová lingvistika zahrnuje mnoho dílčích oborů, z nichž mnohé nabízejí aplikace využívané v běžném životě, tedy nejen pro výzkumné účely. Tyto obory lze zhruba dělit podle několika kritérií, např. práce s psaným jazykem a práce s mluveným jazykem; porozumění textu (analýza) a vytváření textu (syntéza); práce se standardním textem a s textem nestandardním (textem obsahujícím chyby, českým

textem bez diakritiky, textem vytvořeným nerodilým mluvčím aj.). Nejvíce různých úloh se řeší v oblasti analýzy psaného textu, mezi mnoha dalšími je to třeba identifikace významu jednotlivých slov, automatické odpovědi na zadané otázky, automatické shrnutí (sumarizace) textu a vyhledávání informací v rozsáhlých textech, zjišťování postoje pisatele k popisovanému předmětu nebo nástroje pro vyhledání a opravy chyb v textu (korektor gramatiky, doplňování diakritiky apod.).

Některé aplikace kombinují více dílčích úloh do jedné komplexní úlohy, například aplikace pro automatický překlad mluvené řeči z jednoho jazyka do mluvené řeči jiného jazyka využívají obvykle nejprve převod mluvené řeči do psané podoby, dále automatický překlad textu z jednoho jazyka do druhého, nejprve tedy analýzu původního vstupu, pak syntézu cílového jazyka (v textové podobě), nakonec převod psaného jazyka do mluvené řeči.

Nástroje počítačové lingvistiky využívané jinými obory (např. historie, sociologie, psychologie) pomáhají získávat z textů informace, které by jinak nebyly dostupné, a to buď kvůli velkému rozsahu textů, nebo kvůli detailům v textu, které je obtížné určovat „ručně“. Může jít o zjištění změn v užití vybraných slov a slovních spojení v čase (historie) nebo třeba hledání souvislostí mezi užitím určitých jazykových prostředků a psychickým stavem respondenta (psychologie). Tomuto hledání se věnuje tato publikace.

## 2.2

# Komputační lingvistika v kontextu psychologického výzkumu: Současná témata a vybrané přístupy

*Dalibor Kučera*

Zpracováním a analýzou lidské komunikace a hledáním souvislostí, které tuto komunikaci definují a ovlivňují, se dnes zabývá nepřehledné množství oborů. Díky dynamickému rozvoji informačních sítí, nárůstu výpočetní kapacity počítačových systémů i zjevnému komerčnímu potenciálu, který s vyhodnocováním komunikace souvisí, není už tato oblast zdaleka jen doménou vědecké komunity, ale stala se klíčovou např. i pro velké komerční subjekty či pracovníky bezpečnostních složek. Sociální média, například Facebook, Twitter, messengery či různé blogy neustále vstřebávají ohromné množství zpráv a informací ve snadno přístupné datové podobě, kterou lze dále využít. Pomocí analýzy akustické řeči a jejímu automatizovanému převodu na psaný text se pak i tento typ materiálů stává snadno komputačně (výpočetně) zpracovatelným. Uplatnění takto získaných výstupů je velké – od sledování spotřebního chování, politických názorů, preferencí, změn ve společnosti, až po personalistické či psychologické aplikace. Pro naši studii jsou zásadní právě aplikace v oblasti psychologie, zejména pak psychodiagnostiky.

### 2.1.1 Úvod do problematiky

---

Dorozumívání pomocí řeči a slov je jedním z pilířů lidské sociální interakce. První propojování vokalizace s interakcí „tváří v tvář“ se objevuje přibližně před dvěma miliony let, tedy ve výrazně dřívější době, než došlo ke vzniku lidského druhu *homo sapiens* (Arib, 2005). Komunikace, která o něco více odpovídá dnešnímu pojetí jazyka, tzv. cvakavé a mlaskavé jazyky, se pak začíná rozvíjet cca před 50–100 tisíci

let. Je zřejmé, že tento způsob sdělení byl v mnoha směrech poměrně limitovaný. Pracoval sice s určitou znakovou zásobou, tedy s popisem předmětů a jevů charakteristickým projevem, např. právě vokalizací, nicméně gramatická stránka promluvy byla velmi primitivní (Workman, 2014). V návaznosti na rozvoj složitějších sociálních interakcí byl také jazyk postupem času obohacován o další, složitější parametry a začala se tvořit jeho ustálená struktura (systém jazyka, gramatika), která už zahrnovala pravidla pro skládání lexikálních jednotek (syntax) a jejich morfologii (ibid.).

Složitě procesy vývoje jazyka se začaly projevovat ve dvou zřejmých benefitech: Prvním přínosem byla možnost úspěšně a srozumitelně předávat informační obsah, který chce komunikující osoba sdělit recipientovi. Vzhledem k tomu, že slovní zásoba průměrného jedince (v západní společnosti) čítá cca 45000 slov (Nagy & Anderson, 1984), dokáže poměrně dobře postihnout většinu předmětů, jevů atd., se kterými jedinec běžně přichází do kontaktu a o nichž chce recipientovi referovat. Druhou výhodou je variabilita a flexibilita komunikačního nástroje. Slovní zásoba většiny současných jazyků totiž obsahuje velké množství lexikálních jednotek, které jsou vzájemně (více či méně) zaměnitelné, a je tak na komunikujícím, které označení v komunikaci zvolí. A nejen to; i gramatická podoba jednoho konkrétního sdělení může být zcela odlišná. Pro názornost uveďme příklad v českém jazyce: Komunikující chce recipientovi sdělit informaci, že má hlad. Může tak učinit pomocí jednoduché věty *Mám hlad.*, nebo ji rozšiřovat a měnit – *Já mám hlad jako vlk!*, *Máte taky takový hlad?*, *Co si takhle dát jídlo?*, *Hodilo by se něco k snědku!* apod. V tomto případě se mění celá gramatická struktura výpovědi; komunikující používá jiné osoby, čísla a času, ale také výrazně upravuje pořadí a zastoupení větných členů, přestože věcný obsah výpovědi zůstává víceméně stejný. Díky variabilitě, kterou je možné využít při formování každého sdělení, může být výsledná podoba (i zcela běžné) komunikace velmi originální a pro komunikujícího vysoce specifická.

Myšlenka, že volba formy sdělení (tedy nikoliv jen „co chci říct“, ale „jak to chci říct“, tj. styl promluvy) není náhodná a že je do určité míry vázaná na osobnost či např. motivy komunikujícího, se v odborné

literatuře objevuje již řadu desetiletí. Je ostatně běžné, že na základě určitých prvků řeči implicitně usuzujeme například na charakter, intelektuální kapacitu či sociální status. Pionýrskou mezi studiemi je práce Dorothee McCarthyové (1929), která na souboru dětí (2–15 let) prokazuje, že rozdíly v užívání jazykových prostředků vznikají nejen v závislosti na situačních podmínkách, ale též v závislosti na osobnostních rysech mluvčích. Fillmore Sanford (1942) myšlenku dále rozvíjí a popisuje verbální projev jako specifický osobnostní marker. Někteří autoři pak verbální chování považují přímo za osobnostní rys (např. Cheng, 2011) a odkazují na jeho propojení s nonverbálními formami projevu i na jeho úlohu v intra- a interpersonálních procesech (viz např. Furnham, 1990; Scherer & Giles, 1979; Weintraub, 1989).

Studium psychologických kontextů verbální komunikace vychází především z disciplinární a metodologické báze, která definuje jeho zaměření i konkrétní techniky výzkumu. Zásadním oborem je v tomto směru **psycholingvistika** (*psycholinguistics*). Psycholingvistika vznikla v polovině 20. století spojením jazykovědy a psychologie. Je to disciplína, zabývající se např. produkcí a recepcí řeči či vztahem ke kognitivním procesům (Nebeská, 1992, s. 8). Tento obor pracuje s pestroutou paletou nástrojů a metod, nicméně z psychologického či sociologického pohledu je jednou z klíčových metod obsahová analýza (*content analysis*). Obsahová analýza se soustřeďuje především na konkrétní, explicitní komunikační obsah, resp. na produkt komunikace (text), který pomocí rozličných kritérií posuzuje (Berelson, 1952, s. 18). Předmětem zájmu obsahové analýzy je např. kontext, ve kterém byl jazykový produkt vytvořen, motivace komunikátora, jeho cíle a efekt, který má jeho sdělení na recipienta textu (ibid). Obsahová analýza může být prováděna kvalitativně či kvantitativně (např. komputačně), konceptuálně (*conceptual*; sledující existenci/frekvenci určité jazykové jednotky) nebo relačně (*relational*; popisující vztahy mezi výskyty jednotky, zejm. mentální modely) (Carley, 1993, s. 78–80), případně se zaměřením na obsah (*content-focused*; sémantiku, tj. „co“ je komunikováno) nebo na styl (*style-focused*; formu/styl, tj. „jak“ je komunikováno) (Mehl, 2006b, s. 144–145).

Při pohledu na současné trendy v oblasti psycholingvistických metod nemůžeme přehlédnout evidentní nárůst popularity počítačích (kvantitativních) přístupů; tedy výpočetního zpracování velkého množství textu (Altmann, 2006). Zde se jako významný obor etablovala především **komputační psycholingvistika** (*computational psycholinguistics*), obor zaměřený na využití výpočetních přístupů při studiu kognitivních mechanismů a reprezentací souvisejících se zpracováním jazyka. Základními tématy jsou osvojování, porozumění a produkce jazyka, zejména na úrovni kognitivních procesů a z hlediska fungování mozku (Crocker, 2012).

V následujících kapitolách bude hlavní pozornost věnována studii vztahu psychologicky relevantních charakteristik mluvčích (jejich osobnosti, motivaci, prožívání, aktuálnímu stavu apod.) k podobě textu, který tyto osoby produkují, konkrétně k výsledku kvantitativní, komputační lingvistické analýzy jejich textu. Tento postup zde označujeme jako komputačně-psycholingvistický. Je zřejmé, že takovéto vymezení je pouze výsečí z celkového záběru psycholingvistického bádání a zároveň, že se poněkud vzdaluje běžnému vymezení cílů a témat komputační psycholingvistiky (viz výše). V rámci naší studie však považujeme zařazení do rámce komputační psycholingvistiky za správné a logické; což navazuje nejen na některé zahraniční texty, které popsanou specializaci vnímají jako komputačně-psycholingvistickou (např. Krahmer, 2010), ale především s ohledem na tři základní komponenty našeho výzkumu – psychologické proměnné, textové (jazykové) parametry a proces výpočetního zpracování, tedy psychologicko-lingvisticky-komputační metodologický komplex. Pro doplnění uvedme, že v zahraniční literatuře v tomto směru chybí konsenzus a hledání psychologických markerů v textu (tedy ona „komputačně-lingvistická psychodiagnostika“) je často řazené do oboru sociální psychologie (viz např. Holtgraves, 2014), nebo do interdisciplinárního mixu oborů lingvistiky, sociolingvistiky, neurovědy, psycholingvistiky, komputační lingvistiky atd. (Pennebaker, 2011, s. 8).



### 2.1.2

## Možnosti zpracování textu pro potřeby psychologického výzkumu

---

Současné přístupy kvantitativní (komputační) jazykové analýzy, využívané pro psychologické a obecně společenskovední účely, je možné zjednodušeně rozdělit do dvou základních skupin – tzv. **uzavřené přístupy** (*closed approaches*, nebo také *closed vocabulary analysis*; Schwarz et al., 2013) a tzv. **otevřené přístupy** (*open approaches*, resp. *open vocabulary analysis*; *ibid.*). Uzavřené přístupy, v současnosti převažující, jsou reprezentované počítáním frekvence výskytu předdefinovaných slov, která jsou součástí určitého slovníku, databáze (např. databáze morfologických kategorií, slov spojených s emocemi či sociálně relevantních slov; Park et al., 2015). Druhá skupina, tzv. otevřených přístupů, je naopak zacílená na nekategorizovanou extrakci jazykových jevů (slov, sousloví, interpunkce, emotikonů, témat apod.), které nejsou a priori definované a vychází pouze z analyzovaného textu. Principiálně se zde jedná o obdobu sémantické LSA (*latent semantic analysis*), která vyhledává výskyty klíčových slov v určitém korpusu (Landauer, Foltz, & Laham, 1998). Oba přístupy mají řadu výhod i úskalí; otevřené přístupy jsou např. ideální volbou při práci s velkými daty a různými jazyky, zatímco uzavřené přístupy umožňují lépe definovat a diskutovat základní teorie (viz Grimmer & Stewart, 2013; O'Connor, Bamman, & Smith, 2011; Yarkoni, 2010). Stále populárnější je také propojování zmíněných přístupů (např. Schwarz et al., 2013), případně využívání umělé inteligence při extrakci informací z textu. Ostatně právě zde zaznamenává komputační lingvistika v posledních letech poměrně velké změny a aplikace neuronových sítí a *deep learning-based methods* (tzv. metody hlubokého učení) vykazují při analýzách nebyvalou úspěšnost (viz např. Majumder, Poria, Gelbukh, & Cambria, 2017).

Pro naši studii jsme jako klíčový nástroj analýzy zvolili uzavřený přístup. Důvodem byla nejen vysoká transparentnost procesu analýzy textu a přehlednost výstupů, ale také žádoucí kompatibilita s podobně zaměřenými zahraničními výzkumy. Psychologické využívání uzavřených kvantitativních procedur v posledních dvaceti letech vý-

razně facilitovala dostupnost textově-analytického softwaru **LIWC** (*Linguistic Inquiry and Word Count*; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007), představeného v polovině devadesátých let J. W. Pennebakerem (ibid.). LIWC pracuje na principu vnitřního (uzavřeného) slovníku, s jehož obsahem je porovnáván analyzovaný text. Jednotlivé výskyty znaků v textu jsou skórovány a fixovány v absolutním/relativním počtu a následně řazeny do sémantických a formálních (tj. nesémantických, stylových) kategorií. Jedná se o čtyři základní kategorie: souhrnné jazykové proměnné (*summary language variables*; např. počet slov), lingvistické dimenze (*linguistic dimensions*; např. slovní druhy), jiná gramatika (*other grammar*; např. kvantifikátory) a psychologické procesy (*psychological processes*; slovník zahrnující slova spojená např. s afektivními procesy či orientací na čas). Software byl od doby svého vzniku průběžně optimalizován a v poslední verzi LIWC 2015 je možné analyzovat 83 proměnných, z čehož 29 proměnných pokrývá gramatika (viz Pennebaker, Boyd, Jordan, & Blackburn, 2015). Mimo LIWC je zde samozřejmě dostupná i řada dalších aplikací, které je možné k sémantické a/nebo formální analýze využít – např. *General Inquirer* či *Thematic Content Analysis*, tedy systémy hodnotící zejména motivaci komunikátora (Stone et al., 1966; Smith, 1992), *Gottschalk–Gleser Method*, zaměřenou na textový obraz klinických diagnóz (jako např. anxiety či hostility; Gottschalk et al., 1969) či *Regressive Imagery Dictionary* (Martindale, 1987), věnující se především kognitivním procesům. Jejich stručný přehled uvádíme v tabulce 2.1.2.

Tab. 2.1.2 – Přehled významných aplikací zaměřených na analýzu jazyka v psychologickém výzkumu (převzato dle Mehl, 2006b, s. 147).

Orig. název aplikace	Autor/autoři	Sledované lingvistické proměnné (textové parametry)	Kód.*	Zaměření analýzy
Thematic Content Analysis	Smith (1992)	Motiv moci, výkonu a afiliace, explanatorní styl, integrativní komplexita	LP	obsah (sémantika) / formální (styl)

General Inquirer	Stone et al. (1966)	Harvardský III. psychosociologický slovník, Stanfordský politický slovník, Slovník <i>Need-Achievement</i>	C	obsah (sémantika) / formální (styl)
Gottschalk–Glesser Method	Gottschalk et al. (1969)	Klinické fenomény (např. anxieta, hostilita, sociální odcizení, deprese, kognitivní deficit)	LP / C	obsah (sémantika)
Regressive Imagery Dictionary	Martindale (1987)	29 primárních kategorií (tzv. <i>primary process cognition</i> ), 7 sekundárních kategorií (tzv. <i>secondary process cognition</i> ), 7 kategorií emocí	C	obsah (sémantika)
Analysis of Verbal Behavior	Weintraub (1981)	15 dimenzí vč. zájmen ( <i>I, we, me</i> ), negací (např. <i>not, no, never</i> ), modifikátorů (např. <i>kind of</i> ), exprese pocitů (např. <i>love</i> ) a adverbálních intenzifikátorů ( <i>really, so</i> )	LP	formální (styl)
TAS/C	Mergenthaler (1996)	Emoční naladění, abstrakce, reference	C	formální (styl)
DICTION	Hart (1984)	5 základních proměnných (jisota, optimismus, aktivita, realismus, podobnost) s celkem 35 lingvistickými subdimenzemi	C	formální (styl)
LIWC	Pennebaker et al. (2001)	82 proměnných; standardní lingvistické dimenze (např. zájmena, člen), psychologické procesy (např. emoční slova, kauzalita), relativita (minulý čas, inkluzivní slova), osobní zájmy (např. škola, náboženství, sexualita)	C	obsah (sémantika) / formální (styl)

LSA	Landauer et al. (1998)	n/a (bez proměnných); 2 strategie zaměřené na slova s nízkou (obsah, sémantika) nebo vysokou (formální, styl) frekvencí	C	obsah (sémantika) / formální (styl)
-----	------------------------	---	---	-------------------------------------

\* Kód. – způsob kódování; LP – lidsší posuzovatelé; C – počítač

### 2.1.3

#### Detekce psychologicky relevantních charakteristik pomocí počítační lingvistické analýzy

Využívání kvantitativní textové analýzy pro psychologicko-dia-  
gnostické účely není ve společenskovědním výzkumu novým tématem. Výraznější posun tato oblast zaznamenala v sedmdesátých letech dvacátého století, což souviselo především s nárůstem kapacity systémů, které umožnily účinně provádět automatizované zpracování textu, tedy s rozvojem výpočetních (tj. počítačích) technologií a potřebného softwaru. Díky ohromnému množství dat, která bylo možné do výpočtů zahrnout, ale také díky modernímu statistickému zpracování, se postupně začala objevovat cesta k psychologické interpretaci výsledků analýz i k tomu, aby se mezi daty identifikovaly skutečně podstatné a relevantní informace. Stále větší pozornost zároveň získávala analýza formální (nesémantické) stránky textu a řada autorů tyto do té doby relativně opomíjené parametry dokonce prohlásila za klíčový psychologicko-textový faktor (např. Pennebaker, 2011). Formálními parametry jsou zde označována především ta specifika textu, která jsou založena na stanovení gramatické kategorie slov a jejich kvantity.

Je pozoruhodné, že podle těchto kritérií se jako nejperspektivnější ukázala slova, která byla v předchozích výzkumech bez nadsázky zanedbávána (dostávalo se jim např. označení *junk words* – odpadová slova), konkrétně neplnovýznamové slovní druhy, jakými jsou zájmena, předložky, členy, spojky a slova víceméně pouze doplňující větnou syntax (např. spony u sloves) (Chung & Pennebaker, 2007, s. 347). Tyto druhy jsou jen minimálně spojeny s věcným obsahem promluvy a lidé se jim učí nejpозději (z hlediska ontogeneze). Dle Pennebaker (2011)

však mají tato slova implicitní sociální a psychologický význam. A svou teorii dokládá i na poznatcích z oblasti neurofyziologie. Uvádí např., že je přinejmenším pozoruhodné, že ve výzkumech pacientů s postižením tzv. Brocova centra (souvisejícího s verbální expresí) byly zjištěny výrazné změny v produkci jmen a sloves, ale nikoliv funkčních slov. Při postižení tzv. Wernickeho centra (souvisejícího s verbální percepcí) pak výrazně klesá produkce významových druhů slov a výrazně narůstá užívání slov funkčních. Je tedy patrné, že i na této neurofyziologické úrovni mají funkční slova poměrně výsadní roli (ibid.).

Společně s dalšími autory Pennebaker publikoval desítky studií, poukazujících na vztah užívání jazyka s osobnostními charakteristikami, nejčastěji klasifikovanými v rámci modelu Big Five (Costa & McCrae, 1992). Tyto výstupy se týkají např. extraverte, která zřejmě souvisí s častějším užíváním slov pozitivně emočně zabarvených (např. *skvělý, šťastný, úžasný*; Pennebaker & King, 1999), což potvrzují i výzkumy Schwartze et al. (2013) či Yarkoniho (2010), s využíváním delších slov či akronym (Holtgraves, 2011) a s menším výskytem slov vyjadřujících rozlišování (např. *na rozdíl, kromě*; Pennebaker & King, 1999), neuroticismem, kdy lidé s vyšším skórem užívají více 1. osoby singuláru (např. *já, můj, mě*) a méně pozitivně emočně zabarvených slov (ibid), otevřenosti ke zkušenosti, charakterizované vyšším výskytem odkazů na sociální procesy a vyšší frekvencí citací (Sumner, Byers, & Shearing, 2011) či přívětivosti, kdy se v textu objevuje větší počet členů a větší množství slov v 1. osobě singuláru (viz také neuroticismus). Kromě dimenzí modelu Big Five byla popsána i řada souvztažností s dalšími osobnostními charakteristikami, např. sociální desirabilitou, která zřejmě souvisí s vyhýbáním se přivlastňování (např. nižším počtem výrazů *můj, váš*) (Knapp et al., 1974), úzkostností, kdy je u úzkostných jedinců výskyt slov v 1. osobě vyšší než slov v osobě 2. a 3. (Pennebaker & Stone, 2003) či interpersonální orientací, která je charakterizována nižším výskytem zájmen v 1. osobě a vyšším výskytem zájmen označujících relaci (např. *kdokoliv, jež*) u osob, které jsou orientované na sociální interakci (tzv. *interaction involvement*; Cegala, 1989).

Mimo osobnostních charakteristik se někdy výzkumy dotýkají i situačního kontextu, resp. konativních procesů, např. lhaní, které je obvykle doprovázeno nižším výskytem slov v 1. osobě singuláru a slov vylučujícího významu (*kromě, bez, nikdy*; Newman et al., 2003).

Rozdílnosti ve způsobu, jakým je tvořen text, byly zjištěny také s ohledem na pohlaví komunikátora; takže zatímco ženy užívají více zájmen 1. osoby singuláru a odkazů na sociální interakci, muži užívají více členů a jejich text obsahuje více formálních a ryze informativních slov (Pennebaker & Stone, 2003; Argamon et al., 2003; Mukherjee & Liu, 2010; Rao et al., 2010). Další zajímavé vztahy byly nalezeny v souvislosti s věkem komunikátora, kdy starší lidé preferují zájmena v plurálu, budoucí čas sloves, pozitivně emočně zabarvená slova, menší množství referencí na svou osobu a nižší výskyt negace (Pennebaker & Stone, 2003; Chung & Pennebaker, 2007; Schler et al., 2006). Další doklady o propojení jazyka s osobností můžeme nalézt i v pracích řady jiných autorů (např. Cohen, Minor, Baillie, & Dahir, 2008; Fast & Funder, 2008; Hirsh & Peterson, 2009; Lee, Kim, Seo, & Chung, 2007; Mehl et al., 2006).

Důležitou otázkou ve vztahu k uvedeným poznatkům je míra, s jakou je verbální chování (komunikace) jedince diskriminativní, tj. variuje-li, nebo naopak konzistentní v rámci různých situací (např. Shoda & Mischel, 1994). Pokud bychom měli citovat J. W. Pennebaker, je konzistence textových parametrů mezi různými typy promluv (například tématy a časem) pozoruhodně vysoká a srovnatelná s tradičními psychologickými dotazníky (Pennebaker & King, 1999, s. 1309). Řada studií však uvádí i jiný pohled a jejich autoři zmiňují, že diferenciací textů a komunikačních registrů (viz např. Biber, 1993) je velmi významná z hlediska interpretace výstupů analýz, resp. z hlediska jejich zobecnitelnosti. Vliv komunikačního kontextu na podobu a variabilitu projevů jedné osoby (tzv. *language style*) byl opakovaně popsán v rovině užívání jazyka jako takového (např. Chen & Bond, 2010), ale i při konkrétních manipulacích se slovy (např. Newman, Groom, Handelman, & Pennebaker, 2008; Ireland & Mehl, 2014). Je patrné, že přestože se zde jedná o skutečně zásadní téma, nepanuje v otázce variability textu, v rámci produkce konkrétního jedince–mluvčího, příliš velká shoda.

Pro současný výzkum by tedy mělo být prioritní oblast této intersituace – ní variability sledovat a poskytnout k ní další výzkumnou evidenci.

#### **2.1.4 Čeština: Jazyková specifika a východiska výzkumu**

---

Výše popsaná analýza verbální komunikace je do značné míry vázána na charakter a specifika cílového jazyka. Angličtina je zde, alespoň co do počtu publikovaných výzkumů, zcela dominující. Přesto je možné nalézt příbuzné studie také v čínštině, arabštině, španělštině, holandštině, francouzštině, němčině, italštině a turečtině a k dispozici je také několik studií pro ruštinu a srbštinu (např. Bjekić et al., 2012; Sboev et al.; 2016, Sikos et al., 2014; Litvinova, Seredin, Litvinova, & Lyell, 2017). Srovnatelnost slovanských výzkumů je však poměrně problematická jak z hlediska jejich zacílení, tak z hlediska použité metodologie či formátu výstupů (viz Panicheva, Ledovaya, & Bogolyubova, 2016; Kartelj, Filipović, & Milutinović 2012). V rámci sledovaného tématu tak nejsou momentálně k dispozici dostatečné podklady, které by umožnily ověřit či korigovat výsledky anglofonních studií ve směru k slovanským jazykům. Je přitom zřejmé, že dominantní zaměření psycholingvistických výzkumů na anglický jazyk, byť jde o hlavní světový jazyk a jazyk do jisté míry univerzální, není vhodným řešením. Morfoloická, lexikální a stylistická struktura jazyků z jiných rodin může být totiž poměrně odlišná a/nebo může zprostředkovat jevy, které v angličtině nejsou přímo identifikovatelné.

Tato situace se týká i českého jazyka. Čeština, jako zástupce západoslovanské jazykové skupiny, se od angličtiny liší v mnoha ohledech, například z pohledu flexe (čeština je vysoce flektivní jazyk, ve kterém např. afixy kumulují gramatické funkce), lexikologie (např. česká deminutiva) či syntaxe (pořadí slov není tak fixované jako v angličtině) (Hornová, 2003). Čeština je sice jazyk s relativně malým počtem mluvčích, řadí se až na 83. místo v rámci dalších světových jazyků, avšak kupříkladu v prostředí internetu se v roce 2007 zařadila na překvapivé 15. místo (Parkvall, 2007).

Prvním českým výzkumem zaměřeným na komputační analýzu formálních aspektů verbální komunikace, z pohledu vztahu textových parametrů s osobností komunikujícího, je výzkum CPACT (*Computational Psycholinguistic Analysis of Czech Text*; Komputační psycholinguvistická analýza českého textu), kterému je tato publikace věnována. Pomineme-li vlastní předvýzkumy a pilotáže (Kučera & Havigerová, 2015a) a práci Linkova a Šmerka (2009), zaměřenou na hodnocení klamání v on-line komunikaci, nejsme si vědomi podobně zaměřeného výzkumu prováděného na českém jazyce. Svým rozsahem projekt zároveň přesahuje odborný záběr výzkumů realizovaných na dalších slovanských jazycích. Jeho přínos je tedy zásadní nejen z perspektivy domácího výzkumu, ale i v širším, světovém kontextu.



## 2.3

# Problematika dotazníkového sebesposouzení a posouzení druhým v psychologickém výzkumu

*Dalibor Kučera*

Kapitola je zaměřena na problematiku současného psychologického testování, konkrétně na využívání psychologických dotazníků a na limity jejich aplikace. Pozornost je věnována nejen tematice a mechanismům zkreslování odpovědi respondenty, ale také specifikům administrace, při které se dotazování v rozdílných rolích vyjadřují směrem k jiné, cílové osobě. Mimo oblasti klasického dotazování ve formě sebesposouzení (*self-report*; osoba posuzuje sama sebe) se tedy zaměříme i na variantu posouzení druhým (*other-report*), ať už v roli obeznámené či blízké osoby, nebo v roli nezaujatého pozorovatele (*stranger-report*) (Wood, Harms, & Vazire, 2010). Text se věnuje rovněž asymetrii (diskrepanci), ke které může v této souvislosti mezi různými posuzovateli docházet. Je důležité zmínit, že daná témata jsou v kontextu výzkumu CPACT velmi důležitá a že byla mj. klíčová při koncepci jeho psychodiagnostické části. Dílčí zhodnocení uvedené problematiky, již zcela v rovině našeho výzkumu, pak bude představeno v kapitole 3.4.2 (Výsledky zpracování variant sebesposouzení a posouzení druhou osobou).

### 2.3.1 Úvod do problematiky

---

Využívání dat z dotazníkového sebesposouzení je dnes běžně využívaným psychodiagnostickým postupem. Dotazníky jsou „psychodiagnostické metody založené na subjektivní výpovědi vyšetřované osoby o jejich vlastnostech, citech, postojích, názorech, zájmech a způsobu reagování v nejrůznějších situacích (atd.). Jejich konstrukce vychází z principu introspekce, úkolem zkoumané osoby je zatrhnout nebo ji-

ným způsobem označit výpovědi, které podle jejího názoru nejlépe vystihují zkoumaný znak“ (Svoboda, 1999, s. 221). Přestože historie jejich využívání sahá už do počátku 20. století, resp. k době první administrace testu *Woodworth's Personal Data Sheet* (Shepherd, 1919), jejich popularita nijak výrazně neklesá. Oproti například rozhovoru nabízí dotazníková metoda celou řadu výhod, mezi něž patří praktická forma administrace, která může být hromadná, velmi dobře kontrolovaná (např. co do kontroly prostředí), přinášející navíc možnost snadného vyhodnocení, zejména v případě strojového zpracování. Nevýhodou je pak zřejmá redukce komplexity získávané informace, limity spojené s mechanismem introspekce i související omezené možnosti korekce dat z hlediska vědomého či nevědomého zkreslení skutečnosti respondentem (tzv. *response bias*, Kaplan & Saccuzzo, 2017).

### 2.3.2

#### **Limity psychologických dotazníků a asymetrie mezi různými variantami posouzení**

---

Z pohledu tradičního psychologického dotazování jsou zásadními podklady pro zhodnocení osobnostních charakteristik člověka informace, které nám sám poskytuje. Tato teze vychází z předpokladu, že nejlepším expertem na sebe je člověk sám a že „poznání druhého člověka nikdy není celistvé, máme-li k dispozici pouze útržky, které s námi sdílí. Ať už jsou to myšlenky, názory nebo pocity, vždy je možné nahlédnout pouze na střípky jeho duševního světa“ (Allport, 1937, s. 47). Ostatně např. 70 % studií publikovaných do roku 2003 v *Journal of Research in Personality* bylo založeno právě a výlučně na sebesposouzení (Varize, 2006). Navzdory tomuto pohledu se však objevuje stále rozsáhlejší výzkumná evidence k tvrzení, že kvalita sebesposouzení (relevance náhledu na sebe, sebereflexe) je v mnoha směrech ovlivňována spíše sebezpojetím jedince (často nepřesným) a že např. ostatní lidé nás mohou vnímat, popisovat a posuzovat přesněji a uceleněji, než my vnímáme sami sebe (Vazire & Mehl, 2008).

Pomineme-li zkreslení výpovědní hodnoty (sebe)posouzení vinou zřejmé formulační nekvality otázek a zadání, nedbalosti či záměrného

narušení odpovědí respondenty, zjevnou roli hrají např. **vlivy sociální desirability** (také tzv. *self-serving bias*, *self-enhancement* apod.), **problematická interpretace konstruktů a termínů respondentem** a **specifika samotné introspekce či extraspekce** (Furnham, 1986).

Následující text tedy bude zaměřen na popis a interpretaci těchto aspektů, stejně jako na dostupná metodologická řešení, která umožňují uvedená omezení do určité míry redukovat.

### 2.3.2.1 Vliv sociální desirability

---

Sociální desirabilitu bychom mohli definovat jako „snahu odpovídat líbivě bez ohledu na pravdu“ (Kubička & Csémy, 1999) či „snahu jevit se ve společensky příznivějším světle“ (Svoboda, 2001). Tato tendence samozřejmě vzbuzuje řadu oprávněných otázek k její prevalenci a především efektům sociálně desirabilních odpovědí na testování (Alliger, Lilienfeld, & Mitchell, 1996; Donovan et al., 2003). O to více, že je řadou autorů považována za relativně výrazný a do značné míry konstantní projev osobnosti (např. Furnham, 1986), který vyniká zejména v kontextu tzv. *high-stakes assessments*, tj. kdy je respondent silně motivován dosáhnout určitého testového skóru, např. při ucházení se o práci (Barrick, Mount, & Judge, 2001). Je vhodné zmínit, že může existovat i obrácená tendence, kdy respondent směřuje spíše ke svému podcenění (tzv. *self-handicapping bias*). Principiálně se však jedná o podobný mechanismus, neboť manifestované sebepodcenění má většinou za následek pozitivně laděnou korekci ze strany okolí, kterou respondent očekává (např. Strube, 1986).

Většina autorů se shoduje přinejmenším na dvou rovinách, pomocí kterých můžeme popsat vliv sociální desirability – na dimenzi „**klamání sebe**“ (*self-deception*) a dimenzi „**klamání ostatních**“ (*other-deception* nebo také *impression management*) (viz Figurová, 2007), přičemž obě dimenze jsou pochopitelně spojeny s odlišnými motivačními vzorci (McFarland & Ryan, 2000).

Je řada postupů, které se pokoušejí omezit vliv sociální desirability při užívání dotazníků. Pomineme-li opakované testování, je zřejmě tím nejjednodušším způsobem **odrazování respondenta** od zkreslení odpovědi (např. hrozbou), které sice může přinést lehký pozitivní efekt, např. vyšší sebekritičnost (Dwight & Donovan, 2003), ovšem spojený s rizikem nežádoucí reakce ze strany dotazovaného a s negativním efektem v konstruktové validitě takového měření (Vasilopoulos, Cucina, & McElreath, 2005). Jinou možností je prezentace položek dotazníku tak, aby byly pro respondenta **obtížně dešifrovatelné** z hlediska jejich skutečného zacílení, případně cíleně upravovat jejich pořadí (např. neseskupovat otázky stejné škály). Na rozdíl od možnosti úpravy pořadí (McFarland, Ryan, & Ellis, 2002) empirické poznatky tezi o nižším zkreslení „nezřetelných“ položek z hlediska vlivu sociální desirability příliš nepodporují (tj. že by tyto položky měly vyšší výslednou validitu; např. Viswesvaran & Ones, 1999; Holden & Jackson, 1981).

Další alternativou je pokusit se **formulovat odpovědi ve výběru** tak, aby byla jejich případná (ne)desirabilita srovnatelná. Tedy nastavení, kdy danou škálu definují položky, které jsou ve všech alternativách pro respondenta přijatelné. Byť je možné takový postup v určitých testech využít, není obecně doporučovaný – zejm. vzhledem k riziku zhoršených psychometrických vlastností takto navrženého testu (např. Meade, 2004).

Metodou, která je pak v tomto směru zřejmě nejpoužívanější, je využití **samostatných dotazníků** (či jejich položek), které se snaží detekovat míru sociální desirability a poskytnout tak např. podklady k úpravě váhy, kterou přisuzujeme dalším odpovědím respondenta (resp. úpravě odhadované validity testování). Řada odborníků však považuje tyto testy/položky za málo stylisticky zacílené, tj. že neodrážejí respondentův obvyklý „zkreslovací styl“, ale spíše popisují konkrétní a specifické situace nebo saturují zcela jiné rysy. To ostatně potvrzuje řada výzkumů, které zdokumentovaly např. vysoké korelace mezi skóry sociální desirability a skóry svědomitosti a emoční stability (Swanson & Ones, 2002).

### 2.3.2.2

#### Problematická interpretace termínů a konstruktů respondentem

---

Řada autorů se domnívá, že nedokonalost (sebe)posouzení je způsobena především rozdílnými soudy a asociacemi, které různí respondenti k popisovaným konstruktům mají (např. Griffin & Buehler, 1993; Griffin & Ross, 1991). Například v testu předložená otázka „Projevujete se v diskusi spíše dominantně?“, zjišťující určitý aspekt interpersonálního profilu, může být jedním respondentem interpretována jako míra žádoucí aktivity v komunikaci (např. iniciování rozhovoru) a zvolí tak souhlasnou odpověď, zatímco jiný respondent, projevující jinak zcela shodné charakteristiky, může „komunikační dominanci“ interpretovat jako uzurpování si komunikace a neschopnost přijmout kritiku, a na odpověď odpoví záporně. Tento jev, označovaný jako *differential item functioning* (Holland & Wainer, 1993), je samozřejmě nejvýraznější u položek, které pro respondenta nejsou zcela srozumitelné, případně umožňují vágní interpretaci.

Pro omezení vlivu odlišné interpretace otázky či obecně psychologického konstruktů je v některých novějších výzkumech využíváno metody tzv. **ukotvujících vinět** (*anchoring vignettes*; King et al, 2004) Jak uvádí Voňková (2012, s. 28), „základní myšlenkou metody ukotvujících vinět je nechat respondenty mimo sebehodnotící otázky hodnotit v dané oblasti i hypotetickou osobu popsanou v krátkém příběhu (ukotvující viněti). Vzhledem k tomu, že všichni respondenti hodnotí jednu a tutéž hypotetickou situaci, může být rozdílnost hodnocení ukotvujících vinět interpretována jako rozdílnost v jejich využívání škály. Pomocí takto odhadnuté rozdílnosti můžeme provést „očistění“ sebehodnocení (sebehodnocení) a následně je mezi skupinami respondentů porovnat.“ Tento způsob kalibrace odpovědí je velmi příhodnou metodou, nicméně i on má své nevýhody. Těmi jsou například odborná i časová náročnost (je třeba navrhnout a prezentovat vhodnou hypotetickou situaci a respondenti musejí mít na její pochopení dostatek klidu a času), a dále skutečnost, že to, že je někdo vhodným posuzovatelem předložené situace, nemusí znamenat, že je i dobrým posuzovatelem sebe sama.

Při úvahách o interpretační rovině při výběru odpovědí respondentem je důležité zvážit také vliv prostředí, ve kterém se respondent pohybuje, a vliv sociálního srovnávání, stejně jako míru, do jaké je daný konstrukt náchylný k interpretačnímu zkreslení (Cantor & Mischel, 1979). I zde je patrný vliv sociální desirability, kdy je interpretace konstruktů respondentem značně závislá na tom, zda může svému chování v rámci daného rysu přisoudit **pozitivní konotace**. To dokládají například změny výsledků v (re)testovém sebeposouzení respondentů po jejich informování o tom, že je měřený rys spojený se sociálně žádoucími aspekty (např. „dominance“ s kariérním úspěchem). Respondenti po této informaci obvykle aktualizují, resp. mění interpretaci vlastního chování do více žádoucí podoby (Kunda & Sanitioso, 1989; Dunning & McElwee, 1995). Stojí za pozornost, že i samotná **volba škály**, která má přesněji vystihnout přítomnost nějaké charakteristiky, může být důvodem takto zkreslené odpovědi. Například respondenti hodnotící úspěšnost svého života volí spíše nižší hodnoty v případě, je-li škála značená „0–10“, oproti škále značené „-5–+5“, kde je odpověď „-5“ zřejmě více asociována s neúspěchem než neutrální hodnota „0“ (Schwarz, Knauer, Hippler, Noelle-Neumann, & Clark, 1991).

### 2.3.2.3 Specifika introspekce a extraspekce

---

Zaměříme-li se na problematiku limitů introspekce, resp. na riziko nedostatečného náhledu respondenta na vlastní chování a prožívání, dostáváme se na poněkud tenký led současného odborného diskurzu. Zatímco někteří autoři v duchu Augustiánské a Descartevské tradice považují člověka za nejlepšího znalce sebe sama (viz uvedená citace Allporta či častá teze „nikdo o sobě nemá více dostupných informací, než člověk sám“; Paulhus & Vazire, 2007) a data ze sebeposouzení tak za zcela klíčovou informaci (Funder & Colvin, 1991; Mehl, Gosling, & Pennebaker, 2006; Pronin, Kruger, Savitsky, & Ross, 2001), jiní autoři namítají, že sebeposouzení je značně limitované jak v oblasti popisu emocí, motivace, preferencí, tak i v rovině popisu samotného chování a že vhodní pozorovatelé jsou v tomto směru přinejmenším stejně dobrými hodnotiteli (Vazire & Mehl, 2008).

V roce 1955 Luft a Ingham představili model zvaný *Johari window*, tvořící čtyři kvadranty pokrývající limity a parametry sebeposouzení (S) a posouzení druhým (O), resp. *self- and other-perceptions*. První kvadrant zahrnuje aspekty osobnosti přístupné jak S, tak O (tzv. aréna), druhý kvadrant aspekty přístupné pouze S (tzv. fasáda), třetí aspekty jsou známé pouze O (tzv. slepé místo) a čtvrté aspekty, které nejsou přístupné ani jedné straně (tzv. neznámo) (Luft & Ingham, 1955).

Je zřejmé, že identifikace konkrétních aspektů osobnosti zařazených v jednotlivých kvadrantech by byla přinejmenším žádoucí, a to nejen pro oblast psychologické diagnostiky. Zvláště zajímavá je pak „aréna“, ke které je možné získat obě relevantní informace (S/O), ať už shodné, nebo rozdílné. Řada odborníků se tomuto tématu věnovala s cílem vyjádřit **přítomnost či míru asymetrie** (diskrepance posouzení), která se tak často objevuje mezi S a O (tzv. *self-other asymmetry*, resp. *actor–observer asymmetry*). Jejich studie dokládají nejen „prosté“ odlišnosti ve výsledcích obou skupin (např. Beer & Watson, 2008), ale poukazují také na to, že obě respondentní perspektivy mohou efektivně **predikovat jinou rovinu chování**. Kolar et al. (1996) poukázali například na úspěšnější predikci chování při interakci s cizí osobou z perspektivy S (sebeuposuzovatelem) v rovině hodnocení projevovaného klidu (*calmness*), oproti lepší predikci humornosti, sympatičnosti či arogance z pozice O. Ve výzkumu Vazire a Mehla (2008) pak pozice S predikovala například vyšší míru argumentace (*arguing*), oproti úspěšnější predikci družení (*socializing*) ze strany posouzení O. Velmi zajímavý je pak výzkum Fiedlera et al. (2004), který porovnával výsledky dotazníků zaměřených na psychopatologické charakteristiky vojenských rekrutů a jejich eventuální předčasné uvolnění ze služby. Vyšší predikční schopnost mělo S v rovině paranoidních, hraničních a vyhýbavých rysů osobnosti, zatímco perspektiva O byla pozitivněji spojena s antisociálními, schizoidními, schizotypálními a histrionskými rysy rekrutů.

Vztahování výsledků posouzení k určitému kritériu, které by potvrdilo jejich validitu, např. pozorovatelnému chování, je v tomto směru velmi důležité, protože samozřejmě může nastat situace, kdy v rámci S i O sice získáme shodující se výsledky (tj. *self-other agreement*), ov-

šem víceméně irelevantní; „protože dva posuzovatelé mohou být spolu v perfektní shodě, avšak oba se zcela mýlit.“ (Funder & West, 1993, s. 458). Zde je tedy třeba cílit spíše na tzv. *self-other accuracy*, **přesnost hodnocení**, než na míru shody, a mimo vztahování výsledků k výše uvedeným kritériím, doplnit hodnocení také např. o **sumarizaci většího počtu hodnocení** (tzv. *multiple others rating*).

Je třeba zmínit, že při odhadu relevance a přesnosti posouzení druhou osobou musíme uvážit i další aspekty, týkající se např. obecných tendencí spojených se sociální percepcí. Těmi jsou například **naivní realismus** posuzovatelů či jejich iluzorní **superiorita** (Hoorens, 1993). První z uvedených tendencí vychází z domněnky respondentů, že jejich vnímání reality skutečně reflektuje to, jaké věci jsou, a nepřipouští si (jinak přirozenou a pochopitelnou) omylnost. Druhá pak spočívá v podceňování relevance hodnocení druhých lidí, oproti vlastnímu pohledu na ně (Pronin, Kruger, Savitsky, & Ross, 2001). Tyto okolnosti komplikují zejm. možnost vyjádřit určitou „jistotu v hodnocení“, kterou bychom mohli použít pro kritické hodnocení či relativizaci respondentova posouzení – např. v případě, kdy s popisovanou osobou není dostatečně obeznámen a měl by tento deficit v hodnocení reflektovat a deklarovat. Studie Biesanz & Human (2010) nicméně dokládá, že pokud je hodnotitel přesvědčen (nebo je k tomu pobídnut), že výstižnost hodnocení je obzvláště důležitá, dosahuje při posuzování lepších výsledků, avšak s rizikem, že bude při posouzení augmentovat nepodstatné detaily (ibid.).

#### 2.3.2.4 Východiska asymetrie mezi různými variantami posouzení

---

Není pochyb, že asymetrie mezi sebeposouzením a posouzením druhým člověkem je z hlediska komplexního zhodnocení osobnostních charakteristik poměrně zásadní informací (viz. Andersen, Glassman, & Gold, 1998; Pronin et al., 2001). Její modelové uchopení je však přinejmenším komplikované a souvisí s mnoha sociálně-percepčními mechanismy. Funder (1999) zde např. stanovil čtyři podmínky k úspěšnému zhodnocení sledovaných charakteristik: (1) relevance (charakteristika existuje a poskytuje indikátory), (2) dostupnost (infor-



mace o charakteristice je přístupná hodnotiteli), (3) detekce (informace je povšimnuta/zaznamenána) a (4) utilizace (informace je správně interpretována). Tyto parametry mohou být využity při hodnocení výstupů v obou formách posouzení (S a O), nicméně teorie již nepracuje s jejich vzájemným porovnáním, resp. s odlišnou perspektivou a jiným typem informací. Respondenti například dávají při formování (detekci a utilizaci) své charakterizace větší váhu svým myšlenkám a pocitům, oproti posouzení stejných charakteristik vnějšími posuzovateli, kteří se logicky více soustředí na manifestní projevy (Andersen, Glassman, & Gold, 1998). Také proto jsou sebeposuzovatelé přesnější kupř. ve zhodnocení míry vlastní depresivity, která má sice řadu vnějškových projevů, avšak ty zde nejsou těmi nejrelevantnějšími indikátory (Mehl, 2006a). Na druhou stranu může vést důraz sebeposuzovatelů na myšlenky a pocity k upozadění behaviorální informace v situaci, kdy je právě chování tím nejrelevantnějším indikátorem. Je rovněž možné, že behaviorální informace není sebeposuzovatelům ani dostatečně dostupná (např. výraz jejich tváře), což může dále posílit popisovanou S/O asymetrii.

Z uvedených poznatků vychází **model SOKA** (*Self-Other Knowledge Asymmetry*) (Vazire, 2010). Model SOKA vidí důvod asymetrie mezi S a O jednak v rovině této „odlišné informační perspektivy“ (viz výše; např. vnějškově zřejmé, či více skryté charakteristiky osobnosti), jednak v rovině „motivační signifikance“ (např. ego-relevance hodnocení). Tento dvojí, informační a motivační, rozměr hodnocení ostatně zmiňuje i řada dalších prací (např. Watson, Hubbard, & Wiese, 2000; Funder, 1999). Východiska odlišné informační perspektivy model SOKA definuje takto: „Pozorovatelé vědí více o charakteristikách dané osoby, jedná-li se o vysoce **pozorovatelné** rysy. Sebeposuzovatelé pak vědí více o svých charakteristikách, pokud jsou tyto rysy hůře pozorovatelné“, motivační signifikanci pak: „Pozorovatelé vědí více o charakteristikách dané osoby, jedná-li se o vysoce **evaluativní** (posouditelné) rysy. Sebeposuzovatelé vědí více o svých charakteristikách, jsou-li tyto rysy méně evaluativní“ (Vazire, 2010, s. 284–285). Příbuznost charakteristik motivační signifikance s fenoménem sociální desirability (viz výše) je samozřejmě vysoká – i zde mají respondenti

tendenci k užívání ego-protektivního (obv. pozitivnějšího) pohledu na sebe sama. Tuto tendenci ale můžeme do určité míry sledovat i u pozorovatelů, kdy jejich hodnocení druhého člověka (O) může být pozitivnější než jeho sebehodnocení (S), např. z důvodu citové blízkosti k posuzovanému či pro jeho vnímanou podobnost s posuzovatelem. Pozitivnější hodnocení O například dokládá výzkum Vazirové (2006), ve kterém měla být posouzena fyzická atraktivita jedince jednak ve formě S (jedinec hodnotil svou atraktivitu), jednak formou O (byl hodnocen kamarádem). Výsledky ukázaly, že S bylo méně přesné než O, ale také to, že hodnocení O bylo signifikantně pozitivnější (přisuzovalo hodnocenému vyšší atraktivitu). Přes tento příklad je však ego-protektivní zkreslení z pohledu O obvykle nižší než u pohledu S, kde je efekt ego-protекce a ego-angažovanosti (*ego-involvement*) výraznější (ibid.).

Je zřejmé, že míra asymetrie mezi výsledky sebehodnocení a posouzení jinou osobou je ovlivňována jak posuzovanou charakteristikou, tak osobnostními charakteristikami posuzovatele a posuzovaného, resp. jejich **vzájemnou interakcí**. V tzv. Realistickém modelu správnosti (*Realistic Accuracy Model*, RAM; Funder, 2001) jsou načrtnuty čtyři moderující proměnné, které upravují přesnost hodnocení: „dobrá cílová osoba“ (*good target*), „dobrý rys“ (*good trait*), „dobrá informace“ (*good information*) a „dobrý posuzovatel“ (*good judge*).

Jako „**dobrý rys**“ je v tomto směru označován, v souladu s modelem SOKA, rys pozorovatelný a evaluativní; např. proto je častěji správně odhadována extraverte, průbojnost či komunikativnost, než plachost, tvořivost či zvědavost (Funder, 2012). Např. i Connolly, Kavanagh a Viswesvaran (2007) zjistili, že nejvyšší hodnoty shody S/O (*self-other agreement*) vykazuje posouzení extraverte, oproti nižší míře shody v rámci dalších Big5 charakteristik. Taktéž jiné výzkumy (např. Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008) potvrdily vyšší míru shody u rysu extraverte než např. u neuroticismu. Do určité míry však záleží na metodě, kterou je rys popisován; a tak může rys měřený dvěma metodami (např. dotazníky) vykazovat jinou míru shody S/O (viz např. Watson, Hubbard a Wiese, 2000). Tato rozdílnost může být způsobena např. rozdílnými instrukcemi či formulací a stylizací po-

ložek dotazníků (Allik, Realo, Mõttus, Esko, Pullat, & Metspalu, 2010).

Pokud se zaměříme na téma „**dobrého posuzovatele**“, můžeme identifikovat řadu aspektů, které zvyšují míru shody mezi posouzením S/O. Cronbach (1955 in Watson, Hubbard & Wiese, 2000) například uvádí, že vysokou shodu často vysvětluje prostá asociace posuzovatele k člověku, který je posuzovateli podobný. Následně tak může nastat situace, kdy posuzovatel spíše popisuje sám sebe, avšak je v tomto posouzení úspěšný. Přesnost „dobrého posuzovatele“ je do značné míry závislá také na vztahu a míře obeznámenosti mezi popisovaným a popisujícím (viz např. Letzring, Wells, & Funder, 2006). Shoda posouzení u lidí, kteří se znají dobře a relativně dlouho, je zřejmá a statisticky významná (Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008). Řada výzkumů poukazuje na to, že blízké osoby by měly být schopné lépe vystihnout méně pozorovatelné (vnitřní) charakteristiky popisovaného, než jsou toho schopni cizí posuzovatelé (např. Colvin & Funder, 1991). Někteří autoři pak uvádějí, že přesnost v hodnocení vysoce evaluativních a snadno hodnotitelných charakteristik (např. spokojenosti) má naopak s mírou blízkosti vztahu pouze minimální spojitost (Conelly & Ones, 2010). V této oblasti však nepanuje široká shoda a např. Costa a McCrae k tomuto zmiňují, že (paradoxně) blízcí příbuzní mají potíže reálně odhadnout např. svědomitost (tj. relativně evaluativní charakteristiku) blízkého člověka, a to z důvodu jejich specifického vztahu (Hřebíčková, 2003). Další autoři doplňují, že klíčový je zde především emoční charakter vztahu mezi popisovaným a popisujícím, zejm. v rovině tzv. emočního vkladu (*emotional investment*), který definuje jejich vzájemnou citovou polohu (např. Kenny, 2004). Další výzkumnou evidenci k roli blízkosti vztahu přinášejí autorky modelu SOKA (Beer & Vazire, 2017), které reflektují předchozí informace a uvádějí, že přátelé (blízké osoby) přesněji predikují myšlenky a pocity posuzované osoby, avšak přátelé i cizí lidé jsou stejně úspěšní v predikci chování.

### 2.3.3 Závěr

---

Tato kapitola byla věnována problematice využití dotazníků při testování v psychologii, a to z pohledu zkreslování výsledků testování

vlivem sociální desirability, omezené (sebe)reflexe, nevhodnou interpretací otázek či zkreslením pod vlivem role a perspektivy, kterou posuzovatel k posuzovanému zaujímá. Klíčovým tématem byla v tomto ohledu asymetrie, ke které právě v těchto případech dochází mezi různými variantami posouzení (např. mezi výsledky sebeposouzení a posouzení jiným člověkem).

Přestože mohou zmíněné mechanismy významně ovlivňovat validitu a reliabilitu testování, nejsou v rámci domácí psychodiagnostiky dostatečně diskutovány. Je proto nezbytné se na ně více zaměřit a propagovat zahrnutí adekvátních metodologických postupů do současných (nejen dotazníkových) studií, např. v podobě obohacení tradičního sebeposouzení o další varianty dotazování či další referenční informace. Jedním z výzkumů, který se k tomuto tématu pokusil přispět, byl i náš výzkum CPACT, který mimo běžného sebeposouzení využívá také varianty posouzení druhou osobou (zde deklarující blízký vztah s posuzovaným). Podrobné informace o zvoleném řešení a vybraných výsledcích pak naleznete v následujících kapitolách.

3

# Výzkum CPACT

## 3.1

# Úvod do výzkumu CFACT

*Dalibor Kučera*

Výzkum CFACT, Komputační psycholinguistická analýza českého textu (*Computational Psycholinguistic Analysis of Czech Text*), je projektem základního výzkumu zaměřeným na studium mezilidské komunikace. Ve výzkumu byly sledovány vztahy mezi osobností člověka a slovy, která užívá. Výzkum probíhal od ledna 2016 do prosince 2018 na Pedagogické fakultě Jihočeské univerzity v Českých Budějovicích pod vedením řešitele Dalibora Kučery (autora kapitoly). Projekt finančně podpořila Grantová agentura ČR grantem č. 16-19087S.

V této kapitole budou představena východiska a cíle projektu, pilotní studie, průběh výzkumu, technické zázemí, personální zajištění a etické aspekty výzkumu. Na text pak navazuje kapitola 3.2 zaměřená na podrobný popis výzkumného souboru (resp. výzkumných souborů) a použitých metod, včetně diskuse.

### 3.1.1 Východiska a cíle výzkumu CFACT

---

Projekt CFACT byl plánován a kontinuálně připravován přibližně od poloviny roku 2011, kdy se řešitel seznámil se zahraničními studiemi Jamese W. Pennebaker a jeho spolupracovníků (Pennebaker, Mehl, & Niederhoffer, 2003; Pennebaker, 2011; Pennebaker & King, 1999). V těchto textech byly popsány aktuální výzkumy zaměřené na vztah vybraných charakteristik mluvčích (jejich osobnosti, pohlaví, zdravotního stavu apod.) k podobě textu, který tyto osoby produkují, resp. k výsledku kvantitativní lingvistické analýzy.

Vzhledem k atraktivitě tématu a potenciální široké uplatnitelnosti výsledků v rámci mnoha oborů byla nasnadě otázka, zda by podobné mechanismy spojené s produkcí verbálního komunikátu nemohly být popsány také v českém jazyce. Na základě rešerše bylo zjištěno,

že v dané době není dostupná žádná relevantní literatura či studie, která by se detekci osobnostních charakteristik mluvčích pomocí kvantitativní lingvistické analýzy textu komplexněji věnovala; ať už na úrovni českého, nebo obecně západoslovanského jazyka (pozn.: V kapitole 2.2 jsou zmíněny některé projekty, které jsou tematicky nebo metodologicky blízké, případně se na výše uvedené anglofonní výzkumy odkazují.). Bylo tedy zřejmé, že dané téma je v domácím kontextu poněkud přehlíženo a že nabízí unikátní příležitost k vědeckému bádání.

V souvislosti s rozhodnutím otevřít tuto problematiku dalším zkoumáním bylo třeba sestavit původní výzkumný projekt, stanovit jeho cíle a zvolit vhodnou metodologickou koncepci. Výsledný plán měl podobu kvantitativního extenzivního výzkumu a pracoval jak s verifikačními, tak s deskriptivními a exploračními postupy (viz Reichel, 2009).

**První dva cíle výzkumu** byly formulovány následovně:

1. V rámci českého jazyka testovat hypotézy, které vycházející z anglofonních výzkumů a které předpokládají vztah mezi psychologicky relevantními charakteristikami osob a formálními (tj. nesémantickými) textovými parametry jimi produkovaného komunikátu. Zde byl výzkum orientovaný verifikačně.
2. Pomocí analýzy dat identifikovat další proměnné a vztahy, které dosud nebyly v zahraničních studiích publikovány. V tomto bodě byl výzkum orientovaný na deskripci a exploraci.

Přestože jsou základní cíle poměrně jasně a srozumitelně vymezeny, volba způsobu, jakým jich dosáhnout, byla výrazně komplikovanější. Zahraniční studie pochopitelně využívají různé přístupy, které často není možné kombinovat či zcela přesně replikovat. Pro deskripci a exploraci bylo zároveň vhodné definovat i jiné, specifitější postupy, které by mohly přinést nové a žádoucí výsledky. Zásadními výzvami se pak staly především tyto oblasti:

- Volba zdroje a typu analyzovaného textu;
- volba vhodného výzkumného souboru;
- definice vhodných lingvistických a psychologických (příp. i dalších) proměnných a způsob jejich operacionalizace;

- zvážení srovnatelnosti výsledků získaných na českém jazyce s anglofonními výzkumy.

Tyto a další parametry byly postupně řešeny a vyjasňovány v rámci dodatečných rešerší, odborných úvah a pilotních studií. Výsledek tohoto procesu je pak popsán v následujícím textu a dále v kapitole 3.2.

### 3.1.2 Pilotní studie

---

Prvním krokem ke konkrétnějšímu uchopení celého projektu bylo provedení pěti pilotních studií realizovaných v období 2012–2015 pod názvem QPA-FPT (*Quantitative Psycholinguistic Analysis of Formal Parameters of Czech Text*; Kvantitativní psycholingvistická analýza formálních parametrů českého textu). Na jejich realizaci se mimo osoby řešitele podíleli i dva členové budoucího týmu CPACT – Jana M. Havigerová (FF MU) a Jiří Haviger (FIM UHK). Výhodiska a výsledky studií pak byly prezentovány v rámci čtyř konferencí (Kučera & Haviger, 2013; Kučera & Havigerová, 2014a; Kučera & Havigerová, 2014b; Kučera & Haviger, 2014a; Kučera & Haviger, 2015) a současně byly publikovány čtyři sborníkové příspěvky (Kučera & Haviger, 2014b; Kučera & Havigerová, 2015a; Kučera & Havigerová, 2015b; Kučera & Havigerová, 2015c). Lingvistická data byla zpracována Tomášem Jelínkem z ÚTKL FF UK. Zároveň byly v tomto období podány tři nepodpořené grantové návrhy (GAČR 13-03691S, 14-19158S a 15-16009S).

V rámci pilotních studií byla zjišťována především vhodnost konkrétních postupů sběru dat, využitelnost psychologických testů a volba textových parametrů, které měly být předmětem lingvistické analýzy. Zde uvádíme některé konkrétní poznatky, které z pilotáží vyplynuly:

#### 3.1.2.1 Textové podklady

V rámci pilotáží byly využity různé scénáře, podle kterých účastníci vytvářeli text. Ten vznikl na počítači, obvykle v domácím prostředí. Bylo zjištěno, že různé scénáře, které definuje různá úroveň formálnosti či vztahu autorů k zamýšlenému recipientovi sdělení, vykazují odliš-



né lingvistické kvality (viz např. Eder, Rybicki, & Kestemont, 2016). Pro výzkum CPACT bylo tedy důležité zajistit pestrost sady připravených scénářů (tj. pestrost výsledných textových registrů) a každý scénář (registr) konkrétně vymežit. Vliv registrového nastavení je totiž pro následnou analýzu klíčový (viz např. Cleuziou, & Poudat, 2007; Golcher, & Reznicek, 2011). Dále byla identifikována rizika spojená s neomezenou délkou textu, a to i přes následné zpracovávání textu ve formě vyjádření relativní četnosti textových parametrů vzhledem k celkové délce textu. Jako příklad můžeme uvést např. prodlouženou stať u delších textů, oproti více proporčnímu členění úvod–stať–závěr u textů kratších, při kterém se pak zvýrazňují textové parametry typické právě pro úvod a závěr (viz např. Corral, Boleda, & Ferrer-i-Cancho, 2015). Další riziko souvisí s nekontrolovaným prostředím, ve kterém byl text tvořen – jak bylo zjištěno, někteří autoři totiž text psali nesoustředěně, případně po částech nebo za spolupráce další osoby.

### 3.1.2.2 Soubor

Pilotáží se účastnili většinou studenti vysokých a středních škol, jejichž komunikační dovednosti a slovní repertoár jsou specifické a ne zcela reprezentativní vzhledem k běžné populaci. Při srovnání s lidmi jiné věkové kategorie, zejména se základním vzděláním, bylo patrné, že jazykový projev této skupiny se v mnoha aspektech výrazně liší (viz např. Pennebaker & Stone, 2003; Mulac, Studley, & Blau, 1990).

### 3.1.2.3 Psychologické proměnné

Důležitou podmínkou kvality výzkumu bylo sestavení komplexní a efektivní testové baterie, a to jednak s ohledem na potřebnou kompatibilitu zvolených škál se zahraničními výzkumy, jednak s ohledem na možnost nalézt další relevantní vztahy. V rámci pilotáží bylo administrováno velké množství různých dotazníků, u kterých se sledovalo nejen jejich zacílení, ale také parametry a efektivita jejich administrace. V poslední pilotáži byla dotazníková varianta ve formě sebesouzení (*self-report*) doplněna také variantou „posouzení druhým“ (*other-report*) a tento model byl následně přijat jako funkční a přínosný (viz východiska v kapitole 2.3).

### 3.1.2.4 Lingvistické proměnné

Z hlediska popisu lingvistických proměnných byly pilotáže nejprve zacíleny na analýzu ryze morfologických proměnných (konkrétně 27 izolovaných parametrů, např. slovního druhu, čísla, času apod.) vyjádřených pomocí systému pozičních morfologických tagů (viz kapitola 3.3.2). Bylo však zřejmé, že bude nezbytné sadu proměnných rozšířit do podoby, která by lépe korespondovala s proměnnými, které jsou sledované v anglofonních výzkumech (např. kombinace několika proměnných) a zároveň umožnila vyhledávání vztahů, které by mohly být specifické pro český jazyk.

### 3.1.3 Výzkumné moduly a další cíle projektu

---

Při úvahách o finální podobě projektu byly zvažovány nejen výše uvedené výzkumné oblasti a různé varianty realizace šetření, ale také možná rozšíření výzkumného záměru. Zde bylo rozhodnuto, že mimo uvedené hlavní výzkumné části (tzv. prvního modulu) bude CPACT zahrnovat další dva moduly. Jejich zformování souviselo jednak s atraktivitou souvisejících výzkumných témat, jednak s dostupností odlišného výzkumného souboru a možností více zužitkovat připravené procedury.

Hlavní část výzkumu, modul označený P200 (práce se souborem 200 osob), tak zahrnoval sběr dat v rámci nespécifického souboru, resp. souboru kvótně reprezentujícího populaci ČR v kategoriích pohlaví, věku a vzdělání. Jak bylo uvedeno výše, tyto parametry jsou z hlediska podoby komunikátu zvlášť významné (viz např. Pennebaker & Stone, 2003; Gleser, Gottschalk, & John, 1959; Newman, Groom, Handelman, & Pennebaker, 2008). Účastníci v tomto souboru deklarovali, že jsou rodilými mluvčími češtiny, jsou duševně zdraví a aktuálně nejsou medikováni psychofarmaky (viz P20+). Tento modul sledoval základní projektové záměry a tvořil těžiště projektu.

Druhý modul, označený P20+ (dle původního plánu práce se souborem 20 osob), zahrnoval sběr dat od populace, která byla diagnostikována s depresivní (F31.3–5, F32, F33.1–3) a/nebo úzkostnou

(F40.0–.1, F40.1.–.2) symptomatikou (ÚZIS ČR, 2008). Zatímco první modul navazoval na výše uvedené (a počtem převažující) studie pracující s běžnými soubory, druhý modul umožnil dále rozšířit zacílení a význam projektu pro klinickou populaci. Byl tak formulován **třetí cíl výzkumu**:

3. Pomocí analýzy vztahů mezi proměnnými popsat užívání textových parametrů u osob diagnostikovaných s depresí a/nebo úzkostí a ověřit, zda skladba těchto parametrů vykazuje signifikantní odlišnosti od běžné populace (viz níže), případně zda je možné ji detekovat. Zde byl výzkum zaměřený na deskripci, exploraci i verifikaci.

Východiskem pro vytvoření druhého modulu (P20+) byly především výzkumy popsané v kapitole 3.4.5, které odkazují na specifické textové parametry, které klinická populace při tvorbě komunikátu užívá (např. Demjén, 2014; Holtgraves, 2014). Sběr dat musel být v tomto případě upravený vzhledem požadavkům a potřebám respondentů – tento modul tak vykazuje určité metodologické odlišnosti oproti hlavnímu modulu P200. Celkový počet respondentů se v průběhu realizace výzkumu zvýšil z plánovaných 20 osob na 72 účastníků.

Třetí projektový modul (P2; dle původního plánu využívajícího dvou osob posuzovatelů) byl zaměřen na popis objektivně obtížně kvantifikovatelných charakteristik textu. Klíčová zde byla snaha získat další referenční data, která by nevycházela z využití automatizovaných výpočetních systémů či psychodiagnostických testů, ale ze subjektivního lidského úsudku soustředěného na laickou interpretaci textu jako izolovaného produktu (tj. bez vazby na další informace např. o autorech textů) (viz např. Küfner, Back, Nestler, & Egloff, 2010). Tento záměr byl realizován designací školených osob, tzv. posuzovatelů. Tito posuzovatelé (školení laici, resp. ne-psychologové) měli číst všechny texty a následně je, v souladu s referenčním manuálem (Havigerová, Haviger, & Kučera, 2016), pomocí vlastního odhadu, zkušenosti či dojmu ohodnotit prostřednictvím připravených škál v elektronickém rozhraní. Jejich hodnocení zahrnovalo nejen sledování explicitních a implicitních charakteristik jednotlivých textů (např. míru humornosti), ale také zhodnocení (odhad) osobnostních charakteristik samotných autorů textu (např. úzkostnost autora). Z hlediska cílů výzkumu byl

třetí modul, resp. **čtvrtý cíl výzkumu**, definován následovně:

4. Porovnat výsledky škálování (získané od posuzovatelů textu) s výsledky administrovaných psychologických testů a s výsledky počítační lingvistické analýzy (např. srovnání vyjádření originality textu strojově a posuzovatelem) s cílem identifikovat možné signifikantní vztahy. Zde převládalo deskriptivní výzkumné zaměření.

Východiskem pro vymezení třetího modulu byly nejen teze tzv. folkové psychologie o diagnostické kompetenci laiků (např. Tellegen, 1993), ale především výstupy několika studií, které již s podobným konceptem pracovaly (viz např. Gill, Oberlander, & Austin, 2006; Rodriguez, Holleran, & Mehl, 2010). Je rovněž možné zmínit určitou návaznost třetího modulu na problematiku posouzení druhým, konkrétně tzv. *stranger-report* (viz Vazire & Carlson, 2011), která je popsána v kapitole 2.3. Konkrétní informace o procesu škálování uvádí následující kapitola 3.2. V rámci projektu prováděli škálování celkového počtu 1200 textů (šesti typů textu od 200 účastníků z modulu P200) dva placení posuzovatelé, které doplnilo 8 neplacených posuzovatelů hodnotících pouze část textů. Presentace výsledků této části výzkumu bude součástí samostatné studie.

Tab. 3.1.3 – Shrnutí: Přehled výzkumných modulů projektu CPACT

Modul	Soubor	Cíle
1. P200	Běžná populace	Verifikace výsledků zahraničních studií pro český jazyk
		Explorace a deskripce dalších vztahů text–osobnost
2. P20+	Klinická populace	Explorace a deskripce specifik populace ve vztahu k textu, verifikace výsledků zahraničních studií
3. P2	Běžná populace a školení laici (posuzovatelé)	Deskripce a porovnání výsledků škálování textů (data od posuzovatelů) s výsledky ostatních metod (resp. s daty P200)

### 3.1.4 Harmonogram a průběh výzkumu

---

V období od ledna do srpna 2016 byl sestaven plán výzkumu a připraveny veškeré postupy. V období od září 2016 do prosince 2017 pak probíhal sběr dat a rok 2018 byl věnovaný jejich zpracování, evaluaci a tvorbě publikačních výstupů.

Zásadní podmínkou úspěšné realizace výzkumu bylo jednak připravit vhodný plán sběru dat, jednak získat kvalitní a kompletní výzkumný soubor. V druhém případě se jednalo o velmi náročný proces. Oslovování participantů pro modul P200 probíhalo mnoha způsoby; pomocí sociálních sítí, plakátů, letáků, přímého oslovování institucí (domovy seniorů, občanská sdružení a kluby, univerzita třetího věku, úřady práce, probační a mediační služba apod.), propagace v rámci jiných akcí (sportovní utkání apod.) či prostřednictvím studentů (kteří informace předávali např. svým příbuzným). Zároveň bylo vytvořeno propagační video „Výzkum CPACT – Hledáme mezi lidmi a slovy“ (Kučera, 2017a), stránky projektu na síti Facebook (Kučera, 2017b) a zveřejněn článek ve formě rozhovoru v regionálním tisku (Skálová, 2017). Přes tuto snahu byla nedostatečná naplněnost některých kvót (např. skupiny starších mužů se základním vzděláním) stále velkým problémem, který výrazně komplikoval průběh sběru dat zejména v roce 2017. Situaci znesnadňovalo také to, že pro výzkum byly hledány dvojice osob, ve kterých se oba lidé vzájemně znali, a nikoliv pouze jednotlivci. Získání tohoto souboru tak bylo jednou z největších výzev, které realizace projektu přinesla. V rámci modulu P20+ již byla situace méně náročná a oslovování participantů a práci s nimi kontinuálně zajišťovaly asistentky projektu, které měly s cílovou populací přímý kontakt. Získání osob posuzovatelů pro modul P2 bylo také bezproblémové a spolupráce s nimi probíhala úspěšně v průběhu celé doby řešení projektu.

Samotný sběr dat byl v modulu P200 realizován v podobě tzv. výzkumných sezení. Pomocí elektronického rozhraní se dvojice účastníků, tzv. registrující a partner (viz kapitola 3.2), přihlásily do výzkumu a následně jim byla zaslána nabídka termínů, ve kterých se mohly výzkumu zúčastnit. Kapacita pro daný den byla maximálně 10 účastníků, tj. 5 dvojic. Absolvování výzkumného sezení vyžadovalo celodenní

účast rozdělenou do dvou bloků; dopoledne byla administrována sada testů zaměřená na sebeposouzení účastníka, psány dva texty (dopisy) a absolvován jeden nahrávaný rozhovor, přičemž odpoledne byl průběh shodný, avšak sada testů byla tentokrát zaměřena na posouzení druhého účastníka ve dvojici. Pořadí dopisů a rozhovorů bylo náhodně losováno. Výzkumné sezení probíhalo o víkend, nejčastěji mezi 9.–16. hodinou v prostorách Pedagogické fakulty JU v budově Jeronýmova. Na začátku každého sezení byli účastníci prostřednictvím projektové asistentky, která je provázela celým výzkumným sezením, podrobně seznámeni s průběhem výzkumu a svůj souhlas s účastí vyjádřili podepsáním informovaného souhlasu. V průběhu výzkumného sezení jim bylo umožněno dělat si libovolné přestávky, obracet se na asistentku s jakýmkoliv dotazy a mezi oběma bloky byla naplánována také větší pauza na oběd, kdy bylo možné opustit prostory fakulty. Za účast ve výzkumu bylo dvojici vyplaceno 2500 Kč formou uzavřené DPP. Z každého dne výzkumu (výzkumného sezení) byl pořízen podrobný protokol s dokumentací času a průběhu všech procedur, obsahující také přiřazení unikátních účastnických kódů (které následně nahradily všechny osobní identifikátory).

V rámci modulu P20+ probíhal sběr dat prostřednictvím 3 asistentek, které zajistily sběr dat od klinického souboru v rámci individuálních sezení. V tomto případě nebylo využito elektronického rozhraní (resp. data do něj byla vkládána expost) a administrace testů i sběr textů probíhaly ve formě papír–tužka přímo na místě. V tomto modulu nebyly realizovány nahrávané rozhovory, došlo k redukci rozsahu testové baterie, která byla administrována pouze ve variantě sebeposouzení. U těchto účastníků byla součástí dokumentace také informace o diagnóze a formě léčby. Za účast v modulu P20+ bylo účastníkům vyplaceno 1200 Kč formou uzavřené DPP. Také z této části byl pořízen protokol a účastníkům byly přiřazeny účastnické kódy sloužící k denonymizaci poskytnutých dat.

V rámci modulu P2 probíhalo škálování textů osobami posuzovateli kontinuálně, resp. v závislosti na sběr textů v modulu P200. Posuzovatelé využívali elektronické rozhraní (viz níže), které jim pod kódy předkládalo aktuálně dostupné texty (dopisy i přepisy celých rozhovo-

rů). Za jeden posouzený text byla vyplacena odměna cca 21 Kč formou uzavřené DPP.

V závěrečné fázi projektu byla rovněž realizována 4 setkání s účastníky výzkumu, o kterém byli participanti v předstihu informováni formou hromadného e-mailu. V rámci dvouhodinového programu na půdě PF JU pak byl představen výzkum CPACT, předány výsledky psychologických testů a poskytnut odborný komentář k jejich interpretaci. Výsledky byly předány pouze těm osobám, které se prokázaly platným průkazem totožnosti, případně předložili úředně ověřenou plnou moc od konkrétního účastníka výzkumu. Účastníci měli možnost pokládat libovolné otázky, komentovat průběh výzkumu a poskytl výzkumnému týmu řadu cenných podnětů. Setkání byla samozřejmě nepovinná, bezplatná a dostavilo se na ně celkem 64 účastníků výzkumu.

### **3.1.5 Elektronické rozhraní výzkumu**

---

Pro potřeby výzkumu bylo navrženo a naprogramováno rozsáhlé elektronické rozhraní, které pokrývalo téměř všechny fáze výzkumu a související sběr dat a jejich export. Rozhraní bylo dostupné pomocí počítače (zařízení) připojeného k internetu prostřednictvím domény [www.cpact.cz](http://www.cpact.cz) a jeho zabezpečení zajišťovala víceúrovňová hesla. Rozhraní poskytovalo především tyto služby:

- Zveřejňování informací o projektu prostřednictvím internetových stránek;
- registrace účastníků výzkumu ve dvojici registrující/partner (dynamicky kontrolující parametry zájemců, zejm. naplněnost kvót v jednotlivých kategoriích, viz kapitola 3.2);
- přihlášení uživatele (dvojice účastníků) ke konkrétnímu termínu výzkumného sezení a nastavení přihlášení (log-in) do počítače přímo na výzkumném sezení;
- přiřazení unikátního účastnického kódu ve formátu „číslo“ a „role“ (registrující/partner), např. 123R a 123P;
- vyplňování psychologických testů účastníkem a psaní textů do textové aplikace (aplikace průběžně ukládala všechna data, aby nebyla ohrožena např. výpadkem sítě).

Pro projektový tým pak byly dostupné další služby:

- Editace a export všech dat;
- rozhraní pro škálování textů hodnotiteli;
- rozhraní pro vkládání přepisů rozhovorů.

### 3.1.6 Personální zabezpečení projektu

---

Realizaci projektu zajišťoval tým výzkumníků, tzv. projektový tým, a tým složený z asistentů a technického a administrativního personálu. Projektový tým sestával z odborníků z oblasti psychologie, statistiky a lingvistiky, celkem ze šesti různých institucí. Psychologická sekce zůstala beze změn po celou dobu trvání projektu, lingvistická sekce doznala ve třetím roce (2018) lehké úpravy související se změnou projektové agendy (operacionalizace textových parametrů a zpracování textů vystřídala práce s registrovou variabilitou textů). Složení projektového týmu a podíl na projektové agendě shrnuje tab. 3.1.6a, popis práce asistentů a dalšího personálu pak tab. 3.1.6b.

Tab. 3.1.6.A – Projektový tým CPACT

<b>Člen týmu (specializace)</b>	<b>Pracoviště v roce 2018</b>	<b>Projektová agenda</b>
Dalibor Kučera (psycholog)	Katedra pedagogiky a psychologie, Pedagogická fakulta, Jihočeská univerzita v Českých Budějovicích	Řešitel projektu, design a management projektu, koordinace týmu, propagace projektu, publikace
Iva Stuchlíková (psycholožka)	Katedra pedagogiky a psychologie, Pedagogická fakulta, Jihočeská univerzita v Českých Budějovicích	Konzultační a metodologická podpora, supervize a optimalizace výzkumných procesů, podpora a rozvoj spolupráce
Jana Marie Havigerová (psycholožka)	Psychologický ústav, Filozofická fakulta, Masarykova univerzita	Metodologické řešení projektu, sběr materiálů a dat, podpora při zpracovávání dat, statistická analýza, druhý koordinátor projektu



Tomáš Urbánek (psycholog)	Psychologický ústav, Akademie věd České republiky, v.v.i.	Konzultační a metodologická podpora, podpora v rámci procesů textové analýzy, pod- pora a rozvoj spolupráce
Jiří Haviger (statistik)	Katedra informatiky a kvantitativních metod, Fakulta informatiky a managementu, Univer- zita Hradec Králové	Metodologické řešení pro- jektu, úprava a zpracování projektových dat, aplikace analytických procesů, podpo- ra při zveřejňování výsledků
Tomáš Jelínek (lingvista)	Ústav teoretické a komputační lingvisti- ky, Filozofická fakulta, Univerzita Karlova v Praze	Zpracování výzkumných tex- tů, konzultační a metodolo- gická podpora, aplikace ana- lytických procesů, publikační spolupráce, spolu-koordinace lingvistické sekce
Milena Hnátková (lingvistka)	Ústav teoretické a komputační lingvisti- ky, Filozofická fakulta, Univerzita Karlova v Praze	Úprava a zpracování výzkum- ných textů, identifikace tagů a vybraných jevů v textech, práce s korpusy a slovní- ky, příprava podkladů ke statistické analýze (v projektu 2016–2017)
Hana Skoumalová (lingvistka)	Ústav teoretické a komputační lingvisti- ky, Filozofická fakulta, Univerzita Karlova v Praze	Úprava a zpracování výzkum- ných textů, identifikace tagů a vybraných jevů v textech, práce s korpusy a slovníky, příprava datových podkladů (v projektu 2016–2017)
David Lukeš (lingvista)	Ústav Českého národ- ního korpusu, Filozofic- ká fakulta, Univerzita Karlova v Praze	Zpracování výzkumných textů, aplikace analytických procesů, práce s korpusy a slovníky, příprava podkladů ke statistické analýze, publi- kační spolupráce (v projektu od 2018)

Zuzana Komrsková (lingvistka)	Ústav Českého národního korpusu, Filozofická fakulta, Univerzita Karlova v Praze	Práce s korpusem a slovníky, příprava podkladů ke statistické analýze, příprava publikačních výstupů, interpretace textových dat a parametrů (v projektu od 2018)
Václav Cvrček (lingvista)	Ústav Českého národního korpusu, Filozofická fakulta, Univerzita Karlova v Praze	Interpretace výsledků, aplikace analytických procesů, publikační spolupráce, konzultace a metodologická podpora, koordinace lingvistické sekce (v projektu od 2018)

Tab. 3.1.6.B – Asistenti a další personál v projektu CPACT

Kategorie	Projektová agenda
Primární asistenti	Asistenti provádějící administraci testů, sezení s účastníky, pořizování nahrávek, přípravu podkladů a přenos dat (roli těchto asistentek plnily zejm. studentky psychologie)
Asistenti herci	Asistenti zajišťující vedení řízených rozhovorů s účastníky a pořizování audiovizuálních materiálů v rámci výzkumu
Technický asistent	Asistent zajišťující přípravu a programování aplikace elektronického rozhraní, technickou podporu, údržbu, instruktáž a export dat
Jazykoví asistenti	Asistenti zajišťující překlad projektových materiálů a publikačních výstupů, zejm. do anglického jazyka
Asistenti textu	Role asistent textu pokrývala trojí agendu spojenou s textem: asistenti korektoři (zajišťovali úpravy textu pro následné zpracování), asistenti přepisovatelé (zajišťovali přepis nahrávek rozhovorů) a asistenti posuzovatelé (provádějící škálování všech psaných a mluvených textů).
Administrativní pracovník	Zajištění administrativní agendy projektu na řešitelském pracovišti (PF JU)

Na realizaci projektu CPACT se v průběhu tří let (tj. období 2016–2018) podílelo celkem 34 pracovníků, z čehož 11 osob tvořilo projektový tým, 2 osoby byly administrativními pracovníci a 21 osob působilo v roli asistentů projektu. Většinu pracovníků tvořily ženy (24) a mezi asistenty bylo 5 studentů.

### **3.1.7 Etické aspekty výzkumu**

---

Výzkum CPACT byl realizován ve snaze splňovat kritéria vysoké odborné kvality a společensky prospěšných a významných cílů. Výzkum probíhal v souladu s etickými principy Helsinské deklarace (WMA) a Etickým kodexem JU a byl schválen Etickou komisí Jihočeské univerzity v Českých Budějovicích. Výzkum nebyl spojen s rizikovou činností z hlediska ochrany života, zdraví a majetku.

Účastníci se do výzkumu zapojili dobrovolně a za sjednanou finanční odměnu. Po úvodní prezentaci projektu a po seznámení se s informovaným souhlasem (viz příloha 3.1.7) účastníci svým podpisem deklarovali, že jsou se všemi specifiky výzkumu srozuměni a nemají k nim výhrad. Zároveň udělili souhlas se zpracováním svých osobních údajů, s pořízením zvukových a obrazových materiálů v průběhu výzkumných sezení a zavázali se ke svědomitému a pravdivému vyplnění předkládaných materiálů. Stvrдили rovněž to, že jsou fyzicky i psychicky připraveni se výzkumu zúčastnit a nejsou si vědomi toho, že by jejich účastí mohly vzniknout jakékoliv škody na zdraví či majetku.

Výzkum byl registrován Úřadem pro ochranu osobních údajů (ÚOOÚ) pod registračním číslem 6027 (evidenčním číslem 12). Veškeré poskytnuté materiály byly považovány za osobní a bylo s nimi takto nakládáno. Výstupy výzkumu byly publikovány pouze v anonymizované podobě.

### **3.1.8 Propagační materiály**

---

Jak bylo zmíněno v předchozím textu, oslovování potenciálních účastníků probíhalo pomocí různých informačních a propagačních prostředků. V přílohách prezentujeme některé z nich:

- 3.1.8.A – Logo projektu
- 3.1.8.B – Ukázka webové stránky projektu
- 3.1.8.C – Ukázka propagačního posteru
- 3.1.8.D – Ukázka prezentace na Facebooku
- 3.1.8.E – Ukázka videa na kanálu YouTube
- 3.1.8.F – Ukázka rozhovoru v Českobudějovickém deníku

## 3.2

# Výzkumný soubor a metody

*Dalibor Kučera*

V kapitole 3.1 byl představen výzkum CPACT, Komputační psycholinguvistická analýza českého textu (*Computational Psycholinguistic Analysis of Czech Text*), především z hlediska celkového zaměření projektu a jeho rozdělení do výzkumných modulů. Tato kapitola je věnována podrobnému popisu výzkumného souboru (resp. výzkumných souborů), procedur a metod, kterých bylo v průběhu šetření využito. Text je rozdělen do tří částí, přičemž každá z nich je věnována konkrétnímu výzkumnému modulu. Kapitulu uzavírá oddíl věnovaný diskusi ke koncepci výzkumu a jeho metodologii.

### 3.2.1 Výzkumný modul P200

---

První modul, P200, byl zaměřen na studium komunikátu obecné (nespecifické) populace, resp. na hledání obecných vztahů mezi osobností člověka a produkovaným textem napříč demografickými kategoriemi. Vzhledem k východiskům popsaným v předchozí kapitole bylo pro nás zásadní sestavit dostatečně reprezentativní soubor osob, který zahrne obě pohlaví, různé věkové kategorie, různé úrovně vzdělání a zároveň je klinicky intaktní. Druhým úkolem pak bylo připravit kvalitní testovou baterii, která efektivně a korektně charakterizuje osobnost respondentů. Pro zvýšení přesnosti testování jsme proto využili nejen variantu dotazníkového sebezposouzení, ale také variantu posouzení druhou osobou, která popisovaného dobře zná (viz kapitola 2.3). Tento model si vyžádal specifické nastavení nejen v rámci administrace testů, ale i v rámci realizace celého výzkumu. Výzkumného sezení se totiž současně účastnily dvojice participantů, nikoliv jednotlivci. Modul P200 tvořil těžiště výzkumu a víceméně určoval koncepci dalších modulů projektu CPACT. Jeho výstupy je možné nalézt v samostatných studiích a kapitolách 3.4.2 a 3.4.4.

### 3.2.1.1 Soubor

Velikost souboru pro modul P200 byla stanovena za využití Fisherovy transformace (Fisher, 1921) pro  $\alpha=0,05$ ,  $\beta=0,8$  a  $r=0,3$ , min.  $n=85$ . Tato hodnota byla zaokrouhlena na soubor  $n=100$  mužů a  $n=100$  žen. Reprezentativnost souboru byla podpořena metodou kvótního výběru, a to s vědomím výhod i omezení tohoto přístupu (viz např. Gravetter, & Forzano, 2011). Kritéria pro zařazení do souboru (tzv. *inclusion criteria*) byla věk, pohlaví a dosažená úroveň vzdělání (pouze dokončeného). Konkrétní kvótní hodnoty jsou odvozeny z údajů o populaci ČR dle Českého statistického úřadu (2015).

Tab. 3.2.1.1 – Soubor P200 s indikátory odvozenými z hodnot ČSÚ (2015) ( $n=200$ )

<b>Muži</b> ( $n=100$ )			
Věk	Nejvyšší dokončené vzdělání*		
	ZŠ	SŠ+SOU	VOŠ+VŠ
15-24	6	7	1
25-34	1	11	5
35-55	3	26	7
55+	8	20	5
<b>Ženy</b> ( $n=100$ )			
Věk	Nejvyšší dokončené vzdělání*		
	ZŠ	SŠ+SOU	VOŠ+VŠ
15-24	5	6	1
25-34	1	11	5
35-55	2	23	6
55+	10	24	6

\* Zahrnuje i neukončené základní vzdělání; ZŠ – základní škola a neukončené základní vzdělání; SŠ+SOU – střední škola ukončená maturitou nebo ujučným listem, VOŠ+VŠ – vzdělání ukončené získáním titulu

Účast na výzkumu byla dále podmíněna znalostí českého jazyka (účastníci v tomto souboru deklarovali, že jsou rodilými mluvčími češ-

tiny, českým státním občanstvím a duševním zdravím (účastníci prohlášovali, že nejsou léčeni s psychickými/psychiatrickými obtížemi a že neužívají léky s přímými účinky na duševní stav, např. antidepressiva). Prohlášení o splnění podmínek účastníci provedli podpisem v rámci informovaného souhlasu.

Poslední podmínka vycházela z designu testové části výzkumu. Zde bylo nezbytné, aby se daného výzkumného sezení účastnily vždy dvojice participantů (první v roli tzv. registrujícího, druhý v roli tzv. partnera) a aby se obě tyto osoby ve dvojici dobře znaly a dokázaly se vzájemně podrobně a s nadhledem popsat. Byli proto preferováni blízcí rodinní příslušníci, dlouhodobí partneři nebo dobří přátelé, kteří mezi sebou mají častý a upřímný kontakt. Naopak nebylo vhodné, aby se tyto lidé znali jen zběžně, například byli spolužáky nebo se potkávali jen jednou za čas. Dvojicí mohla být žena a muž, dva muži nebo dvě ženy.

Proces oslovování a registrace účastníků byly popsány v části 3.1.4, zde je však třeba zmínit i průběh naplňování definovaných kvót a procesní komplikace, které s tím byly spojené. Ty souvisely především se snahou dodržet proporce výběru, tj. rovnoměrně zaplňovat všechny kvóty, a zároveň zachovat model dvojic (účasť dvojic blízkých/obeznámených osob), což od účastníků vyžadovalo nadstandardní kooperaci – jak při volbě druhé osoby do dvojice (aby oba ve dvojici spadali do některé z nezaplňovaných kvót), při následné registraci (která mohla být pro některé účastníky komplikovaná), tak při plánování společné účasti na výzkumném sezení ve stanovený den a čas. Je proto zřejmé, že i přes finanční odměnu nebylo toto nastavení pro všechny oslovované osoby atraktivní, případně že nenašli ve svém okolí druhou osobu, která by splňovala potřebná kritéria (zejm. spadala do nezaplňované kvóty). Kvóty s nejproblematičtější zaplněností zahrnovaly muže starší 35 let a ženy starší 55 let se základním a středoškolským vzděláním. V druhé části sběru proto bylo nezbytné uvolnit pravidla registrace a společně s „žádanými“ účastníky (tj. těmi, kteří spadají do ještě nezaplňované kvóty) přijímat i účastníky, kteří primárně plnili roli testových posuzovatelů (tj. poskytovali testové hodnocení „žádoucí“ osoby). Výzkumné nastavení bylo nicméně pro všechny osoby totožné a všichni (i „nežádaní“) účastníci v rámci výzkumného sezení absolvovali kompletní

sadu procedur. Tímto způsobem byla získána data od celkového počtu 248 účastníků, tj. plánovaná kapacita souboru P200 byla překročena o 48 osob. Skupina přesahující kvóty byla tvořena především vysokoškolskými studenty, kteří na výzkum přivedli své příbuzné („žádané“ účastníky). Data všech osob byla archivována, nicméně pro další práci byl tento soubor selektivně redukován opět na 200 osob, které odpovídají všem kvótním parametrům.

### 3.2.1.2 Procedura

Pro sběr dat byla zvolena forma řízeného testování, která je sice organizačně náročná, avšak umožňuje kontrolovat většinu důležitých aspektů testování – zejm. autorství, vnější podmínky, motivaci a seriózní přístup respondenta k plnění úkolů. Účastníci výzkumu tedy vyplňovali testy a psali texty (dopisy) ve vyhrazené počítačové učebně, kde byl klid a kde byla vždy k dispozici asistentka, která mohla s čímkoliv poradit. Administrace testů (viz níže) byla v dopolední i odpolední části výzkumného sezení rozdělena do tří sad a mezi každou testovou sadou bylo elektronickým rozhraním náhodně zobrazeno zadání jednoho ze čtyř dopisů. Mezi popsánymi činnostmi proběhly rovněž dva nahrávané rozhovory (jeden dopoledne a druhý odpoledne, náhodně). Plán výzkumného sezení zobrazuje tab. 3.2.1.2.

Tab. 3.2.1.2 – *Procedury P200: Plán výzkumného sezení*

<b>Blok</b>	<b>Seznam testů</b>	<b>Dopisy</b>	<b>Rozhovory</b>
1. Dopolední (testy k sebeposouzení)	BFI, SMS, DASS1+2, STAIX2, SPSBS	1. dopis	1. rozhovor
	MMG, PSSI1+2, IAS		
		2. dopis	
	SHQ, SSI, BOBR		
Polední přestávka			



	BFI, SMS, DASS1, STAIX2		
2. Odpolední (testy k posouzení druhé osoby)		3. dopis	2. rozhovor
	PSSI1, IAS, DBV		
		4. dopis	
	SHQ, SSI, BOBR		

### 3.2.1.3 Psychologické testy

Podrobný popis testů v testové baterii je uveden v kapitole 3.3.1. Zde se zaměříme především na jejich výběr, úpravu a administraci.

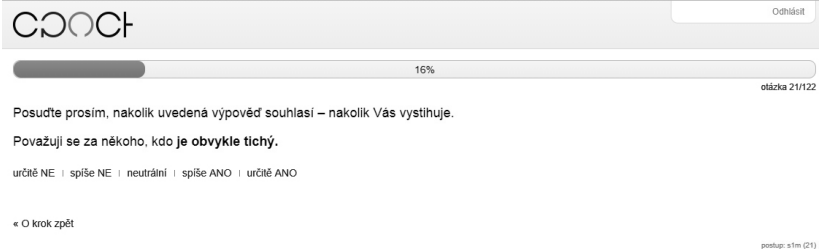
Testy byly v souladu s projektovými cíli vybrány jednak na základě jejich kompatibility s testy využitými v referenčních zahraničních výzkumech (tj. z hlediska potenciálu replikace), jednak na základě jejich relevance vzhledem ke sledované problematice. Testová baterie tak zahrnovala testy zaměřené na obecné rysy osobnosti (BFI), na interpersonální charakteristiky (IAS, SSI), implicitní motivy (MMG) a na charakteristiky související s patologií (PSSI1+2, STAIX2, DASS1+2, SPSBS). Tyto testy doplnily dotazníky zaměřené na humor (SHQ), sebedemonitoring (SMS) a vnímání vzhledu (BOBR). Volba uvedené trojice dotazníků vycházela nejen z atraktivity a potenciálu případných nově objevených vztahů, ale také z kontextu dalších částí výzkumu (např. SHQ a třetí výzkumný modul). Celou baterii uzavíral test DBV, který měl pomoci vyjádřit charakter a blízkost vztahu mezi oběma účastníky ve dvojici. Finální testová baterie tak zahrnovala celkem 361 položek (z 11 psychologických testů) zaměřených na sebeposouzení, tj. varianta S, a 321 položek (z 9 testů) zaměřených na posouzení druhé osoby, tj. varianta O.

Využití varianty O, kdy respondent popisuje druhého účastníka ve dvojici, si logicky vyžádalo úpravy v podobě a skladbě testů v testové baterii. Některé testy bylo možné využít téměř v nezměněné podobě (např. test BFI, kde bylo upraveno pouze zadání „Považuji se za někoho, kdo je...“ a všechny položky testu zůstaly beze změn), jiné bylo třeba upravit z hlediska užívání gramatické osoby (např. test SSI, kde

bylo původní „Velmi rád se stýkám s lidmi.“ nahrazeno „Velmi rád se stýká s lidmi.“). U některých testů či jejich škál (testy MMQ, SPSBS a škály v PSSI2, DASS2) však využití ve variantě O nebylo možné z hlediska samotného obsahu a zacílení otázek, nikoliv jen formulace. Zhodnocení dané škály druhou osobou („vnějším posuzovatelem“) by nebylo proveditelné či adekvátní. Příkladem jsou introspektivní otázky „Je mnoho věcí, ze kterých se mohu radovat.“ (PSSI2), „Pociťoval jsem sucho v ústech.“ (DASS2) či „Snažím se tvářit, že jsem v pohodě.“ (SPSBS). Tyto testy byly proto z testové baterie varianty O vyřazeny.

Jak bylo zmíněno, testy byly administrovány v rámci elektronického rozhraní (viz 3.1.5), které bylo za tímto účelem vyvinuto. Rozhraní po přihlášení kontinuálně předkládalo testy v jednotlivých dopoledních či odpoledních blocích, s možností vyplňování testů kdykoliv přerušit, např. z důvodu občerstvení či z důvodu účasti na jiné aktivitě (zejm. absolvování výzkumných rozhovorů). Otázky (položky testů) byly zobrazovány vždy jednotlivě, popořadě, v okně internetového prohlížeče. Volba se prováděla klikem myši, přičemž systém účastníkům umožňoval návrat na předchozí položku, nikoliv však dále. Čas na odpověď nebyl omezen. Všechny otázky se vypisovaly v podobě, která odpovídala pohlaví respondenta nebo popisované osoby (zejm. v užívání rodu v 1. a 3. osobě). Rozhraní obsahovalo také grafickou osu, která ukazovala pokrok ve vyplňování testů. V případě potřeby se mohl účastník kdykoliv obrátit na asistentku výzkumu, která byla přítomná v místnosti, kde vyplňování testů probíhalo.

### *Obr. 3.2.1.3 – Elektronické rozhraní pro testování (ukázka)*



The screenshot shows a web-based test interface. At the top left, there are four circular icons representing different test sections. A progress bar below the icons shows 15% completion. The question text reads: "Posudte prosím, nakolik uvedená výpověď souhlasí – nakolik Vás vystihuje. Považují se za někoho, kdo je obvykle tichý." Below the question, there are radio buttons for response options: "určitě NE", "spíše NE", "neutrální", "spíše ANO", and "určitě ANO". At the bottom left, there is a navigation button labeled "◀ O krok zpět". At the bottom right, it says "postup: s1m (21)".

Elektronické rozhraní umožňovalo zajistit jak export hrubých dat (odpovědi na položky), tak vyhodnocení z hlediska výsledků jednotlivých škál. Tyto podklady pak tvořily výchozí materiál pro následné statistické zpracování.

#### 3.2.1.4 Textové materiály

Na základě provedených pilotáží bylo postupně definováno šest typů textu, které byly využity ve výzkumu CPACT. Texty byly účastníky vytvářeny na základě samostatně zadaných scénářů a pokrývaly jak písemnou, tak ústní komunikaci. Podrobné informace o profilu jednotlivých textů poskytuje kapitola 3.4.3. V této části pak bude pozornost věnována především psychologickým a metodologickým aspektům textů a koncepci jejich sběru.

##### 3.2.1.4.1 Písemné komunikáty – dopisy

Písemné komunikáty, resp. dopisy, byly zaznamenávané prostřednictvím elektronického rozhraní, ve kterém bylo zobrazeno zadání úkolu (scénář) a pod ním hlavní textové pole, do kterého účastníci psali svůj text. Ve spodní části bylo dále viditelné počítadlo počtu slov, které bylo nastaveno na minimum 180 slov a maximum 200 slov (viz zdůvodnění v části 3.1.2). Bez splnění tohoto kritéria nebylo možné text odeslat. Tento model byl shodný pro všechny účastníky modulu P200; výjimku tvořilo pouze několik účastníků–seniorů, kterým byla nabídnuta možnost záznamu textu ve formě ručně psaného dopisu, který byl až následně přepsán do elektronického rozhraní asistentkou.

### Obr. 3.2.1.4.1 – Elektronické rozhraní pro psaní dopisů (ukázka)

COOCT

Odhlásit

Do připraveného textového pole napište dopis, který odpovídá uvedenému scénáři. Napište ho prosím tak, jak byste právě Vy, ve skutečné situaci, dopis vytvořil a odeslal. Věnujte také pozornost potřebné délce textu, která se zobrazuje pod textovým polem.

**Nalezl jste nabídku práce, která Vás velmi zaujala a moc stojíte o to, abyste právě toto místo získal. Píšete proto řediteli dané firmy, abyste reagoval na jeho nabídku a pokusil se ho přesvědčit o tom, že právě Vy jste ten pravý kandidát na dané zaměstnání.**

Vážený pane řediteli,  
obracím se na Vás

Aktuální počet slov: 7  
Rozsah od 180 do 200 slov

Uložit a zavít

postup: dopis1 (1)

Obecný úvod před zadáním scénáře byl vždy stejný pro všechny čtyři dopisy: „Do připraveného textového pole napište dopis, který odpovídá uvedenému scénáři. Napište ho prosím tak, jak byste právě Vy, ve skutečné situaci, dopis vytvořil a odeslal. Věnujte také pozornost potřebné délce textu, která se zobrazuje pod textovým polem.“ Pod ním následoval tučný text zadání scénáře, který definoval konkrétní komunikační situaci (registr). Jeho formulace byla s ohledem na maximální srozumitelnost přizpůsobena pohlaví účastníka (níže uvádíme obourodé formulace). Výsledné texty měly ve výzkumu označení TXT1–TXT4.

1. **Motivační dopis** (TXT1): „Nalezl(a) jste nabídku práce, která Vás velmi zaujala a moc stojíte o to, abyste právě toto místo získal(a). Píšete proto řediteli dané firmy, abyste reagoval(a) na jeho nabídku a pokusil(a) se ho přesvědčit o tom, že právě Vy jste ten pravý kandidát na dané zaměstnání.“
2. **Dopis z dovolené** (TXT2): „Trávíte čas na úžasné dovolené. Vše vychází podle plánů a užíváte si naplno Vašich oblíbených činností. Rozhodli jste se proto napsat svému kamarádovi (své kamarádce) a přesvědčit ho (ji), aby za Vámi přijel(a) a užíval(a) si ty krásné dny s Vámi.“

3. **Stížnost** (TXT3): „Až donedávna jste celkem spokojeně žil(a) ve svém bytě (domě) a nic Vám nechybělo. Před časem se ale začaly objevovat problémy, které Vám z příjemného bydlení udělaly pekelo. Přestože jste se původně snažil(a) situaci vyřešit „po dobrém“, nepomohlo to. Rozhodl(a) jste se proto napsat na příslušná místa oficiální stížnost.“
4. **Dopis s omluvou** (TXT4): „Provedl(a) jste věc, která velmi ublížila Vašemu vztahu s člověkem, se kterým jste si byli po dlouhou dobu velmi blízcí. Slíbil(a) jste něco, co jste nedodržel(a). Mrzí Vás to, víte, že jste udělal(a) chybu. A protože ho (ji) nechcete ztratit, rozhodl(a) jste se napsat mu (jí) dopis.“

Volba konkrétních scénářů (registrů) vycházela především ze zkušeností s předchozími sběry dat v rámci pilotních studií (viz kapitola 3.1). Byl zde navíc předpoklad, že všechny čtyři texty mezi sebou vykazují zřejmé jazykové odlišnosti – zatímco dopis z dovolené a omluva (TXT2 a TXT4) může být vytvořen v hovorové či obecné češtině, text motivačního dopisu a stížnosti (TXT1 a TXT3) bude zřejmě spíše spisovný a celkově stylově formálnější. Také interpersonální rámec textů byl odlišný – zatímco TXT2 a TXT3 vycházejí z relativně rovnocenného vztahu mezi komunikačními partnery, příp. z určité dominance autora, který je proaktivní, TXT1 a TXT4 jsou založené na nerovnocenném vztahu mezi autorem a adresátem, příp. na určité submisivitě autora, který se musí mnohem více odkrýt či obhájit. Potřebnou ekologickou validitu výsledných textů (zejm. pak podobnost s adekvátními textovými registry běžné komunikace) následně podpořily i výsledky lingvistické analýzy, jejíž dokumentaci je možné nalézt v kapitole 3.4.3.

#### *3.2.1.4.2 Ústní komunikáty – rozhovory*

Jako zdroj dalších dvou komunikátů, tentokrát ústních, byly zvoleny přepisy dvou polostrukturovaných rozhovorů, které účastníci absolvovali s osobami školených herců (asistentů). Délka rozhovorů byla nastavena na 5–15 min. a jejich obsah rovněž respektoval předem daný scénář. Pořadí rozhovorů (tj. který bude realizován dopoledne či odpoledne) bylo pro daný den zvoleno náhodně.

Scénáře rozhovorů byly označeny jako motivační pohovor (TXT5) a vyprávění o příjemném zážitku (TXT6). Jejich zadání uvádíme níže; v uvozovkách je uvedena informace, kterou účastníci získali bezprostředně před započetím rozhovoru od asistentky (před jejich příchodem do místnosti s hercem), za ní je popis situačního nastavení.

5. **Motivační pohovor (TXT5):** „Nyní Vás čeká první (druhý) rozhovor. Představte si následující situaci: Byl jste pozvaný na pracovní pohovor k pozici, o kterou velmi stojíte. Za dveřmi Vás čeká pan ředitel a je na Vás přesvědčit ho, že právě Vy jste tím pravým kandidátem pro danou pozici.“ **Nastavení:** Ve strohé, potemnělé místnosti (nevybavené učebně) sedí za stolem „ředitel firmy“ (starší, autoritativní muž), kamera je umístěna na stativu před účastníkem, jeho směrem je namířeno silné studiové osvětlení a „ředitel“ mezitím s probandem komisně provádí motivační pohovor.
6. **Vyprávění o příjemném zážitku (TXT6):** „Nyní Vás čeká první (druhý) rozhovor. Za dveřmi sedí jedna milá, starší paní. Může to být Vaše známá, sousedka nebo jen sympatická dáma, kterou jste nedávno potkal. A slíbil jste jí, že jí povyprávíte o pěkných zážitcích z Vaší dovolené. Už se na Vás moc těší!“ **Nastavení:** Ve světlé místnosti s křesílky a stolkem (odlišné od předchozí) sedí milá starší žena, která nabízí občerstvení a vesele vítá účastníka. Prosí ho, aby jí vyprávěl nějaký pěkný zážitek, na který si vzpomene. Do vyprávění nezasahuje, pouze podporuje příjemnou atmosféru, uvolněnou komunikaci a motivuje k ní účastníka.

Jak je zřejmé, zaměření obou rozhovorů do značné míry korespondovalo se scénáři psaných textů, kdy TXT1 (motivační dopis) byl obsahově podobný TXT5 (motivační pohovor) a TXT2 (dopis z dovolené) korespondoval s TXT6 (vyprávění o příjemném zážitku). Zároveň spolu oba rozhovory, z hlediska obsahu komunikační situace, poměrně výrazně kontrastovaly. Zatímco motivační pohovor nebyl pro většinu účastníků příliš příjemnou konverzací (v řadě z nich probouzel nejistotu, nervozitu či ostych), vyprávění o příjemném zážitku hodnotila většina účastníků velmi kladně. Oba dva komunikáty tak vykazovaly svébytnou komunikační modalitu, která pokrývala jak uvolněnou konverzaci, tak konverzaci ve formálních a/nebo nepříjemných situacích.

Cílem obou rozhovorů bylo scénář maximálně přiblížit reálným podmínkám, avšak při zachování kontroly prostředí a shodného průběhu předdefinovaných skriptů pro všechny účastníky. Podklady k oběma rozhovorům byly proto velmi důkladně připravovány a testovány a herci byli ve vedení rozhovoru opakovaně školeni. Role herců byly označeny „Ředitel“ (TXT5) a „Babička“ (TXT6). Každý z herců měl zajištěnou alternaci (proškoleného náhradníka), nicméně nebylo jí téměř třeba (Babička vedla rozhovor se všemi účastníky, Ředitel měl alteraci pouze u 23 rozhovorů). Níže uvádíme nejdůležitější parametry a zásady, které s oběma rozhovory souvisejí.

### **Motivační pohovor (TXT5)**

- **Ředitel** (herec): Muž, 67 let, vysokoškolsky vzdělaný

Charakteristiky komunikace:

- Ředitel je výborně upravený, v tmavém obleku, nepoužívá parfém.
- Tváří se vždy neutrálně (neusmívá se, ani nemračí).
- Není příliš angažovaný (nejeví přehnaný zájem, ani nepovzbuzuje).
- Pozorně sleduje upřeným pohledem účastníka.
- Na vše reaguje klidně, spíše pomalu, nekomentuje, je maximálně stručný.
- Krátké ticho mu nevádí (pouze pozorně sleduje účastníka).
- Nepohybuje se, vstane pouze na konci při loučení a podání ruky (které iniciuje on), případně vstane, pokud na začátku účastník vyžaduje podání ruky.

Osnova rozhovoru:

- Dobrý den.
- Hned se Vám budu věnovat. (15–30 sekund si něco hledá, píše si poznámky, pohledem průběžně sleduje probanda apod.)
- Mohl byste se v krátkosti představit?
- Mohl byste zmínit nějakou svou pracovní zkušenost?
- O jakou pracovní pozici byste měl zájem?
- Co můžete naší společnosti nabídnout?
- A jaké jsou Vaše slabé stránky?
- Opravdu si myslíte, že na tuto práci stačíte? (Proč myslíte?)

- Napadá Vás ještě něco, co byste o sobě mohl říct?
- (V tom případě) Vám děkuji.

Doporučené fráze:

- Řekněte mi víc.
- Nebud'te stručný.
- Zkuste se zamyslet.
- Máte na odpověď čas.
- Co Vám přijde důležité.
- Co ještě?
- Proč myslíte?
- Proč zrovna toto?
- Pokuste se prosím uvolnit a soustředit na náš rozhovor.  
Je to nezbytné. (v případě výhrad účastníka)

### Vyprávění o příjemném zážitku (TXT6)

- **Babička** (herec): Žena, 68 let, středoškolsky vzdělaná

Charakteristiky komunikace:

- Babička je přiměřeně upravená, v pohodlném ošacení.
- Usmívá se, živě gestikuluje, pozorně sleduje a přikyvuje, nonverbálně podporuje.
- Občas se může naklonit, jinak sedí klidně – nenarušuje osobní zónu účastníka.
- Uvitá účastníka tím, že vstane, pozdraví a podá mu ruku.
- Snaží se minimalizovat dlouhé pauzy – povzbuzuje doplňujícími otázkami, reakcemi.

Osnova rozhovoru:

- Dobrý den, to jsem ráda, že Vás vidím!
- Pojd'te si ke mně sednout! Nedáte si čaj nebo vodu?
- Slyšela jsem, že jste byl na jedné skvělé dovolené.  
To mě moc zajímá. Povídejte mi o tom!
- Na co nejraději vzpomínáte?
- Cestuje rád sám, nebo s někým?
- Co se Vám na tom líbí?
- A plánujete něco dalšího, nějakou další cestu nebo výlet?



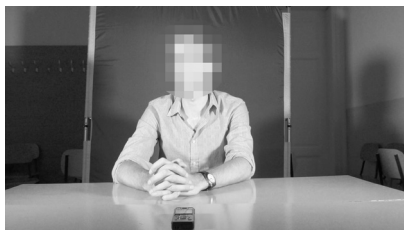
- Tak Vám děkuju, bylo to moc příjemné.

Doporučené fráze:

- To mě zajímá.
- To mi musíte říct.
- To zní moc zajímavě.
- Povídejte mi o tom.
- To zní výborně, no a co třeba...
- A jak to bylo?
- Nedělejte si s tím hlavu, udělejte si pohodlí a něco mi povězte. (v případě výhrad účastníka)

Jak bylo zmíněno, rozhovory se odehrávaly ve dvou různých místnostech. Místnost vyhrazená pro **motivační pohovor** byla strohou učebnou se zataženými závěsy, ve které byla ve středu umístěna řada čtyř lavic a na opačných koncích řady postavené židle. Vzdálenější židle byla obsazena hercem–Ředitelem, který měl na stole před sebou postavenou 4K kameru s externím mikrofonem a položený diktafon. Za židlí pro účastníka bylo umístěno zelené plátno (zvolené pro případné klíčování pořízeného záběru). Po stranách lavic byly umístěny kompaktní studiové reflektory, které mířily na účastníka z obou stran. Děj v místnosti nahrávala z pravé stěny také malá HD kamera se širokouhlým objektivem. Důvodem pro specifické světelné nastavení místnosti byla především snaha vyvolat určitý komunikační diskomfort, spočívající např. v nemožnosti navázat oční kontakt s osobou za reflektory. Oboustranné nasvětlení zároveň umožnilo zajistit kvalitní videozáznam, využitelný v rámci dalšího výzkumu.

*Obr. 3.2.1.4.2.A – Nastavení záznamu motivačního pohovoru (přední a boční kamera)*



Místnost vyhrazená pro **vyprávění o příjemném zážitku** byla odděleným prostorem Vysokoškolské psychologické poradny JU. Vybavení tvořila dvě pohodlná křesílka, mezi nimi konferenční stůl. Místnost byla přirozeně osvětlená velkými okny. 4K kamera i HD kamera zde byly umístěny v opačném rohu místnosti, aby neodváděly pozornost účastníka (případně ho nerušily či neznervózňovaly). Diktafon byl umístěn pod deskou stolu. Herečka–Babička účastníkům nabízela teplý čaj nebo vodu. Vzhledem k uvolněné atmosféře probíhaly tyto rozhovory obvykle delší dobu než motivační pohovor.

*Obr. 3.2.1.4.2.B – Nastavení záznamu vyprávění o příjemném zážitku (diagonální a boční kamera)*



Uvedené nastavení bylo v rámci obou místností zachováno po celou dobu sběru dat a téměř pro všechny účastníky. Výjimku tvořilo pouze 23 rozhovorů, které musely být relokované do dvou odlišných, avšak srovnatelných prostor. Úpravu prostor pravidelně zajišťovaly výzkumné asistentky.

Po pořízení rozhovorů byly audionahrávky předány asistentům–přepisovatelům, kteří na základě připraveného manuálu přepsali všechny rozhovory do písemné podoby a vložili je do elektronického rozhraní výzkumu. Formát přepisu byl definovaný zejména potřebou text následně komputačně zpracovat, z čehož vyplývalo např. i odstranění promluv herců a zachování pouze komunikátu produkovaného účastníkem.

Přepis rozhovorů měl mj. tyto parametry:

- Označení promluv herce a účastníka specifickým kódem.
- Nefonetická transkripce se zachováním charakteru slov obecné češtiny či zkomolenin.
- Doplnění interpunkce dle pravidel českého pravopisu.
- Zaznamenání akustického (nonverbálního) přerušení výpovědi křížkem #, apoziopuze trojtečkou a srozumitelných citoslovcí tak, jak jsou pronesena.

Export přepisů pro následné zpracování byl zajištěn pomocí exportní funkce elektronického rozhraní výzkumu. Průměrná délka rozhovorů byla (bez započítání promluv herců) 760 slov u motivačního pohovoru (TXT5) a 1408 slov u vyprávění o příjemném zážitku (TXT6).

### **3.2.2 Výzkumný modul P20+**

---

Modul P20+ byl zaměřen na studium vztahu komunikátu k osobnostním charakteristikám a psychickému stavu specifické populace, konkrétně účastníků, kteří vykazovali diagnózu depresivní (F31.3–.5, F32, F33.1–.3) a/nebo úzkostné poruchy (F40.0–.1, F40.1–.2) (ÚZIS ČR, 2008). Tento soubor bude dále označován jako klinický. Design modulu vycházel z parametrů P200, nicméně vzhledem k povaze základního souboru, zejm. jeho nižší dostupnosti, zatíženosti a etickým aspektům bylo třeba řadu procedur upravit. Mezi nejdůležitější změny patřila redukce testové baterie, změny v administraci testů a odebrání varianty O (posouzení druhým) z plánu testování. Vzhledem k rozdílným místům, kde byla data sbírána, a problémům souvisejícím s transportem techniky a herců nebylo zároveň možné zajistit srovnatelné (zejm. kontrolované) nastavení ústních rozhovorů. V tomto modulu proto probíhal pouze sběr písemných komunikátů. Část výstupů modulu P20+ je možné nalézt v kapitole 3.4.5.

#### **3.2.2.1 Soubor**

Klinický soubor tvořilo 72 osob s depresivní a úzkostnou symptomatikou dle vyjádření odborného lékaře nebo psychologa. Tyto osoby

byly získány příležitostným výběrem. Mimo uvedených diagnóz byli do výzkumu zařazeni i pacienti s dalšími komorbidními diagnózami (např. porucha osobnosti), kteří vykazovali známky deprese a úzkosti. Vzhledem k zapojení několika asistentů, zajišťujících sběr dat z této populace, byl v průběhu realizace výzkumu počet postupně navýšen z plánovaných 20 osob na celkových 72 osob.

Sběr dat probíhal v rámci několika klinických pracovišť v místech Hradec Králové a Kroměříž, kde byli potenciální účastníci oslovováni pomocí letáků a/nebo prostřednictvím osoby psychologa. Do výzkumu byly zařazeny pouze osoby, které svobodně a s plným vědomím souhlasily s účastí ve výzkumu a svůj souhlas stvrdily podpisem informovaného souhlasu.

V souboru P20+ bylo přítomno 52 osob s diagnózou deprese (F3) a 36 osob s diagnózou anxiety (F4); 16 osob zde vykazovalo smíšenou symptomatiku. 40 osob v souboru bylo aktuálně medikováno psychofarmaky. Nejpočetnější skupinu z hlediska věku tvořili účastníci ve věku 35–55 let (43,1 %), z hlediska pohlaví ženy (45, tj. 62,5 %) a z hlediska vzdělání SŠ vzdělaní (46, tj. 64 %). U 9 respondentů nebyly údaje o vzdělání k dispozici.

Tab. 3.2.2.1 – Soubor P20+ (n=72)

<b>Muži (n=27)</b>				
Věk	Nejvyšší dokončené vzdělání*			
	ZŠ	SŠ+SOU	VOŠ+VŠ	Neuvedeno
15-24	0	4	0	2
25-34	1	3	3	0
35-55	1	7	2	0
55+	0	4	0	0
<b>Ženy (n=45)</b>				
Věk	Nejvyšší dokončené vzdělání*			
	ZŠ	SŠ+SOU	VOŠ+VŠ	Neuvedeno
15-24	0	4	1	1
25-34	2	5	3	3

35-55	2	15	2	2
55+	0	4	0	1

\* Zahrnuje i neukončené základní vzdělání; ZŠ – základní škola a neukončené základní vzdělání; SŠ+SOU – střední škola ukončená maturitou nebo výučním listem, VOŠ+VŠ – vzdělání ukončené získáním titulu

### 3.2.2.2 Procedura

Výzkumná sezení s účastníky P20+ probíhala nejčastěji v podobě individuálních schůzek s asistentem výzkumu, obvykle v prostorách daného klinického zařízení. Po představení projektu CPACT byl účastníkem podepsán informovaný souhlas a přistoupilo se k administraci testové baterie a psaní výzkumných textů, v obou případech metodou papír–tužka. Tato činnost byla opět rozdělena do dvou bloků (1T a 2T), mezi kterými byla přestávka. Jeden blok trval cca 45–60 minut a v některých případech byly oba bloky rozděleny delším časovým úsekem (např. i vícedenním), dle potřeb a možností konkrétního účastníka výzkumu. Pořadí dopisů bylo určeno náhodně. Plán výzkumného sezení zobrazuje tab. 3.2.2.2.

Tab. 3.2.2.2 – Procedury P20+: Plán výzkumného sezení

Blok	Seznam testů	Dopisy
1. blok 1T	1T-1: BFI, DASS1+2	1. dopis
	1T-2: SHQ, SPSBS	2. dopis
Přestávka		
2. blok 2T	2T-1: DASS1, MMG	3. dopis
	2T-2: SSI	4. dopis

### 3.2.2.3 Psychologické testy

Stejně jako u modulu P200, i zde odkazujeme na podrobný popis testů v kapitole 3.3.1. Testy pro modul P20+ byly vybrány zejména s ohledem na cílové charakteristiky, ale také s ohledem na jejich délku a náročnost vyplňování, která by mohla být pro některé účastníky limitující. Výsledná testová baterie opět zahrnovala testy zaměřené na obecné rysy osobnosti (BFI), na interpersonální charakteristiky (SSI), implicitní motivy (MMG) a na charakteristiky související s patologií (DASS1+2, SPSBS), které doplnil dotazník zaměřený na humor (SHQ). Finální testová baterie zahrnovala celkem 206 položek (ze 6 psychologických testů) zaměřených výhradně na sebeposouzení.

Jak bylo zmíněno, testy byly administrovány metodou papír–tužka, se záznamem odpovědí do archu o pěti stranách. Zadání bylo umístěno na sedmi samostatných stranách. Čas na odpověď nebyl omezen, všechny otázky byly prezentovány v obourodém zadání. Po odevzdání testů byla data (odpovědi) asistentem vložena do elektronického rozhraní.

Obr. 3.2.2.3 – Administrace testů P20+ – zadání a záznamový arch (ukázka)

#### ZADÁNÍ CPACT P20 - 2T-1

Platilo pro Vás tato tvrzení?


**Použití prosím, do jeho míry vystihují uvedená tvrzení Vaše povahu a postování RŮZNĚ UPOVĚŘENO VÝNO.**

vůbec ne  málo  často  téměř vždy


89. Bylo pro mě těžké se zkládat.
90. Zdálo se mi, jako bych neokázal(a) prokázat žádné přímé povy.
91. Bylo pro mě obtížné najít v sobě část něčeho dítěte.
92. Měl(a) jsem sklon podávat v některých situacích přeháně.
93. Chlil(a) jsem, že jsem nespjatý (naspjatá) a nervozní.
94. Měl(a) jsem pocit, že se nemám na co tědit.
95. Přistihl(a) jsem se, že jsem rozrušený (rozrušená).
96. Bylo pro mě těžké relaxovat (odpočívat).
97. Chlil(a) jsem se silně a pomatěle.
98. Nevěděl(a) jsem, když mĚ něco odvdělo od práce.
99. Neokázal(a) jsem se pro své nadchnout.
100. Měl(a) jsem pocit, že jako člověk za moc nestojím.
101. Chlil(a) jsem se docela podřaděn.
102. Připadalo mi, že žít nemá smysl.

**Všimte se role člověka osobu na obrázku a použijte, jestli se tvrzení k dané situaci hodí, nebo nehodí. Vyberte si tvrzení (připadne) i více tvrzení, či žádná, které se podle Vás HODÍ.**

103. Obrázek 1



104. Obrázek 2



#### ZÁZNAMOVÝ ARCH 2T-1

Svou odpověď prosím zaznamenejte vyčíslením správného počtu.

0  1  2  3  4  5

**Příklad se chystá opovřit, najít výměrně opovření počítá a to špatně proto vidělat překřátat.**

0  1  2  3  4  5

**Opravovl jte svou odpověď na variantu 5.**

Identifikační číslo:

Datum vyplnění:

Čas:

**Vyčíslete písmena těch tvrzení, která se podle Vás k dané situaci HODÍ (může jich být více). Nevyčíslete tvrzení zranovná, že se jako tvrzení k situaci NEHODÍ (může jich být více, nebo jedno, nebo žádná).**

3

103. Obrázek 1
  - Člověk je rád, že potkal družbu.
  - Žáde mláde člověk ztratil respekt.
  - Ohavně se mocí jinych.
  - U sličtu úloh myslí člověk na nedostatek schopenosti.
  - Žáde raději tĚže úlohy nevděti hod.
  - Člověk dovd, že se v druhém mláti, mlyt sám přivezene iniciativu.
  - Žáde mláde člověk sňat vĚti respekt.
104. Obrázek 2
  - Žáde mláde člověk ztratil respekt.
  - U sličtu situacích vĚti ve vlastni úpěch.
  - Žáde mocou člověk druzi leke odmlit.
  - U sličtu úloh myslí člověk na nedostatek schopenosti.
  - Ohavně se mocí jinych.
  - Člověk má strach, že drubeho mud.
105. Obrázek 3
  - Člověk je rád, že potkal družbu.
  - Žáde mláde člověk ztratil respekt.
  - Ohavně se mocí jinych.
  - Žáde raději tĚže úlohy nevděti hod.
  - Člověk dovd, že se v druhém mláti, mlyt sám přivezene iniciativu.
106. Obrázek 4
  - V sličtu situacích vĚti ve vlastni úpěch.
  - Žáde mocou člověk druzi leke odmlit.
  - U sličtu úloh myslí člověk na nedostatek schopenosti.
  - Člověk počítáe hodot, že sece skláde.
  - Člověk má strach, že drubeho mud.
  - Žáde raději tĚže úlohy nevděti hod.
  - Žáde mláde člověk získat vĚti respekt.
107. Obrázek 5
  - Člověk je rád, že potkal družbu.
  - Žáde mláde člověk ztratil respekt.
  - Žáde mocou člověk druzi leke odmlit.
  - U sličtu úloh myslí člověk na nedostatek schopenosti.
  - Člověk počítáe hodot, že sece skláde.
  - Žáde raději tĚže úlohy nevděti hod.
  - Člověk dovd, že se v druhém mláti, mlyt sám přivezene iniciativu.
108. Obrázek 6
  - V sličtu situacích vĚti ve vlastni úpěch.
  - Žáde mocou člověk druzi leke odmlit.
  - U sličtu úloh myslí člověk na nedostatek schopenosti.
  - Člověk počítáe hodot, že sece skláde.
  - Člověk má strach, že drubeho mud.
  - Žáde raději tĚže úlohy nevděti hod.
  - V sličtu situacích čer mláde člověk.

76 |

Výzkum CPACT: Komputační psychologická analýza českého textu

### 3.2.2.4 Textové materiály

Sběr textových materiálů P20+ vycházel z modulu P200, avšak pouze v rovině práce s písemnými texty (čtyřmi dopisy), nikoliv rozhovory. Smysluplnou realizaci výzkumných rozhovorů zde nebylo možné zajistit, už jen vzhledem ke specifickým podmínkám prostředí i samotného výzkumného souboru. Odlišné nastavení P20+ si zároveň vyžádalo změny ve způsobu, jakým byly texty zaznamenávány. Vzhledem k tomu, že často nebylo možné využít výpočetní techniku, tj. ani elektronické rozhraní, byly texty účastníky vytvářeny klasickým způsobem; ručně psaným písmem na připraveném oboustranném listu. Na úvodní stránce bylo zadání scénáře, které přesně kopírovalo text z elektronického rozhraní (viz P200), pod ním pak osmimilimetrové linkování, do kterého účastník psal svůj text, a které pokračovalo na druhé straně. Poslední věta úvodní informace z materiálů P200 byla nahrazena textem: „Váš text by měl mít délku alespoň dvě strany“, což definovalo očekávaný rozsah textu. Délka textu napříč výzkumným souborem však poměrně viditelně variovala, což bylo dáno jednak rukopisem účastníka, jednak jeho vyjadřovacími schopnostmi a ochotou komunikovat. Po sběru všech listů byly texty asistentem přepsány do elektronického rozhraní, s vypnutou limitací délky textu.

Obr. 3.2.2.4 – Zadání textu P20+ s příkladem zápisu (ukázka poloviny strany)

#### ZADÁNÍ TEXTU - MOTIVAČNÍ DOPIS

1

Na tento list napište dopis, který odpovídá uvedenému scénáři. Napište ho prosím tak, jak byste právě Vy, ve skutečné situaci, dopis vytvořil(a) a odeslal(a). Váš text by měl mít délku alespoň dvě strany.

„Nalezl(a) jste nabídku práce, která Vás velmi zaujala a moc stojíte o to, abyste právě toto místo získal(a).“

Píšete proto řediteli dané firmy, abyste reagoval(a) na jeho nabídku a pokusil(a) se ho přesvědčit o tom, že právě Vy jste ten pravý kandidát na dané zaměstnání.“

Vážný pan řediteli  
Při předložení doporučení firmy jsem mnohokrát  
na internetu viděl podobné. Sám jsem do  
Váší kolektivní zaměstnání pravdě  
podobně nechtěl jít.  
Děkuji si za Vaši nabídku a děkuji mnohokrát.  
Uvěřím si, že budu moci být pro Vaši  
firmu v mnohém dobrým zaměstnancem.

### 3.2.3 Modul P2

---

Relativně samostatnou částí výzkumu byl modul P2 – hodnocení výzkumných textů formou jejich škálování lidskými posuzovateli. Východiska tohoto procesu jsou popsána v části 3.1.3. Zde se zaměříme na realizaci tohoto procesu a konkrétní materiály, na základě kterých byly texty škálovány.

Modul P2 pracoval s posuzováním všech textů pořízených v rámci modulu P200 (viz výše). Zde se tedy jednalo o 800 textů dopisů a 400 textů přepisů rozhovorů (s průměrnou délkou 1084 slov). U rozhovorů byl text, pro vyšší srozumitelnost obsahu komunikace, doplněn i o promluvy herce, které se jinak ve výzkumu nezpracovávaly.

Vzhledem k tomu, že posuzovatelé měli hodnotit texty, jejichž rozsah přesahoval půl milionu slov, jednalo se o časově enormně náročný proces. Navíc bylo velmi důležité, aby posuzovatel zhodnotil všechny texty, nikoliv jen jejich výseč (z důvodu následného srovnání), a aby byl k textu i samotnému hodnocení maximálně pozorný. Na tuto pozici byly určeny dvě osoby, z hlediska psychologie laici, kteří byli za tuto práci honorováni. Oba posuzovatelé byli seznámeni s referenčním manuálem (Havigerová, Haviger, & Kučera, 2016; viz níže) a pomocí navazujících tréninků potřebně proškoleni a zkalibrováni. Oba posuzovatelé rovněž vyplnili baterii psychologických testů, popisující jejich osobnostní charakteristiky pro potřeby případných navazujících studií.

Charakteristiky posuzovatelů:

- Posuzovatel A: Žena, 65 let, vysokoškolské vzdělání společenskovědního (pedagogického) zaměření
- Posuzovatel B: Muž, 35 let, vysokoškolské vzdělání společenskovědního (pedagogického) a humanitního zaměření

Oba posuzovatelé byli vybráni na základě své spolehlivosti, citlivosti ve směru k textu a schopnosti interpretovat základní psychologické konstrukty, které byly součástí hodnocení. Pomineme-li využívání manuálu, hodnotitelé při svém odhadu nejčastěji vycházeli z východisek tzv. folkové psychologie (viz Tellegen, 1993), čímž do určité míry re-



prezentovali přístup běžné, vzdělané, avšak nep psychologické populace.

Dvojici hlavních posuzovatelů v průběhu let 2017–2018 doplnilo i osm neplacených posuzovatelek (posuzovatel C–J), které si ke škálování zvolily část výzkumných textů. Tímto způsobem bylo posouzeno celkem 5886 textů, tj. průměrně 736 zhodnocených textů na osobu posuzovatelky.

### 3.2.3.1 Metodika škálování pro posuzovatele

Metodika škálování byla navržena s cílem definovat jednak to, co má být v textu konkrétně hodnoceno a pomocí jakých škál, jednak zachovat určitou míru svobodného úsudku posuzovatele ve směru k textu (např. na základě vlastní zkušenosti či dojmu).

Hodnocení textu probíhalo pomocí elektronického rozhraní, ve kterém byly texty označeny pouze kódem, bez jakékoliv dodatečné identifikace autora, tj. dvojitě zaslepeně. Charakteristiky autora (např. věk a vzdělání) bylo tedy možné odhadnout pouze na základě samotného textu.

Z tohoto hlediska bylo hodnocení rozděleno do tří kategorií škál:

- **Dojem z autora textu** (položky A01–A08) – odhad toho, jaký by autor mohl být z hlediska osobnostních charakteristik (tyto škály vycházely především z relevantních osobnostních charakteristik obsažených v psychologických testech modulu P200);
- **dojem z textu** (jako takového) – jak samotný text působí z hlediska jeho originality a humornosti;
- **výskyt konkrétního jevu v textu** – parametry textu, které jsou obtížně kvantifikovatelné pomocí běžných počítačových postupů.

Tab. 3.2.3.1 – Kategorie škálování textu posuzovateli v modulu P2

Kategorie	Škála	Popis škály	Rozsah	Slovní vyjádření
Dojem z autora textu	SDR	Ao1 Samotářský – družný	0–4	velmi samotářský – velmi družný
	PA	Ao2 Pasivní – aktivní	0–4	velmi pasivní – velmi aktivní
	RP	Ao3 Rezervovaný – přívětivý	0–4	velmi rezervovaný – velmi přívětivý
	UU	Ao4 Úzkostný – uvolněný	0–4	velmi úzkostný – velmi uvolněný
	PO	Ao5 Pesimistický – optimistický	0–4	velmi pesimistický – velmi optimistický
	SDO	Ao6 Submisivní – dominantní (sebejistý)	0–4	velmi submisivní – velmi dominantní
	O	Ao7 Otevřený novým zážitkům	0–2	málo otevřený – velmi otevřený
	S	Ao8 Svědomitý	0–2	málo svědomitý – velmi svědomitý
Dojem z textu (jako takového)	OT	Originalita textu	0–2	ne – velmi
	HT	Humornost textu	0–2	ne – velmi
Výskyt konkrétního jevu v textu	MMT	Metafory a metonymie	0–2	ne – mnoho
	UST	Užití slangu	0–2	ne – mnoho
	TET	Typografická a estetická úprava textu	0–2	žádná – nadprůměrná
		Varianta podpisu (na konci textu)	0–8	9 variant (např. celé křestní jméno a iniciála příjmení)

Každá ze škál byla podrobně popsána v manuálu, včetně příkladů, aby byla její interpretace pro posuzovatele co nejvíce srozumitelná. Ukázka konkrétního textu v manuálu je zachycena na obr. 3.2.3.1.A, ukázka elektronického rozhraní pro posuzovatele pak na obr. 3.2.3.1.B.

## Obr. 3.2.3.1.A – Ukázka definice škály v manuálu pro posuzovatele



Manuál ke škálování textů pro výzkum CPACT – Interní metodický materiál

# I. Dojem z osobnosti pisatele – bipolární škály

## 1 A01 Samotářský – družný

0 velmi samotářský      1 spíše samotářský      2 neutrální      3 spíše družný      4 velmi družný

**Charakteristiky samotářský – družný popisují touhu po mezilidských kontaktech, družnost, živost, obecnou společnost, extraverci.**

**Velmi samotářští** jsou spíše tišší, přemýšliví, uzavření, opatrní, raději zůstávají v pozadí. Dávají přednost stabilnímu a známému prostředí, preferují omezený počet blízkých a důvěrných vztahů, potřebují čas, než začnou důvěřovat a sbližují se. V davu se cítí nepříjemně. Dávají přednost samostatné práci.

**Velmi družní** jsou sebejistí, aktivní, komunikativní, hovorní, energičtí, veselí a optimističtí, rychle navazují nové kontakty a známosti. Preferují pobyt ve společnosti, jsou rádi středem pozornosti, mezi druhými lidmi se cítí jako ryba ve vodě, o samotě se záhy začínou cítit nepříjemně. Rádi se účastní diskuzí, vnášejí zájem o ostatní. Častou jsou iniciátory nápadů a hybnou silou ve skupině, rádi organizují, jsou asertivní, nemívají velké zábrany.

- 0 – věcně popisují bez zúčastněných osob, pouze vlastní názory.  
Př: „u sklenky piva na horách jsem zažíval pocity blaženosti, kolem mne jen ptáci“, „mám splněn kurz angličtiny ...“
- 2 – popis míst s osobami, ale bez popisu osob nebo bez vztahu k osobám  
Př: „na horách je krásně, včera jsem byl v hospůdce, pivo skvělé, jídlo také“, „Chodil jsem s kolegy na kurz angličtiny ...“
- 4 – popis kolegů, spolubydlících, důraz na počet osob, projevy komunikace s „cizími“ osobami ...  
Př: „na horách se máme skvěle, včera jsme byli v hospůdce, paní hostinská byla pěkná, pan hostinský byl broučoun, ale dal si s námi taky sklenku. Karel vzal kytaru a s Petrou jsme do rána zpívali“, „Na kurz angličtiny jsem chodil k lektorce Denise, což je roditelka mluvčí z Oxfordu, kde ji učil slavný profesor John W. . Byla fajn, stejně jako její kolegyně Marta a kolega Ondra“

Pozn.: Dotazníky měření rysů extraverte BFI.

## Obr. 3.2.3.1.B – Ukázka elektronického rozhraní pro škálování

**Hodnocení dopisů**

Hodnocení dopisu Přihášen jako: [avatar]

Seznam hodnocení

Otázka: 1

**Zvolte prosím tu odpověď, která nejvíce vystihuje Váš dojem z AUTORA TEXTU.**

Souhlasí?  
A01 Samotářský – družný

velmi samotářský    spíše samotářský    neutrální    spíše družný    velmi družný

H: Dále. Dobrý den, pojďte dál, já vás tady vítám.  
67P: Děkuju, dobrý den.  
H: Posadte se, udělejte si pohodlíčko a mohla bych vám nabídnout trochu brusinkovo-šipkového čaje?

Kompletní data z modulu P2 byla (z důvodu časové náročnosti celého procesu hodnocení) dostupná až na konci roku 2018. Výstupy z této části výzkumu proto budou součástí samostatné odborné studie mimo tuto monografii.

### 3.2.4 Diskuse ke koncepci výzkumu

---

Závěrečná část této kapitoly je věnována obecné diskusi k výzkumu CPACT, jeho modulům a metodologii. Další komentáře, diskuse a shrnutí pak naleznete v navazujících kapitolách, u jednotlivých studií.

#### 3.2.4.1 Soubor

Zajištění kvalitního (reprezentativního) výzkumného souboru je obvykle největší výzvou jakékoliv psychologické studie. Také v našem případě bylo toto téma zásadní koncepční otázkou. Zvolený soubor 200 osob (modul P200) by se mohl zdát jako relativně malý; je jen malou výsečí základní populace, ke které jsme výsledky vztahovali (cca 0,0019 % z počtu všech obyvatel ČR; viz ČSÚ, 2015), navíc je do určité míry spojený s lokalitou Jižních Čech, kde sběr dat probíhal (byť se ho zúčastnilo i mnoho osob z jiných krajů, např. studujících na univerzitě). Argumentem, který naopak zvolenou velikost odůvodňuje, byla náročnost extenzivního testování s řadou výzkumných procedur, které vyžadovaly velké množství času a měly navíc omezenou kapacitu (např. s ohledem na rozhovory a osoby asistentů–herců). Výraznější zvýšení počtu osob v souboru by tak bylo za daných podmínek a v rámci jednoho pracoviště jen obtížně realizovatelné. Modulu P200 se navíc účastnily pouze osoby ve dvojici, což byl další aspekt, který nábor potenciálních zájemců ztěžoval. Nižší počet účastníků tak do určité míry kompenzoval alespoň kvótní výběr, resp. stanovení kvót pro pohlaví, věk a vzdělání. Zmíněný model dvojic, byť trochu paradoxně, rovněž přinesl určité výhody v podobě časté situace, při které zájemce o výzkum k účasti přemluvil nějakou blízkou osobu, která by se jinak, samostatně, výzkumu neúčastnila (např. pod vlivem předsudků, neochoty či obav).

Klinický soubor P20+, zahrnující 72 osob, byl rovněž jen malou výsečí základní populace (cca 0,02 % z pacientů s afektivními a neurotickými poruchami; viz Nechanská et al., 2017). Jednalo se však o populaci, která je jen velmi obtížně dostupná a se kterou musí být spolupráce v mnoha směrech delikátní. Tento počet účastníků navíc postačoval pro smysluplnou aplikaci statistických testů.

Soubor dvou posuzovatelů v modulu P2 samozřejmě neměl ambice reprezentovat jakýkoliv základní soubor čtenářů. A zde by nepochybně bylo žádoucí získat i další posuzovatele, např. jiných osobních či sociálních charakteristik. Množství hodnocených textů (1200 textů) a s tím spojená finanční náročnost však neumožňovaly skupinu posuzovatelů rozšířit. Tento modul je tak spíše impulzem pro další studium problematiky zaslepeného, laického posuzování verbálního komunikátu za účelem odhadu osobnosti jeho autora.

#### 3.2.4.2 Procedury

Procedury v modulu P200 tvořily rozsáhlý komplex zahrnující jak zajištění účasti participantů, rezervaci a úpravu prostor, vyškolení a práci personálu, realizaci výzkumných sezení, tak neustálou kontrolu průběhu výzkumu a případné úpravy jeho harmonogramu. Z organizačního hlediska byl tento proces náročný, avšak nečekaně úspěšný. V průběhu výzkumu nenastal žádný výpadek ohlášeného termínu, v rámci výzkumných sezení nikdy nedošlo k předčasnému ukončení účasti ze strany participanta (nebo k podání jakékoliv stížnosti), podařilo se zajistit stabilní personální složení asistentického týmu (jak asistentů–herců, tak primárních asistentů) a výzkum dodržel všechna etická kritéria i stanovený harmonogram. Nevýhodou procedur tak zůstává pouze jejich délka, rozsah a náročnost. Výzkumné sezení se často protáhlo do pozdních odpoledních hodin, což bylo pro všechny zúčastněné velmi vysilující. To platilo zejména pro starší účastníky výzkumu a účastníky, kteří na výzkumné sezení dojížděli z jiných měst či krajů.

Procedury modulu P20+ sice nebyly tak rozsáhlé jako v P200, z hlediska práce asistentů však byly vysoce náročné. Tím, že byla setkání s účastníky a sběr dat nejčastěji realizovány v terénních podmínkách, vyžadovala jejich organizace značnou flexibilitu a trpělivost. Co se týká standardního nastavení sezení, decentralizace sběru pochopitelně může přinést otázku, zda všechny asistentky (i přes shodný deklarovaný postup) dodržovaly shodné postupy a přístup k účastníkům. Z hlediska plánu výzkumných bloků pak není zcela vhodné, že délka přestávky mezi oběma bloky poměrně výrazně variovala (od desítek minut po několik dní).

Procedury v modulu P2 zřejmě nevyžadují zásadnějších revizí. Výraznějším problémem snad byla pouze vysoká časová náročnost hodnocení textů, která kladla značné nároky jak na posuzovatele, tak na výzkumníka (z hlediska rychlosti, s jakou jsou výsledky k dispozici).

### 3.2.4.3 Materiály: Psychologické testy a výzkumné texty

Při posuzování testové baterie je zřejmé, že jedním z nejproblematičtějších parametrů je její rozsah. Únava účastníků byla v řadě případů vysoká, zejména pak v závěrečných blocích, což mohlo způsobit nižší míru soustředěnosti, a tím i určité zkreslení odpovědí. Potenciální únavu nepochybně podporovala i forma administrace, která byla v hlavním modulu P200 elektronická. Ta sice usnadňuje vyhodnocení testů, avšak dlouhé hledění do monitoru počítače jistě nebylo příliš komfortní. Situaci účastníkům alespoň ulehčovala možnost odpočinku v podobě přestávek.

Z hlediska volby psychologických testů zahrnovala naše baterie množství různých dotazníků, resp. psychologických škál (charakteristik), což souvisí s extenzivním a exploračním profilem výzkumu. V této souvislosti je možné diskutovat o relevanci konkrétních testů, jejich úpravě a zejména o výlučném využití právě dotazníků. Zvolená forma testování může být logicky předmětem úvah a revizí, přičemž doplnění baterie také o nedotazníkové (či jiné dotazníkové) metody by skutečně mohlo přinést nové a cenné výsledky. Z hlediska relevance současných testů však považujeme za významné, že většina dotazníků byla administrována také ve variantě O (posouzení druhou, blízkou osobou), což může pomoci výrazně redukovat tzv. *response bias* (Kaplan & Saccuzzo, 2017), tedy zkreslení v sebesposouzení.

Zaměříme-li se na volbu scénářů pro výzkumné texty (tj. registry) s cílem zajistit jejich variabilitu, dostáváme se do oblasti řady odborných polemik. Zásadní otázkou je zde to, zda je komunikační chování diskriminativní, nebo konzistentní napříč komunikačními situacemi (viz Shoda & Mischel, 1994). Zatímco někteří autoři jsou zastánci relativní konzistence a jejich výzkumy tuto domněnku více či méně podporují (Pennebaker & King, 1999), řada studií uvádí zcela jiné výsledky a jejich autoři zmiňují, že diferenciací textů a komunikačních registrů

(viz Biber, 1993) je pro jakoukoliv analýzu velmi významná. Vliv komunikačního kontextu na variabilitu projevů jedné osoby (tzv. *language style*) byl opakovaně popsán v rovině užívání jazyka jako takového (např. Chen & Bond, 2010), ale i při konkrétních manipulacích se slovy (např. Newman, Groom, Handelman, & Pennebaker, 2008; Ireland & Mehl, 2014). V rámci přípravy našeho výzkumu jsme se přiklonili k druhému pohledu a orientovali se na sběr různých typů komunikátů (textových registrů). Výsledky analýz pak prokázaly, že dané rozhodnutí bylo správné (viz kapitola 3.4.3).

Druhým důležitým aspektem při volbě scénářů (typu textových materiálů, registrů) byla ekologická validita a podobnost výsledných textů s reálnými komunikačními registry. Jak bylo dříve popsáno, veškeré analyzované komunikáty vznikly ve specifických výzkumných podmínkách (kontrolovaných a do určité míry „umělých“). Řada autorů se tomuto modelu brání a preferuje např. sběr on-line dat, případně analýzu materiálů, které primárně nevznikly pro potřeby výzkumu (např. Holtraves 2011; Yarkoni 2010). Po zhodnocení výsledků multidimenzionální analýzy našich textů (MDA, viz kap. 3.4.3) však můžeme potvrdit, že volba daných scénářů byla vzhledem k běžným (nevýzkumným, neelicitovaným) registrům relevantní a funkční. Zásadní nevýhodou tak zůstává především omezená délka získaných komunikátů v podobě psaných textů (dopisů), která znesnadňovala aplikaci lingvisticko-analytických postupů, např. s ohledem na augmentaci výskytu izolovaných textových parametrů, které by v delším textu zanikly.

### **3.2.5 Závěr**

---

Jak bylo zmíněno v kapitole 3.1, výzkum CPACT byl v kontextu českého prostředí unikátním projektem. Tato charakteristika přinášela jak benefity (např. potenciál nových objevů), tak zřejmá rizika. Ta souvisela kupříkladu s rozsáhlým komplexem metod a procedur, které byly v řadě případů vytvořeny „na míru“ projektovým týmem a které neměly vždy oporu v metodologii zahraničních studií. Při realizaci projektu se zároveň objevovalo mnoho problémů, které naznačovaly, že by bylo možné určité části výzkumu provést jiným způsobem, efektivněji

či metodologicky korektněji. Přes tato zjištění je však možné konstatovat, že projekt byl, alespoň z hlediska jeho evaluace projektovým týmem, realizován v souladu s plánem, proběhl úspěšně a přinesl mnoho žádoucích poznatků, které do této doby nebyly nikde publikovány.



## 3.3

# Charakteristika dat a práce s daty

## 3.3.1

# Psychometrické charakteristiky použitých testů

*Tomáš Urbánek, Dalibor Kučera, Jiří Haviger*

Kapitola Psychometrické charakteristiky použitých testů je věnována stručnému představení jednotlivých testů použitých v rámci výzkumu CPACT a jeho prvního modulu P200 (kvótní soubor neklinické populace). Je zaměřena na popis metod a konstruktů, které dané testy sledují. Uvedeny jsou jejich psychometrické vlastnosti – jak ty, které vycházejí z dokumentace původních autorů, tak ty, které byly získány na základě výzkumu CPACT, resp. vlastních analýz.

### 3.3.1.1 Administrace testů

---

Popisu administrace a procedur, které byly využity v rámci výzkumných sezení, se podrobně věnuje kapitola 3.2. Zde shrnujeme pouze nejdůležitější informace, které se vztahují k testové baterii P200.

Všechny testy byly v tomto modulu administrovány přímo ve výzkumných prostorách v rámci jednodenního sezení, pomocí připraveného elektronického rozhraní, které umožňovalo zřetelné zobrazení každé jednotlivé položky a zaznamenání odpovědi respondenta kliknutím myši. Systém účastníkům umožňoval návrat na předchozí položku, nikoliv však dále. Čas na odpověď nebyl omezen. Účastníci mohli vyplňování testů kdykoliv přerušit.

Testová baterie zahrnovala 361 položek (z 11 psychologických testů) zaměřených na sebezposouzení, tj. varianta S, a 321 položek (z 10 testů) zaměřených na posouzení druhé osoby, tj. varianta O. Ve variantě O posuzoval respondent, po dokončení vlastního sebehodnocení, také svého tzv. výzkumného partnera; tedy osobu, se kterou se na testování dostavil a která mu měla být dostatečně blízká a známá. Všechny testy byly ve variantě S i O přizpůsobeny pohlaví hodnoceného, tj. formulace

otázek a odpovědí vycházela z mužského, či ženského rodu, podle toho, byl-li popisovaným člověkem muž, nebo žena.

Administrace testové baterie byla rozdělena do 2 bloků, obvykle oddělených přestávkou (viz tab. 3.2.1.2). 1. blok testů respondenti vyplňovali dopoledne (varianta S) a 2. blok vyplňovali odpoledne (varianta O). Ve variantě S tedy vyplňovali tyto testy: 1. část (BFI, SMS, DASS1+2, STAIX2, SPSBS), 2. část (MMG, PSSI1+2, IAS), 3. část (SHQ, SSI, BOBR). Varianta O tento dopolední model kopírovala, avšak bez škál, které byly určeny pouze pro variantu S. Naopak byl do odpoledního bloku zařazen dotazník DBV.

*Tab. 3.3.1.1 – Seznam testových škál v baterii P200 s vyznačením škál administrovaných pouze ve variantě S (sebeposouzení)\* a škál, které byly administrovány v rámci druhého modulu P20+*

<b>Test**</b>	<b>Škála</b>	<b>Popis škály</b>	<b>S*</b>	<b>P20+</b>
BFI	E	Extraverze		X
BFI	N	Emoční labilita (negativní emoce)		X
BFI	O	Otevřenost vůči zkušenosti (citlivost na estetické podněty)		X
BFI	P	Přívětivost		X
BFI	S	Svědomitost		X
PSSI1	DS	Styl osobnosti: sebejistý – disociální		
PSSI1	SZ	Styl osobnosti: rezervovaný – schizoidní		
PSSI1	NT	Styl osobnosti: pečlivý – nutkavý		
PSSI1	HI	Styl osobnosti: příjemný – histriónský		
PSSI1	OB	Styl osobnosti: ochotný – obětující se		
PSSI2	DP	Styl osobnosti: klidný – depresivní	X	
STAIX2	ANX	Úzkostnost jako rys		
MMG	HA	Očekávání afiliace	X	X
MMG	FR	Strach z odmítnutí	X	X
MMG	HP	Očekávání moci	X	X
MMG	FP	Strach z moci	X	X
MMG	HS	Očekávání úspěchu	X	X

MMG	FF	Strach z neúspěchu	X	X
DASS1	D	Škála deprese		X
DASS1	S	Škála stresu		X
DASS2	A	Škála úzkosti	X	X
IAS	PA	Sebejistý–dominantní		
IAS	BC	Arogantní–vypočítavý		
IAS	DE	Chladný		
IAS	FG	Introvertovaný–rezervovaný		
IAS	HI	Nejistý–submisivní		
IAS	JK	Bezelstný–skromný		
IAS	LM	Vřelý–přívětivý		
IAS	NO	Společenský–přátelský		
SHQ	CD	Humor – kognitivní dimenze		X
SHQ	SD	Humor – sociální dimenze		X
SHQ	AD	Humor – afektivní dimenze		X
SHQ	SHQ	Humor – celkový		X
DBV	BV	Dotazník blízkosti vztahů ***	–	
SPSBS	IMPS	Bezpečnostní chování	X	X
SMS	AMSP	Schopnost měnit svou sebe-prezentaci		
SMS	SEBO	Přecitlivělost na výrazné chování druhých		
SMS	CSV	Schopnost přizpůsobit se různým situacím		
SMS	SMS	Self-monitoring (celkový skór)		
BOBR	SS	Spokojenost s vlastním vzhledem		
BOBR	AD	Žádoucnost vyšší atraktivity		
BOBR	ABI	Aktuální podoba vlastní postavy		
BOBR	DBI	Žádoucí postava		
BOBR	DOSI	Nejpřitažlivější postava opačného pohlaví		
SSI	EE	Sociální schopnosti – nonverbální vyjadřování		X
SSI	ES	Sociální schopnosti – nonverbální vnímavost		X

SSI	EC	Sociální schopnosti – nonverbální regulace (kontrola)	X
SSI	SE	Sociální schopnosti – verbální vyjadřování	X
SSI	SS	Sociální schopnosti – verbální vnímavost	X
SSI	SC	Sociální schopnosti – verbální regulace (kontrola)	X

\* *S* – testy pouze ve variantě pro sebeposouzení

\*\* *BFI* – *BFI-44*: Big Five Inventory (Pětifaktorový inventář osobnosti); *PSSI* – *PSSI*: Persönlichkeits-Stil-und Störungs-Inventar (Inventář stylů osobnosti a poruch osobnosti)(výběr škál); *STAI<sub>X</sub>2* – *STAI X-2*: State-Trait Anxiety Inventory (Dotazník úzkosti a úzkostlivosti); *MMG* – *MMG*: Multi-Motive Grid (Multimotivační mřížka); *DASS<sub>1+2</sub>* – *DASS 21*: Depression, Anxiety, Stress Scale (Škála deprese, úzkosti a stresu); *IAS* – *IAS*: Interpersonal Adjective Scales (Interpersonální adjektivní škála); *SHQ* – *SHQ-6*: Sense of Humor Questionnaire (Dotazník smyslu pro humor); *DBV* – *DBV* – Dotazník blízkosti vztahů; *SPSBS* – *SPSBS*: Social Phobia Safety Behaviors Scale (Škála sociální fobie a bezpečnostního chování); *SMS* – *SMS*: Self-Monitoring Scale (Škála sebemonitoringu); *BOBR* – *BOBR*: Basic Olomouc Body Rating (Škála fyzického vzhledu a „body image“); *SSI* – *SSI*: Social Skills Inventory (Inventář sociálních dovedností)  
\*\*\* *DBV* byl administrován pouze ve variantě posouzení druhé osoby (*O*)

### 3.3.1.2 Zkoumaný soubor a procedura

Zkoumaný soubor P200 zahrnoval  $n(m)=100$  mužů,  $n(f)=100$  žen. Věk účastníků byl zachycen ve čtyřech kvótách:  $n(<25)=26$ ,  $n(25-34)=34$ ,  $n(35-55)=67$ ,  $n(>65)=73$ . Rozložení vzhledem k aktuálně dosaženému vzdělání  $n(ZŠ)=36$ ,  $n(SŠ)=128$ ,  $n(VŠ)=36$  (viz kap. 3.2).

U tohoto souboru byla provedena analýza položek jednotlivých škál a její stručné výsledky jsou uvedeny na příslušných místech této kapitoly za popisy jednotlivých metod. Všechny analýzy, jejichž výsledky jsou uvedeny v této kapitole, byly provedeny v softwaru „R“ (R Core Team, 2018) a knihovně „Psych“ (Revelle, 2018).

### 3.3.1.3 Popis metod (testů)

---

V této části budou podrobněji popsány jednotlivé testy. Ty uvádíme v pořadí, které kopíruje tabulku Tab. 3.3.1.1.

#### **BFI-44: Big Five Inventory** (Pětifaktorový inventář osobnosti)

BFI je pětifaktorovým osobnostním inventářem, resp. dotazníkem měřícím pět obecných dimenzí osobnosti: extraverze (E), přívětivost (P), svědomitost (S), emoční labilita (negativní emoce; N) a otevřenost ke zkušenosti (O).

Autory testu BFI jsou John, Naumann a Soto (2008). Každou dimenzi zastupuje různý počet položek (E=9, N=8, O=10, P=9, S=9) v celkovém počtu 44. Položky mají charakter přídavných jmen používaných k popisu osobnosti. Na položky proband odpovídá výběrem z pětibodové škály typu Likertovy škály (*určitě ne, spíše ne, neutrální, spíše ano, určitě ano*) a vyjadřuje tak svou míru souhlasu s jednotlivými tvrzeními.

V dřívějších výzkumech bylo zjištěno, že BFI-44 má akceptovatelné psychometrické vlastnosti, reprezentované přiměřenou reliabilitou škál a odpovídající retestovou reliabilitou. Verze BFI-44 byla podrobena analýze na adolescentní a dospělé populaci. Reliabilita se pohybuje od 0,65 do 0,83, což lze považovat za uspokojivou míru vnitřní konzistence. Průměrná test-retest stabilita dimenzí BFI-44 po dvou měsících byla  $r = 0,79$  (Hřebíčková et al., 2016).

V rámci výzkumu bylo pracováno se všemi pěti škálami: BFI-E (extraverze), BFI-P (přívětivost), BFI-S (svědomitost), BFI-N (emoční labilita) a BFI-O (otevřenost ke zkušenosti). Test BFI-44 byl administrován v obou verzích, S i O. Česká jazyková varianta byla převzata ze studie Hřebíčkové et al. (2016).

V našem zkoumaném souboru jsme zjistili, že všechny subtesty BFI v obou verzích mají uspokojivou vnitřní konzistenci. V závorkách jsou uvedeny hodnoty Cronbachovy alfy pro jednotlivé subtesty a jejich verze – BFI-E-S (0,88), BFI-E-O (0,85), BFI-N-S (0,87), BFI-N-O (0,88), BFI-O-S (0,79), BFI-O-O (0,84), BFI-P-S (0,73), BFI-P-O (0,79),

BFI-S-S (0,82), BFI-S-O (0,81). V žádné ze škál BFI se nevyskytovaly položky, které by byly z hlediska vnitřní konsistence problematické (např. s nulovou nebo zápornou korelací se zbytkem položek ve škále).

Tab. 3.3.1.3.A – BFI-E: Ukázka položky a možností odpovědi

Varianta	S (sebeuposouzení), muž
Testové zadání	Posudte prosím, nakolik uvedená výpověď souhlasí – nakolik Vás vystihuje.
Položkové zadání	Považuji se za někoho, kdo...
Ukázka položky	je upovídaný.
	určitě NE
	spíše NE
Varianty odpovědi	neutrální
	spíše ANO
	určitě ANO

## PSSI: Persönlichkeits-Stil-und Störungs-Inventar

(Inventář stylů osobnosti a poruch osobnosti)

PSSI je inventářem stylů osobnosti a poruch osobnosti, který určuje relativní vyhraněnost osobnostních stylů, chápaných jako nepatologické varianty poruch osobnosti (viz např. vymezení v DSM-IIIR, DSM-IV, ICD-10). Autory dotazníku jsou J. Kuhl a M. Kazén (1997).

Dotazník obsahuje 140 položek ve 14 škálách. Škály mají dobrou vnitřní konsistenci (Cronbachův standardizovaný koeficient alfa 0,80; Kuhl & Kazén, 2002). Odpovědi respondentů na výroky jsou formulovány do čtyř variant: *ano*, *spíše ano*, *spíše ne*, *ne*.

V rámci výzkumu CPACT nebylo administrováno všech 14 škál, ale pouze 6 škál osobnostních stylů, které byly ve směru k zaměření projektu zásadní: PSS1-DS (sebejistý – disociální), PSS1-SZ (rezervovaný – schizoidní), PSS1-NT (pečlivý – nutkavý), PSS1-HI (příjemný – histriónský), PSS1-OB (ochotný – obětující se), PSS12-DP (klidný – depresivní styl). Mimo škály PSS12-DP byly všechny administrovány ve verzi S a O, škála PSS12-DP byla administrována pouze ve verzi S

(charakteristiky škály neumožňovaly změnu perspektivy na vnějšího hodnotitele). Česká jazyková varianta byla převzata z testového manuálu (Kuhl & Kazén, 2002).

V závorkách za zkratkami jednotlivých škál budou opět uvedeny hodnoty Cronbachovy alfy pro použité škály a jejich varianty – PSSI-DS-S (0,84), PSSI-DS-O (0,86), PSSI-HI-S (0,85), PSSI-HI-O (0,85), PSSI-NT-S (0,81), PSSI-NT-O (0,87), PSSI-OB-S (0,78), PSSI-OB-O (0,78), PSSI-SZ-S (0,79), PSSI-SZ-O (0,80), PSSI-DP-S (0,79). Všechny hodnoty jsou tedy uspokojivé, a hodnoty škál jsou proto vhodné k dalšímu statistickému zpracování.

*Tab. 3.3.1.3.B – PSSI1-SZ: Ukázka položky a možností odpovědi*

Varianta	O (posouzení druhé osoby), žena
Testové zadání	Pokuste se prosím posoudit, nakolik uvedená výpověď souhlasí – nakolik vystihuje Vašeho výzkumného partnera.
Položkové zadání	Souhlasí?
Ukázka položky	Důvěrnost v kontaktu s druhými lidmi je jí spíše nepříjemná.
Varianty odpovědí	určitě NE
	spíše NE
	spíše ANO
	určitě ANO

## **STAI X-2: State-Trait Anxiety Inventory**

(Dotazník úzkosti a úzkostlivosti)

Autory dotazníku STAI, skládajícího se ze dvou variant X-1 a X-2, jsou Spielberger et al. (1983). Ve variantě X-1 dotazník zjišťuje, jak se jedinec cítí v daný okamžik, zatímco varianta X-2 formuluje otázku do podoby obvyklých pocitů a sleduje tak úzkostnost jako rys osobnosti (*anxiousness*).

V rámci projektu CPACT byla zvolena pouze varianta STAI X-2 (úzkostnost), která obsahuje 20 položek. Respondent volí mezi čtyřmi va-



riantami odpovědi: *téměř nikdy, někdy, často, vždy*. Příručka autorů (Müllner, Ruisel, & Farkaš, 1980) obsahuje podrobné normy pro česko-slovenskou populaci; vzhledem k jejich stáří je však jejich praktická použitelnost diskutabilní.

Ve výzkumu je pracováno se škálou STAI<sub>X2</sub>-ANX, tedy s úzkostlivostí jako rysem osobnosti. Ta byla administrována v obou verzích, O i S. Česká jazyková varianta byla převzata z testového manuálu (Müllner, Ruisel, & Farkaš, 1980).

Cronbachova alfa u verze S byla 0,91 a u verze O pak 0,90, což jsou v případě dotazníkových dat velmi vysoké hodnoty svědčící jak o vysoké míře vnitřní konsistence dotazníkového nástroje, tak o poctivém přístupu respondentů při odpovídání na něj.

*Tab. 3.3.1.3.C – STAI<sub>X2</sub>-ANX: Ukázka položky a možností odpovědi*

Varianta	S (sebeuposouzení), žena
Testové zadání	Posuďte prosím, nakolik uvedená výpověď souhlasí – nakolik vystihuje Vaše OBVYKLÉ, běžné pocity.
Položkové zadání	Jak se obvykle cítíte?
Ukázka položky	Jsem klidná, pokojná a rozvázná.
Varianty odpovědí	téměř nikdy
	někdy
	často
	téměř vždy

### **MMG: Multi-Motive Grid** (Multimotivační mřížka)

Test Multimotivační mřížky (MMG) vytvořili Schmalt, Sokolowski a Langens (2010) jako psychodiagnostický nástroj měřící trojici základních implicitních motivů (motiv afiliace, motiv výkonu a motiv moci) vždy v souvislosti s očekáváním a strachem. S každým motivem se tedy pracuje dvakrát, tj. v rámci šesti škál. Metoda je konstruována jako kombinace projektivní techniky a tradičního dotazníku.

Jedinci je předloženo 14 obrázků, které zobrazují každodenní situace, avšak víceznačné, které jsou tematicky zaměřené ve směru ke sledovaným motivům. Ke každému obrázku je uvedeno 4–10 tvrzení, u kterých respondent volí odpověď (*ano, ne*), podle toho, zda se dle něj tvrzení k obrázku hodí. Doba administrace testu je cca 15 min. V roce 2010 byly v rámci testového manuálu zveřejněny normy na českou populaci.

Ve výzkumu jsou použity všechny škály, tj. MMG-HA (očekávání afilice, *hope of affiliation*), MMG-FR (strach z odmítnutí, *fear of rejection*), MMG-HP (očekávání moci, *hope of power*), MMG-FP (strach z moci, *fear of power*), MMG-HS (očekávání úspěchu, *hope of success*), MMG-FF (strach z neúspěchu, *fear of failure*). Celý test byl vzhledem ke svému charakteru administrován pouze ve verzi S. Česká jazyková varianta byla převzata z testového manuálu (Schmalt, Sokolowski, & Langens, 2010).

Přestože konstrukce metody MMG je poněkud odlišná od běžné dotazníkové metody, pro posouzení její reliability byl také zvolen model vnitřní konzistence, která (možná právě z důvodu odlišné konstrukce) vychází v některých případech poněkud nižší. Domníváme se však, že hodnoty Cronbachovy alfy uvedené dále v závorkách nejsou natolik nízké, aby skóry této metody diskvalifikovaly z dalších analýz – MMG-HA (0,58), MMG-FR (0,72), MMG-HP (0,76), MMG-FP (0,74), MMG-HS (0,68) a MMG-FF (0,68). Žádný výrok nebyl na základě analýzy položek vyhodnocen nekonzistentní s ostatními v rámci škál, takže se domníváme, že poněkud nižší vnitřní konzistence např. škály MMG-HA může být způsobena aperceptivní povahou škály.

Tab. 3.3.1.3.D – MMG-HS: Ukázka položky a možností odpovědi

Varianta	S (sebeuposouzení), muž
Testové zadání	Vžijte se do role libovolné osoby na obrázku a posuďte, jestli se uvedená tvrzení (podle Vás) k dané situaci HODÍ, či NEHODÍ.

Varianty odpovědí	Zde může člověk ztratit respekt.
	U těchto úloh myslí člověk na nedostatek schopností.
	Obává se moci jiných.
	V těchto situacích chce mít člověk vliv.

## **DASS 21: Depression, Anxiety, Stress Scale**

(Škála deprese, úzkosti a stresu)

DASS 21 je zkrácenou verzí dotazníku DASS (Lovibond & Lovibond, 1995), který obsahuje 42 položek zaměřených na tři základní negativní dimenze – depresi, úzkost a stres.

Škála deprese zjišťuje špatnou náladu, dysforii, beznaděj, sebedoceňování, setrvačnost či devalvací života. Škála úzkosti měří subjektivní pocit úzkosti, napětí a situační úzkost. Škála stresu zahrnuje netrpělivost, míru podrážděnosti a sklon k přemrštěným reakcím.

Dotazník obsahuje 21 tvrzení, respondent se ke každému z nich vyjadřuje zhodnocením četnosti, s jakou na něj tvrzení platilo v uplynulém týdnu, a to pomocí čtyřbodové škály: *vůbec ne*, *někdy*, *často*, *téměř vždy*.

Ve výzkumu CPACT byly škály DASS1-D (škála deprese) a DASS1-S (škála stresu) administrovány ve verzi S i O. Škála DASS2-A (škála úzkosti) pak byla administrována pouze ve verzi S. Česká jazyková varianta byla vytvořena překladem anglické varianty testu, resp. lehkou úpravou první české varianty, připravené ve spolupráci s Tomášem Vilimovským, autorem první české adaptace testu.

Hodnoty Cronbachovy alfy jsou uvedeny v závorkách za zkratkami jednotlivých subtestů – DASS1-D-S (0,87), DASS1-D-O (0,88), DASS1-S-S (0,83), DASS1-S-O (0,79), DASS2-A-S (0,84), a naznačují, že s daty z tohoto nástroje lze dále výzkumně pracovat.

Tab. 3.3.1.3.E – DASS1-S: Ukázka položky a možností odpovědi

Varianta	O (posouzení druhé osoby), muž
Testové zadání	Pokuste se prosím posoudit, do jaké míry uvedená tvrzení vystihují pocity a prožívání Vašeho výzkumného partnera V POSLEDNÍ DOBĚ, v posledních dnech.
Položkové zadání	Myslíte, že pro něj platilo toto tvrzení?
Ukázka položky	Bylo pro něj těžké se zklidnit.
Varianty odpovědí	vůbec ne
	někdy
	často
	téměř vždy

### IAS: Interpersonal Adjective Scales

(Interpersonální adjektivní škála)

Test IAS, vytvořený J. S. Wiggensem (1979), navazuje na cirkumplexní model interpersonálního chování známý např. z dotazníku Interpersonal Check List (Leary, 1957). Nástroj se zaměřuje na dvě dimenze interpersonálních vztahů: dominanci a starostlivost, přičemž Wiggins ve svém konceptu spojuje chování jedince v rámci sociálních interakcí se stálými charakteristikami osobnosti.

IAS-R je revizí původního testu IAS (Wiggins, 1979), jehož původní počet adjektiv (128) byl redukován na 64. Každá z osmi oblastí (škál) tak obsahuje 8 adjektiv, jejichž relevanci respondent posuzuje na osmi-bodové Likertově škále 1–8: *rozhodně ne, vcelku ne, častěji ne, spíše ne, spíše ano, častěji ano, vcelku ano, rozhodně ano* (v orig. *extremely inaccurate, very inaccurate, quite inaccurate, slightly inaccurate, slightly accurate, quite accurate, very accurate, extremely accurate*, viz níže).

Ve výzkumu CPACT bylo použito všech osm škál: IAS-PA (sebejistý – dominantní), IAS-BC (arogantní – vypočítavý), IAS-DE (chladný), IAS-FG (introvertovaný – rezervovaný), IAS-HI (nejistý – submisivní), IAS-JK (bezelstný – skromný), IAS-LM (vřelý – přívětivý), IAS-NO (společenský – přátelský). Škály byly administrovány v obou verzích, S i O. Česká jazyková varianta byla vytvořena odborným překladem položek z testového manuálu (Wiggins, 1995), při kterém byl finální text konsenzuálním výstupem dvou nezávislých překladatelů, při současném využití slovenské verze testu (Výrost, 1992) jako referenčního zdroje. Znění variant odpovědí (1–8) bylo upraveno s ohledem na standardní zadání dalších testů v baterii a s ohledem na srozumitelnost zadání, zjišťovanou v rámci pilotního testování.

V závorkách za zkratkami jednotlivých subtestů v obou verzích (S i O) jsou opět uvedeny hodnoty Cronbachova koeficientu alfa – IAS-PA-S (0,85), IAS-PA-O (0,82), IAS-BC-S (0,91), IAS-BC-O (0,90), IAS-DE-S (0,92), IAS-DE-O (0,92), IAS-FG-S (0,83), IAS-FG-O (0,84), IAS-HI-S (0,83), IAS-HI-O (0,79), IAS-JK-S (0,63), IAS-JK-O (0,80), IAS-LM-S (0,91), IAS-LM-O (0,93), IAS-NO-S (0,86), IAS-NO-O (0,88). U jediné škály, IAS-JK (bezelstný–skromný) ve variantě S, je hodnota Cronbachovy alfy pod hodnotou 0,7, která je tradičně považována za hranici dostatečné vnitřní konsistence. V případě výzkumné studie není nutné výsledky získané pomocí této škály opomíjet, jen je třeba mít potenciální nereliabilitu na paměti při interpretaci dalších výsledků založených na této škále.

*Tab. 3.3.1.3.F – IAS-DE: Ukázka položky a možností odpovědi*

Varianta	O (posouzení druhé osoby), muž
Testové zadání	Pokuste se prosím posoudit, do jaké míry uvedená tvrzení vystihují pocity a prožívání Vašeho výzkumného partnera V POSLEDNÍ DOBĚ, v posledních dnech.
Položkové zadání	Myslíte, že pro něj platilo toto tvrzení?
Ukázka položky	Bylo pro něj těžké se zklidnit.

	rozhodně NE
	vcelku NE
	častěji NE
Varianty odpovědí	spíše NE
	spíše ANO
	častěji ANO
	vcelku ANO
	rozhodně ANO

## SHQ-6: Sense of Humor Questionnaire

(Dotazník smyslu pro humor)

SHQ je dotazníkem zjišťujícím smysl pro humor, který vytvořil S. Svebak v roce 1996. V původní verzi obsahuje 27 položek, přičemž SHQ-6 je zkrácenou verzí o šesti položkách. Dotazník se zaměřuje na tři dimenze humoru: kognitivní, sociální a afektivní dimenzi. K uvedeným tvrzením vyjadřujeme míru souhlasu/nesouhlasu pomocí čtyřbodové škály: *vůbec ne, někdy, často, téměř vždy* (v orig. v různých variantách odpovědí, např. *not at all, to some degree, to a high degree, yes indeed*). SHQ-6 má souhrnně vnitřní konzistenci na základě Cronbachova koeficientu alfa 0,85 (Svebak, 1996).

Ve výzkumu CPACT byly využity všechny tři škály, SHQ-CD (humor – kognitivní dimenze), SHQ-SD (humor – sociální dimenze), SHQ-A (humor – afektivní dimenze), a souhrnná škála SHQ-SHQ (humor – celkový skór). Dotazník byl administrován v obou verzích, S i O. Česká jazyková varianta byla vytvořena odborným překladem položek z anglické verze dotazníku, při kterém byl finální výstup konsenzuálním výsledkem dvou nezávislých překladatelů. Znění variant odpovědí (1–4) bylo upraveno s ohledem na standardní zadání dalších testů v baterii a s ohledem na srozumitelnost zadání, zjišťovanou v rámci pilotního testování.

Protože každá ze subškál obsahuje pouze dvě položky, nepočítali jsme u nich Cronbachovy alfy, ale pouze Pearsonovy korelace, ze kterých si může čtenář udělat představu o tom, zda (a do jaké míry) měly

respondenti tendenci odpovídat na dvojice položek souhlasně. Pro SHQ-A-S je korelace rovna 0,32, pro SHQ-A-O je 0,35. Pro SHQ-C-S 0,34, pro SHQ-C-O také 0,34. Pro SHQ-S-S 0,45, pro SHQ-S-O 0,50. Celkový skóre SHQ ve verzi S má hodnotu Cronbachovy alfy 0,67 a ve verzi O hodnotu 0,72. Tyto hodnoty patří spíše k nižším, ale dají se vysvětlit tím, že škála je tvořena minimálním počtem položek, ve kterých se autor snažil o co nejrozmanitější obsah, který obecně neprospívá vnitřní konsistenci takto vytvořených škál, aniž by nutně musel znehodnotit reliabilitu. Při interpretaci dalších výsledků je tuto okolnost vhodné mít na paměti.

*Tab. 3.3.1.3.G – SHQ-SD: Ukázka položky a možností odpovědi*

Varianta	O (posouzení druhé osoby), žena
Testové zadání	Posuďte prosím, jak uvedená tvrzení vystihují Vašeho výzkumného partnera.
Položkové zadání	Souhlasí?
Ukázka položky	Myslíte si, že lidé, kteří se snaží být vždy vtipní, jsou nezodpovědní a nedá se na ně spolehnout.
Varianty odpovědi	vůbec ne
	někdy
	často
	téměř vždy

## **DBV: Dotazník blízkosti vztahů**

Pro potřeby výzkumu CPACT byl vyvinut krátký dotazník zjišťující míru blízkosti vztahů mezi účastníky výzkumu. Konstrukt blízkých osobních vztahů, který je v rámci dotazníků měřen, je zde definován jako individuálně významný vztah jedince k jemu nejbližším osobám, jehož prostřednictvím jsou osoby ovlivňované relativně často, silně, dlouhou dobu a v různých sférách činnosti (Kelley, 1982).

Konkrétní položky dotazníku vycházejí z východisek Finkela, Simpsona a Eastwicka (2017), kteří tento vztah definují unikátností, integrací (např. oslabení sebezpojetí a seberegulace ve směru k partnerovi)

a délkou společné trajektorie, která je považovaná také za zásadní ve směru k míře vzájemné obeznámenosti (např. Biesanz, West, & Millevoi, 2007).

Výsledkem syntézy uvedených konstruktů je devítipoložkový dotazník se čtyřmi variantami odpovědí (O1–O4) zjišťující délku vztahu, variabilitu situačních kontextů (např. průnik do soukromí), míru sdílení, vzájemnou identifikaci, míru saturace potřeb a míru závislosti. Dotazník byl administrován se zadáním „Pokuste se posoudit a popsat Váš vztah s výzkumným partnerem“.

Cronbachova alfa tohoto dotazníku má hodnotu 0,81 a korigované korelace jednotlivých položek se pohybují nad hodnotou 0,3, což svědčí o velmi uspokojivé úrovni vnitřní konzistence.

*Tab. 3.3.1.3.H – Položky DBV s variantami odpovědí (O1–O4)*

Č.	Položka – otázka	O1	O2	O3	O4
1	Jak dlouho trvá Váš blízký a osobní vztah s výzkumným partnerem? (uveďte prosím délku blízkého vztahu, nikoliv jen známosti)	méně než 2 roky	2 – 7 let	8 – 17 let	více než 17 let
2	Sdílíte se svým výzkumným partnerem společnou domácnost (bydlíte spolu)?	ne (ani v minulosti)	v minulosti krátce ano	v minulosti dlouhodobě ano	ano
3	Jak často jste v blízkém (osobním) kontaktu se svým výzkumným partnerem? (aktivity ve dvou, povídání si o osobních věcech apod.)	jednou za čas	alespoň jednou týdně	každý den	několikrát denně
4	Je mnoho aktivit, které děláte (sdílíte) pouze se svým výzkumným partnerem? (trávení volného času, společné projekty apod.)	spíše ne	občas takové jsou	je jich poměrně dost	naprostá většina je takových



5	Proniká Váš vztah s výzkumným partnerem do různých oblastí Vašeho života? (rodina, práce, zájmy apod.)	spíše ne	jak kdy	významně	zásadně
6	Ovlivňuje Váš výzkumný partner Vaše vlastní postoje a volby? (např. radou, doporučením či přáním)	spíše ne	jak kdy	významně	zásadně
7	Je možné popsat Váš vztah k výzkumnému partnerovi jako důvěrný či intimní? (svěřování se, sdílení pocitů apod.)	ani ne	jak kdy	spíše ano	naprosto
8	Naplnuje Váš výzkumný partner Vaše očekávání a potřeby?	ani ne	jak kdy	spíše ano	naprosto
9	Je pro Vás obtížné představit si život bez Vašeho výzkumného partnera?	ani ne	jak kdy	spíše ano	naprosto

### **SPSBS: Social Phobia Safety Behaviors Scale**

(Škála sociální fobie a bezpečnostního chování)

Dotazník SPSBS měří míru tzv. bezpečnostního chování (tzn. jakým způsobem člověk reaguje v sociálně náročných situacích, např. v konfliktu, přijímacím pohovoru, při vystupování před větším počtem lidí atp.). Autory testu jsou Pinto-Gouveia, Cunho a Salvador (2003). Dle autorů testu mají lidé se sociální fobií tendenci se v těchto situacích zachovat tak, aby zabránili vzniku nepříjemných pocitů, přičemž tuto tendenci je možné zachytit pomocí strukturovaného dotazování. K testu jsou k dispozici normy pro portugalskou (Cronbach alfa 0,92; Burato, Crippa, & Loureiro, 2009), kanadskou (Cronbach alfa 0,82; Kocovski et al., 2016) a brazilskou (Cronbach alfa 0,95; Pinto-Gouveia, Cunho, & Salvador, 2003) populaci.

Respondent má k dispozici 17 tvrzení, ke kterým volí odpověď (1–4) vyjadřující, jak často se tímto způsobem chová (*téměř nikdy, někdy, často, téměř vždy*). Výsledkem vyhodnocení je jeden souhrnný skór.

Škála SPSBS-IMPS (bezpečnostní chování) byla administrována pouze ve verzi S. Česká jazyková varianta byla vytvořena odborným překladem položek z kanadské studie Kocovski et al. (2016), při kterém byl finální text konsenzuálním výstupem dvou nezávislých překladatelů, při současném využití slovenské verze testu (Výrost, n.d.) jako referenčního zdroje.

Cronbachova alfa dotazníku SPSBS v našich datech je 0,72. Tyto hodnoty snižují nízké korelace některých položek se zbytkem škály, především položek SPSBS\_IMPS\_122 (-0,02; „Než něco řeknu, tak si to promyslím“) a SPSBS\_IMPS\_119 (0,03; „Kontroluji, jestli vypadám dobře“).

*Tab. 3.3.1.3.I – SPSBS-IMPS: Ukázka položky a možností odpovědi*

Varianta	S (sebeuposouzení), žena
Testové zadání	Je mnoho situací, které v lidech vyvolávají NAPĚTÍ a OBAVY – např. přijímací pohovor, vystupování na veřejnosti nebo setkání s osobou, kterou nemůžete vystát. Posuďte prosím, jak se právě Vy obvykle chováte v těchto situacích.
Položkové zadání	Jak často se v náročných situacích chováte tímto způsobem?
Ukázka položky	Mluvím rychleji než obvykle.
Varianty odpovědí	téměř nikdy
	někdy
	často
	téměř vždy

### **SMS: Self–Monitoring Scale** (Škála sebemonitoringu)

Autorem škály sebemonitoringu (SMS) je M. Snyder (1974). Test je zaměřen na sledování vůle a schopností jedince měnit způsob své prezentace pomocí zhodnocení míry identifikace jedince s předloženým tvrzením v dotazníku.

Revize testu (Lennox, & Wolfe, 1984) využila místo tří původních škál (schopnost měnit svou sebe-prezentaci, přecitlivělost na výrazné chování druhých a schopnost přizpůsobit se různým situacím) pouze první dvě zmiňované škály, tj. z původních 25 položek byl počet redukován na 13 (koeficient alfa pro první škálu byl 0,77, resp. pro druhou škálu 0,70). Odpovědi respondent zaznamenává pomocí volby ze šesti výroků (0–6): *určitě ne, většinou ne, spíše ne, spíše ano, většinou ano, určitě ano*.

Ve výzkumu CPACT byla na základě výše zmíněné studie provedena redukce počtu položek originální verze a k těmto 13 položkám doplněna poslední škála (CSV; viz Lennox & Wolfe, 1984, s. 1362). Test tak obsahuje celkem 20 položek, sytící škály SMS-AMSP (schopnost měnit svou sebe-prezentaci, *ability to modify self-presentation*; 7 položek), SMS-SEBO (přecitlivělost na výrazné chování druhých, *sensitivity to expressive behavior of others*; 7 položek), SMS-CSV (schopnost přizpůsobit se různým situacím, *cross-situational variability*; 6 položek) a celkový skór SMS-SMS. Je třeba zmínit, že původní verze dotazníku pracovala pouze s bipolární odpovědí respondentů (*pravda/nepravda*), pro potřeby výzkumu CPACT byl však rozsah možných odpovědí zvýšen na 6. Test byl administrován ve verzi S i O. Česká jazyková varianta byla vytvořena odborným překladem položek ze studie Lennox & Wolfe (1984), při kterém byl finální text konsenzuálním výstupem dvou nezávislých překladatelů.

V závorkách za zkratkami jsou uvedeny hodnoty Cronbachovy alfy jednotlivých škál v obou verzích (S a O) s případnými podrobnostmi: SMS-AMSP-S (0,75), SMS-AMSP-O (0,68), SMS-CSV-S (0,89), SMS-CSV-O (0,85), SMS-SEBO-S (0,78), SMS-SEBO-O (0,86), a celkové skóry SMS-S (0,82) a SMS-O (0,74). Vnitřní konzistence škály SMS-AMSP-O a celkového skóru SMS-O jsou snižovány nízkou korigovanou korelací s celkovou škálou položky SMS\_AMSP\_62\_O („Přestože by to pro něj bylo výhodné, neumí předstírat, že se nic neděje“); –0,06 v případě SMS-AMSP-O a –0,11 v případě SMS-O.

Tab. 3.3.1.3.J – SMS-AMSP: Ukázka položky a možností odpovědi

Varianta	O (posouzení druhé osoby), žena
Testové zadání	Pokuste se prosím posoudit, jak uvedená tvrzení vystihují Vašeho výzkumného partnera.
Položkové zadání	Souhlasí?
Ukázka položky	Dělá jí problém změnit své chování tak, aby se přizpůsobila různým lidem a situacím.
Varianty odpovědi	určitě NE
	většinou NE
	spíše NE
	spíše ANO
	většinou ANO
	určitě ANO

### **BOBR: Basic Olomouc Body Rating**

(Škála fyzického vzhledu a „body image“)

Test BOBR je založen na prezentaci vizuální škály 10 mužských a 10 ženských postav (od velmi štíhlých po obézní figury). Respondenti volí postavu, se kterou se identifikují, na základě svého aktuálního vzhledu a následně na základě pro ně žádoucího (subjektivně desirabilního) vzhledu. Výsledky prezentované studie autorů (Šrámková & Cakirpaloglu, 2015) poukazují na signifikantní rozdíl mezi výsledky obou variant hodnocení jak u mužů, tak u žen.

Ve výzkumu CPACT došlo k modifikaci tohoto nástroje, resp. k rozšíření zadání o položku „Klikněte na tu mužskou/ženskou figuru, která je pro Vás osobně nejpřitažlivější, nejatraktivnější.“ a k doplnění dvou položek, které nejsou spojeny s obrazy postav („Jsem spojený s tím, jak vypadám“ a „Rád bych byl atraktivnější“), ovšem poskytují k vizuálnímu testu dodatečné informace. Na tyto dvě položky jsou k dispozici odpovědi (0–3): *určitě ne, spíše ne, spíše ano, určitě ano.*

Upravený test BOBR tedy pracuje s těmito pěti škálami: dotazníkovými BOBR-SS (spokojenost s vlastním vzhledem, *self satisfaction*) a BOBR-AD (žádoucnost vyšší atraktivity, *attractiveness desirability*)

a třemi vizuálními škálami BOBR-ABI (aktuální podoba vlastní postavy, *actual body image*), BOBR-DBI (žádoucí postava, *desired body image*) a BOBR-DOSI (nejpřitažlivější postava opačného pohlaví, *desired opposite sex image*). Test byl administrován ve verzi S i O.

Výstupy z testu BOBR neumožňují odhad vnitřní konsistence.

*Tab. 3.3.1.3.K – BOBR-SS: Ukázka položky a možností odpovědi*

Varianta	S (sebeuposouzení), muž
Testové zadání	Posuďte prosím, nakolik uvedené tvrzení vyjadřuje Váš názor.
Ukázka položky	Jsem spojený s tím, jak vypadám.
Varianty odpovědí	určitě NE
	spíše NE
	spíše ANO
	určitě ANO.

*Tab. 3.3.1.3.L – BOBR-DBI: Ukázka položky a možností odpovědi*

Varianta	S (sebeuposouzení), muž
Testové zadání	Posuďte prosím, nakolik uvedené tvrzení vyjadřuje Váš názor.
Ukázka položky	Jsem spojený s tím, jak vypadám.
Varianty odpovědí	Obrázky postav 1–10

### **SSI: Social Skills Inventory** (Inventář sociálních dovedností)

Test SSI obsahuje šest škál, měřící komunikační dovednosti na dvou úrovních – emoční (neverbální) a sociální (verbální). Vyjadřování (*expressivity*), vnímavost (*sensitivity*) a kontrola (*control*) jsou hodnoceny na obou úrovních: vyjadřování zahrnuje schopnost jedince sdělovat informace druhým osobám, vnímavost se vztahuje ke schopnosti přijímat a interpretovat informace od druhých a kontrola vyjadřuje schopnost regulovat a řídit komunikační procesy.

Autory testu jsou R. E. Riggio a D. R. Carneyová (2003), českou verzi testu připravila roku 2007 Lenka Krejčová.

Inventář obsahuje 90 položek, respondenti odpovídají na pětibodové stupnici (1–5), do jaké míry koresponduje popsaná situace s jejich charakteristikami (psychologickými konstrukty či chováním): *určitě ne, spíše ne, spíše ano, většinou ano, určitě ano*. Cronbachova alfa pro jednotlivé škály SSI dosahuje výsledků od 0,65 do 0,88 (Riggio & Carney, 2003).

Ve výzkumu CPACT bylo využito všech 6 škál: SSI-EE (neverbální vyjadřování), SSI-ES (neverbální vnímavost), SSI-EC (neverbální kontrola), SSI-SE (verbální vyjadřování), SSI-SS (verbální vnímavost), SSI-SC (verbální kontrola). Test byl administrován ve verzi S i O.

V závorkách za zkratkami jednotlivých subškál jsou uvedeny hodnoty Cronbachovy alfy: SSI-EE-S (0,69), SSI-EE-O (0,69), SSI-ES-S (0,79), SSI-ES-O (0,82), SSI-EC-S (0,74), SSI-EC-O (0,66), SSI-SE-S (0,86), SSI-SE-O (0,88), SSI-SS-S (0,79), SSI-SS-O (0,80), SSI-SC-S (0,85), SSI-SC-O (0,84).

*Tab. 3.3.1.3.M – SSI-EE: Ukázka položky a možností odpovědi*

Varianta	S (sebeposouzení), žena
Testové zadání	Posuďte prosím, nakolik uvedená výpověď souhlasí – nakolik Vás vystihuje nebo nakolik vyjadřuje Váš názor.
Položkové zadání	Souhlasí?
Ukázka položky	Druzí lidé na mně jen těžko poznají, že jsem smutná nebo skleslá.
Varianty odpovědí	určitě NE
	spíše NE
	spíše ANO
	většinou ANO
	určitě ANO

### 3.3.1.4 Závěr

---

Závěrem je možné konstatovat, že analýzy položek jednotlivých škál neodhalily žádný závažný nedostatek, který by naznačoval, že by některá ze škál neměla být použita v dalších analýzách. Pokud se v ojedinělých případech vyskytla nižší hodnota Cronbachovy alfy, je možné toto zjištění zohlednit při interpretaci výsledků analýz – nižší reliabilita totiž může oslabit zjištěné korelace nebo rozdíly mezi skupinami. Celkově je ale testová baterie funkční a jednotlivé testy poskytují akceptovatelné psychometrické podklady.

## 3.3.2

# Lingvistická analýza textů

*Tomáš Jelínek*

Cílem psycholingvistické analýzy popisované v této knize je najít souvislosti mezi výsledky lingvistického rozboru textů, jež vytvořili respondenti za kontrolovaných podmínek, a výsledky psychometrických testů provedených na týchž respondentech. V této kapitole bude popsáno, jaké parametry se v textech zjišťují a jakými postupy. Stručně budou také srovnány dva vybrané textové parametry u zkoumaných textů (souhrnně) se stejnými parametry jiných českých textů (referenčních korpusů Českého národního korpusu).

V textech, které respondenti vytvořili na zadaná témata, hledá lingvistická analýza dílčí parametry především morfosyntaktické, tedy gramatické kategorie jako slovní druh jednotlivých slov, osobu a číslo sloves, pád substantiv aj., a lexikální, tedy příslušnost určitého slova do té či oné lexikální skupiny (modální slovesa, emotivně zabarvená slova, vulgarismy aj.). Pro každý text se pak vytváří celková statistika výskytů slov, které mají ten či onen parametr. Vzhledem k rozsahu textů se všechny tyto práce provádějí automaticky, počítačovými programy vytvořenými pro tento účel.

### 3.3.2.1 Sledované textové parametry

---

Lingvistická analýza textů sleduje u každého slova řadu morfosyntaktických, lexikálních a sémantických parametrů, dále sleduje v textech řadu parametrů týkajících se vět, popř. celého textu. Každý parametr může nabývat různých hodnot, například slovo může mít deset různých slovních druhů, pro neznámá slova je vyhrazen ještě další „slovní druh“ a pro interpunkci a jiné podobné znaky další. Základní morfosyntaktické parametry vycházejí z tzv. pražské morfologické analýzy (Hajič 2001), která každému slovu v textu přiřazuje morfologickou značku o 15 pozicích, resp. 16 pozicích v ČNK (což užíváme i zde).



Každému slovu je také přiřazen jeho základní (slovníkový) tvar: lemma; na základě tohoto lemmatu mohou být slovu přiřazeny další proměnné, rozpoznávají se například modální slovesa, slova s emočním nábojem aj.

Při zpracování textů se dopočítávají i některé parametry, které vycházejí z jiných již zjištěných parametrů. Zčásti jsou to jednoduché statistické parametry (jako celkový počet slov ve větě), zčásti se zjištěné hodnoty kombinují do složitějších parametrů, u nichž se předpokládá, že budou korelovat s parametry psychometrickými.

#### 3.3.2.1.1 Základní statistické kategorie

---

Pro každý text se zjišťují jednoduché statistiky: počet slov, počet vět (přesněji počet souvětí, ukončených samostatných vět a samostatných nevětých výpovědí), počet tokenů (slov, interpunkčních znaků, číslicemi psaných čísel aj.) dohromady.

#### 3.3.2.1.2 Gramatické kategorie

---

Gramatické kategorie vycházejí z jednotlivých pozic morfologických značek, které se zde používají. Jsou to tyto kategorie: slovní druh, podtyp slovního druhu, jmenný rod, číslo, pád, přivlastňovací rod a přivlastňovací číslo, osoba, slovesný čas, stupeň, negace, slovesný rod (aktivum/pasivum), stylová varianta a vid.

Možné hodnoty kategorie **slovní druh** (POS) jsou: podstatné jméno (N), přídavné jméno (A), zájmeno (P), číslovka (C), sloveso (V), příslovce (D), předložka (R), spojka (J), částice (T), citoslovce (I), interpunkce (Z) a neznámé nebo cizí slovo, popř. překlep, chyba (X).

Hodnot kategorie **podtyp slovního druhu** je mnoho (cca 70), u jednotlivých slovních druhů rozlišuje jejich podtypy (např. osobní, přivlastňovací nebo zvrtná zájmena), u sloves rozlišuje slovesný způsob (infinitiv, prézens, minulé přičestí aj.).

Jmenný **rod** může nabýt pěti různých hodnot: rod ženský (F), střední (N), mužský životný (M), mužský neživotný (I) a žádný rod (-) v případě, že nemá smysl jmenný rod určovat.

Kategorie **číslo** má tyto hodnoty: singulár (S), plurál (P), duál (D) a neurčuje se (-).

Hodnoty kategorie **pád** mohou být: nominativ (1), genitiv (2), dativ (3), akuzativ (4), vokativ (5), lokál (6), instrumentál (7) a žádný pád (-) v případě, že pád nemá smysl určovat.

Kategorie **přivlastňovací rod a číslo** jsou relevantní pouze pro přivlastňovací zájmena (*můj, její*) a vztažná přivlastňovací zájmena (*jehož, jejíž*), nabývají stejných hodnot jako jmenný rod a číslo.

Kategorie **osoba** je relevantní pro většinu slovesných tvarů a pro část zájmen. Má hodnoty: první (1), druhá (2), třetí osoba (3) a neurčuje se (-).

Kategorie **čas** se určuje pouze u sloves. Možné hodnoty jsou: minulý čas (R), přezens (P), futurum (F) a neurčuje se (-).

**Stupeň** se týká těch adjektiv a adverbíí, jež mohou být stupňovány, rozlišuje se pozitiv (1), komparativ (2), superlativ (3) a nemá smysl určovat (-).

Hodnoty kategorie **negace** se zjišťují u flektivních slovních druhů (substantiv, adjektiv, sloves, adverbíí, výjimečně u zájmen), určuje se pozitiv (A) a negace (N). U ostatních je hodnota této kategorie opět prázdná (-). U substantiv se však nerozpoznává, zda jde o negovaná substantiva s pozitivním tvarem, nebo o substantiva, která již jako negovaná nejsou pocíťována (*nezávislost, nesmysl*). Substantiva s prefixem *ne-* mají vždy negativní základní tvar a hodnotu tohoto atributu pozitivní (A).

**Slovesný rod** rozlišuje u sloves pasivum (P), ostatní slovesné tvary se označují jako aktivum (A). U jiných slovních druhů se kategorie neurčuje (-).

**Stylová varianta** rozlišuje standardní tvary (-,1), tvary knižní a archaické (2,3), tvary mluvené a obecněčeské (5,6,7), zkratky (8) aj.

**Vid** se označuje jen u sloves, určují se slovesa dokonavá (P), nedokonavá (I) a obouvidá (B). U ostatních slovních druhů se vid neurčuje (-).

### 3.3.2.1.3 Lexikální kategorie

---

Podle automaticky rozpoznávaného základního tvaru se u slov určuje i řada lexikálních kategorií. Zjišťuje se, zda slovesa užitá v textu jsou modální. Všem slovům se přiřazuje informace o jejich frekvenčním ranku. Identifikují se zdrobněliny, hovorová a vulgární slova, slova emočně nabitá.

Lexikální kategorie **modálních slov** rozlišuje jednak vlastní modální slovesa (M): *muset, moci, chtít, smět*, dále široce modální slovesa (m), např. *rozkázat, přinutit, žádat, hodlat* a substantiva s obdobným významem (n), např. *příkaz, zákaz, nařízení*.

Frekvenční **rank** rozlišuje u každého slova, nakolik je časté v referenčním korpusu ČNK SYN2015 (Křen et al., 2015). Uvádějí se číselné hodnoty od 1 do 6, přičemž číslo vyjadřuje mocninu 10 ( $10^x$ ), do jak rozsáhlé skupiny nejčastějších slov v referenčním korpusu dané slovo patří (slovo s rankem 3 patří mezi  $10^3$ , tj. 1000 nejčastějších slov). Číslo 6 označuje všechna slova méně častá než 100 000 nejčastějších slov. U interpunkce a neznámých slov se hodnota neurčuje (-).

**Deminutiva**, tj. zdrobněliny (D), **vulgární slova** (V) a **hovorová slova** (C) se identifikují ve třech nezávislých lexikálních kategoriích podle seznamů vytvořených na základě informací ve slovníku pražské morfologické analýzy (Hajič, 2001), tyto seznamy byly následně manuálně korigovány. Ostatní slova mají hodnotu neurčenou (-).

U **emočně nabitých slov** se rozlišují pozitivní emoce, např. *skvělý* (+), negativní, např. *ohavný* (-) a neutrální nebo bez porozumění věty neurčitelné (\*), například slovo *strašně* může být v některých kontextech pozitivní (velká míra), v jiných negativní. U slov, kde se takový náboj nezjistí, je prázdná hodnota ( ). Používají se dva seznamy (a tedy dvě různé kategorie, pokaždé se čtyřmi možnými hodnotami): jeden širší, převzatý beze změn od autorů českého slovníku subjektí-

vity, využívaného pro automatickou postojovou analýzu (Veselovská & Bojar, 2013), druhý redukovaný, manuálně ověřený.

#### 3.3.2.1.4 Frazémy

---

Součástí automatické lingvistické analýzy je i zjištění, zda slova ve zkoumaném textu tvoří víceslovné lexikální jednotky (rčení, ustálená přirovnání, přísloví aj). Identifikace těchto „frazémů“ byla zařazena na základě předpokladu, že výpovědi využívající ve větší míře ustálených výrazů mohou nějak korelovat s psychickým stavem mluvčího. Ukazuje se však, že texty jsou příliš krátké a výskyt frazémů příliš řídký na to, aby bylo možné z tohoto parametru dělat jakékoliv závěry. Kromě obecnějšího třídění frazémů bylo zařazeno i rozpoznání frazémů s emočním nábojem, u kterých však bylo ještě méně výskytů.

#### 3.3.2.1.5 Kombinované parametry

---

Kromě morfosyntaktických a lexikálních parametrů, které se určují u každého slova jednotlivě, se z dostupných údajů dopočítávají pro každý text další parametry, které celkově charakterizují text. Některé se zaměřují na počet slov splňujících určité podmínky, jiné ověřují souviskyty ve větách aj.

**Počet slov na větu** vyjadřuje, jak dlouhé věty respondent v průměru používá. Je vypočten jako prostý podíl slov ve větě a počtu vět a nevětných uzavřených výpovědí.

**Index lexikálního bohatství** textu ukazuje, jak rozmanitý slovník respondent v textu užil. Počítá se jako počet různých základních tvarů děleno druhou odmocninou celkového počtu slovních tvarů v textu (druhá odmocnina snižuje vliv délky textu na výslednou hodnotu indexu).

**Počet finitních sloves** na větu odpovídá složitosti větných konstrukcí, tedy nakolik respondent používá vedlejší věty či koordinaci v rámci souvětí. Vypočteno jako podíl sloves v určitém tvaru na celkovém počtu vět.

**Počet interpunkčních znamének na větu a počet vykřičníků na větu** se počítá opět jako počet výskytů na počet vět, první parametr by měl odrážet složitost větných konstrukcí, druhý tendenci respondenta vyjadřovat se emocionálně, ovšem výskyt vykřičníků ve sledovaných textech je relativně nízký.

Sleduje se také, nakolik je časté základní, nepříznakové pořadí adjektivní přívlastek – substantivum ve srovnání s příznakovým pořadím substantivum – adjektivní přívlastek. Tento vztah je vyjádřen parametrem **počet preponovaných adjektivních přívlastků** na celkový počet adjektivních přívlastků.

Zjišťuje se **míra užití spojky na prvním místě** ve větě, které vypovídá o strukturaci výpovědi, případně i o expresivnějším vyjadřování mluvčího.

Zjišťuje se **počet vulgárních slov a počet nestandardních**, mluvených **slov** na počet vět.

Dopočítávají se kombinace některých lingvistických parametrů: **kombinace osoby a čísla u jednoho slova** (1S je 1. os. sg., podobně 1P, 2S, 2P, 3S, 3P). U **osobních zájmen** je sledován výskyt **osoby a kombinace osoby a čísla** (PO1S je 1. os. sg. u osobních zájmen, tedy *já, mě, mně, mi, mnou...*, podobně PO1, PO1P, PO2, PO2S, PO2P, PO3). Podobně se zjišťuje výskyt **přivlastňovacích zájmen v kombinaci s osobou a číslem vlastníka** (PP2P je 2. os. pl. vlastníka, podobně PP1, PP1S, PP1P, PP2S, PP3) a **kombinace osoby a čísla u sloves** v určitém tvaru (V1S, V1P, V2). Zaznamenává se **frekvence negovaných přídavných jmen (ANeg), příslovcí (DNeg) a sloves (VNeg)**. U sloves se pak navíc evidují výskyt **první a druhé osoby v negaci (VNeg2)**, v případě první osoby se ještě rozlišuje číslo (VNeg1S, VNeg1P).

### 3.3.2.1.6 Lingvistické parametry textů – závěr

---

Výsledkem lingvistické analýzy je pro každý text každého respondenta 181 možných kombinací parametrů a hodnot s vypočtenými číselnými hodnotami (frekvencemi výskytů a podobně). Korelace mezi těmito lingvistickými parametry a psychometrickými testy pak zkoumají další odborníci.

### 3.3.2.2 Automatická lingvistická analýza textů

---

V této části kapitoly bude popsán postup automatické lingvistické analýzy textů, jímž bylo dosaženo výše představených výsledků, od předzpracování textů po shrnutí výsledků do souhrnných statistických tabulek.

Při práci s flektivním jazykem, jako je čeština, se počítačová lingvistika musí mimo jiné vyrovnávat s homonymií (víceznačností) slovních tvarů, a to jak s homonymií náhodnou (tvar *ženou* může být podle kontextu buď tvarem podstatného jména *žena*, nebo tvarem slovesa *hnát*), tak s homonymií systematickou (dativ a lokál jednotného čísla mužských vlastních jmen na *-ovi*, např. *poslal dopis Karlovi*, se systematicky shoduje s nominativem množného čísla mužského rodu přivlastňovacích přídavných jmen, např. *Karlovi přátelé*). Hlavním úkolem automatické lingvistické analýzy je tak odstranění této víceznačnosti a určení základního, slovníkového tvaru každého víceznačného slova a jeho gramatických kategorií. Popis automatické analýzy je zjednodušený, podrobnější popis automatické lemmatizace a morfologického značkování viz Jelínek & Petkevič, 2011.

#### 3.3.2.2.1 Předzpracování textů

---

Prvním krokem v lingvistické analýze je automatické předzpracování textů. Texty se exportují z databáze do jednotlivých souborů, jsou automaticky označeny názvem podle respondenta a typu textu. V textech se pak nahrazují některé kombinace znaků, například tzv. „smajlíky“, :-), u nichž není žádoucí, aby s nimi systém morfologického značkování

zacházel jako s běžnými interpunkčními znaky. Výsledkem předzpracování jsou krátké textové soubory, které jsou standardním vstupem pro vlastní morfologické značkování.

### 3.3.2.2.2 Segmentace, tokenizace a morfologická analýza

Ve druhé fázi projdou texty tzv. segmentací, tokenizací a morfologickou analýzou. Segmentace rozdělí text na jednotlivé věty, přičemž se opírá o velká a malá písmena, interpunkci, vyznačení konců odstavců aj. Tokenizace rozčlení větu na jednotlivá slova: většinou je slovo vymezené mezerami, někdy je třeba oddělit od slov interpunkční znaménka. Následuje morfologická analýza, která ke každému slovu doplní všechny možné interpretace tohoto slova, jak je například vidět na následujícím úryvku z jednoho z analyzovaných textů: *Myslím, že je tady dostatečný pohodlí.* V tabulce 3.3.2.2.2 jsou u každého slova uvedeny jeho možné základní tvary (lemmata) a možné gramatické kategorie, přičemž jsou zde vypsány kategorie jen slovně, ve skutečném zpracování je u každého možného základního tvaru uvedena také jedna nebo více morfologických značek.

Tab. 3.3.2.2.2 – Lemmata a možné gramatické kategorie

Myslím	<i>mysl</i> substantivum: dativ pl.	<i>myslet</i> sloveso, 1.os.sg.
,	, interpunkce	
že	<i>že</i> spojka	
je	<i>být</i> sloveso: 3.os.sg.	<i>on</i> zájmeno: akuzativ neut.sg. nebo pl.
tady	<i>tady</i> adverbium	
dostatečný	<i>dostatečný</i> adjektivum: nom./ak. sg. mask., v mluvené češtině další pády	<i>dostatečná</i> substantivum: v mluv. češtině, nom./ak.pl. nebo gen. sg.
pohodlí	<i>pohodlí</i> substantivum: mnoho kombinací možného pádu a čísla	
.	. interpunkce	

Morfologická analýza a lemmatizace využívá morfologické značky a slovníky vytvořené na Ústavu formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy (Hajič, 2001). Tyto morfologické značky jsou poziční, každá pozice vyjadřuje jinou morfologickou kategorii (slovní druh, podtyp slovního druhu, jmenný rod, číslo, pád atd.), například tvar *tebou* osobního zájmena *ty* má značku PP-S7--2----- (zájmeno, osobní, rod neurčen, jednotné číslo, sedmý pád (instrumentál), rod a číslo vlastníka neurčeny, osoba druhá, další kategorie neurčeny).

### 3.3.2.2.3 Morfologická desambiguace

---

Dalším krokem ve zpracování textů je morfologická desambiguace (odstranění víceznačnosti): proces, během kterého se u každého slovního tvaru jednoznačně určí jeden základní tvar a jedna morfologická značka, tedy vybere se z nabídky možných lemmat a morfologických značek, jež byly jednotlivým slovům přiřazeny v předchozí fázi značkování. Využívá se kombinace desambiguace založené na lingvistických pravidlech, která provede většinu desambiguační práce, a stochastického taggeru.

Tzv. „pravidlová“ desambiguace je založena na algoritmizovaných lingvistických poznacích, podle nichž v určitých kontextech nejsou některé kombinace morfologických kategorií a lemmat přípustné, tyto kombinace tedy odstraňuje a zachovává pouze morfologické značky, které jsou v daném kontextu gramatické. Například slovní tvar *se*, který může být interpretován buď jako vokalizované předložka *s* (*se světlem*), nebo jako zvrtné zájmeno *se* (*usmál se na ni*), nemůže být považován za předložku, pokud stojí před slovesem (*pomalů se otočil*) nebo slovem začínajícím na samohlásku (*vrátím se autobusem*). Tento postup je podrobněji popsán v článku Vl. Petkeviče (Petkevič, 2006).

V případě citované věty *Myslím, že je tady dostatečný pohodlí* vyřeší desambiguační pravidla celou větu na základě několika dílčích úvah: tvar *Myslím*, po němž následuje čárka a spojka *že* téměř jistě není tvarem slova *mysl* (jako je třeba v kontextu *klíč k jejich myslím*). Tvar *je* v části věty po podřadivé spojce, kde není žádné potenciální sloveso



ani adjektivum s akuzativní valencí nemůže být akuzativem zájmena *oni*, ale jen tvarem slovesa *být* ve třetí osobě jednotného čísla přítomného času. Obecněčeský genitiv slova substantiva *dostatečná* je v tomto kontextu zcela nepravděpodobný, tudíž tvar *dostatečný* bude určen jako adjektivum. U substantiva *pohodlí* lze vyloučit akuzativ, protože ve větě není žádné sloveso s akuzativní valencí a slovo *pohodlí* v tomto kontextu nemůže být ani případem akuzativu okolnostního. Díky (vysoce pravděpodobné) shodě adjektiva *dostatečný* s následujícím substantivem *pohodlí* v rodu, čísle a pádu se určí jmenný rod, číslo a pád adjektiva a číslo a pád substantiva (střední rod, jednotné číslo, nominativ).

Výsledkem je tedy věta, v níž je u každého slovního tvaru jednoznačně určen základní tvar slova a všechny relevantní gramatické kategorie, jak je vidět v následující tabulce. Ve třetím sloupci je uvedena skutečně užitá morfologická značka, ve čtvrtém sloupci zjednodušená slovní interpretace.

Tab. 3.3.2.2.3 – Výsledek morfologické disambiguace

Myslím	<i>mysl</i>	VB-S---1P-AA---I	substantivum, dativ pl.
,	,	Z:-----	interpunkce
že	<i>že</i>	J,-----	spojka
je	<i>být</i>	VB-S---3P-AA--I	sloveso, 3.os.sg.
tady	<i>tady</i>	Db-----	adverbium
dostatečný	<i>dostatečný</i>	AANS1--- -1A---6-	adjektivum, nom. sg. neut., mluv. čeština
pohodlí	<i>pohodlí</i>	NNNS1--- -1A---6-	substantivum, nom.sg.
.	.	Z:-----	interpunkce

Desambiguační pravidla leckdy dokážou správně určit všechny gramatické kategorie všech slovních tvarů ve větě, často se však objevují i případy, kdy u některého slovního tvaru nejde jen na základě kontextu spolehlivě určit vše. Tyto případy řeší stochastický morfologický tagger *Featurama* (viz <https://sourceforge.net/projects/featurama>), jehož autorem je Miroslav Spousta. Morfologický tagger je počítačový

program, který na základě manuálně označovaných dat (tzv. trénovacích dat), jimiž jsou v tomto případě morfologicky označovaná data Pražského závislostního korpusu (Hajič, 2004), vytváří model jazyka, který pak aplikuje na nová, neoznačovaná nebo jen částečně označovaná data. Výhodou taggeru je, že zpracuje s jednoznačným výsledkem jakýkoli vstup, včetně vstupu nestandardního (například s chybějící interpunkcí, s překlady aj.). Nevýhodou je, že funguje jako „černá skříňka“, chybu ve výstupu taggeru nelze jednoduše opravit. Úspěšnost kombinace desambiguačních pravidel a morfologického taggeru se podle různých měření (například Skoumalová, 2011) pohybuje okolo 95 % (cca 5 % slovních tvarů má chybně přiřazen tag), rozpoznání slovních druhů je výrazně úspěšnější, vykazuje jen kolem 0,5 % chyb.

#### 3.3.2.2.4 Automatická anotace frazémů

---

Po morfologické desambiguaci textu je zařazena automatická anotace víceslovných lexikálních jednotek, to znamená ustálená slovní spojení, která se v češtině používají jako taková (nebo s drobnými obměnami), často s odlišným významem než je prostý součet původních významů jednotlivých složek (*bílá vrána*, *mít něčeho dost*). Anotace frazémů se provádí na základě lexikální databáze a s ní spojených programů FRANTA (Hnátková 2011) a novějších databází víceslovných lexikálních jednotek (Petkevič & Hnátková, 2017). Významná část databáze pochází ze Slovníku české frazeologie a idiomatiky (Čermák et al., 2009).

#### 3.3.2.2.5 Doplnění lexikálních informací

---

Po provedení lematizace, morfologické anotace a anotace frazémů se zpětně nahrazují zástupné znaky pro specifické kombinace znaků (smajlíky), které by jinak mohly být v průběhu morfologické anotace chybně interpretovány. I těmto kombinacím znaků se připisují některé gramatické a lexikální kategorie.

Ke slovům v textu jsou pak doplněny další informace převážně lexikologického charakteru, rozlišují se například slovesa modální, slova

emočně nabitá, slova expresivní aj. Informace se přiřazují podle základních tvarů (lemmat) jednotlivých slov, zjišťuje se, zda lemma daného slova patří do toho či onoho slovníku (převzatého či ad-hoc vytvořeného) a podle toho se ke slovu doplňuje další informace. Lexikální kategorie jsou popsány výše (viz 3.3.2.1.2). Mezi elektronické slovníky, z nichž anotační nástroje čerpají, patří například slovník českých emočně nabitých slov (Veselovská & Bojar, 2013), v původní a částečně revidované verzi, frekvenční slovníky excerpované ze synchronního korpusu SYN2015 (Křen et al. 2015) aj.

### 3.3.2.2.6 Postprocessing textů

---

Nakonec jsou texty zpracovány tak, aby bylo možné všechny zjištěné lingvistické parametry využívat statistickými nástroji: na každém řádku je jedno slovo z původního textu, morfologické značky jsou rozděleny na jednotlivé pozice odpovídající jednotlivým gramatickým kategoriím, v samostatných sloupcích jsou pak i další informace o daném slovním tvaru, celkem 26 různých proměnných.

U každého textu je pak pro každou proměnnou (například slovní druh nebo pád) sečten počet výskytů všech možných hodnot (například počet substantiv nebo počet instrumentálů). Dopočítávají se některé kombinované parametry (viz 3.3.2.1.5) a další proměnné závislé na celkovém počtu slov v textu nebo ve větě, například jazyková pestrost textu. Výsledkem je jedna souhrnná tabulka obsahující identifikátory textů (tedy identifikátor respondenta a typu textu) a číselné hodnoty pro 181 lingvistických parametrů.

### 3.3.2.3

#### **Srovnání vybraných gramatických kategorií ve sledovaných textech a v korpusech ČNK**

---

Pro jednoduché zařazení sledovaných celků do širšího lingvistického kontextu je zde uvedeno srovnání dvou morfologických kategorií: podíl výskytu slovních druhů a podíl výskytu pádu u substantiv v souhrnu sledovaných textů (P200), ve třech žánrových skupinách textů

v korpusu SYN2005 (Čermák et al., 2005) a v mluveném korpusu ORAL (Kopřivová et al., 2017). Údaje o výskytech slov s příslušnými gramatickými kategoriemi v korpusu SYN2005 jsou převzaty z publikace Statistiky češtiny (Bartoň et al., 2009, s. 130-132). Podrobnější lingvistická analýza textů pak bude představena v kapitole 3.4.3.

*Tab. 3.3.2.3.A – Podíl slovních druhů*

	P200	BEL	ODB	PUB	ORAL
N	21,8	24,3	34,5	33,8	12,0
A	8,7	8,9	15,4	12,2	3,7
P	15,9	14,9	7,8	8,8	21,5
C	1,4	1,6	3,3	3,3	1,8
V	22,3	21,2	13,9	16,0	21,1
D	9,4	8,4	5,5	6,2	13,3
R	9,9	9,8	10,8	11,5	5,8
J	9,5	8,9	7,5	6,7	11,9
T	1,1	1,8	1,3	1,5	8,4
I	0,1	0,1	0,0	0,0	0,5
%	100	100	100	100	100

Tabulka 3.3.2.3.A ukazuje podíl slovních druhů v analyzovaných textech celkem, ve třech žánrových subkorpusech korpusu psaných textů SYN2005 (BEL: beletrie, ODB: odborná literatura, PUB: publicistika) a v mluveném korpusu ORAL. Slovní druh je označen podle užití morfologické analýzy: N substantivum, A adjektivum, P zájmeno, C číslovka, V sloveso, D příslovce, R předložka, J spojka, T částice, I citoslovce.

Z tabulky je zřejmé, že podíl slovních druhů v analyzovaných textech je nejbližší žánrovému subkorpusu beletrie, s výrazným odstupem od ostatních žánrů a korpusů. Odpovídá to textu psanému (rozdíl oproti mluvenému korpusu ORAL je značný), nicméně psanému relativně spontánně, s vyšší dějovostí (vyšší podíl sloves, nižší zastoupení substantiv a adjektiv) oproti jiným psaným žánrům.

Tab. 3.3.2.3.B – Podíl pádů substantiv

	P200	BEL	ODB	PUB	ORAL
1	23,4	29,8	25,9	31,0	32,6
2	20,2	22,0	32,6	28,0	15,6
3	4,1	3,9	3,6	3,4	2,3
4	27,4	22,5	17,1	18,0	31,1
5	2,6	0,9	0,1	0,1	2,3
6	12,3	11,4	11,8	12,2	10,6
7	10,0	9,5	8,9	7,3	5,5
%	100	100	100	100	100

Tabulka 3.3.2.3.B ukazuje podíl pádů u substantiv v analyzovaných textech celkem, ve třech žánrových subkorpusech korpusu SYN2005 (BEL: beletrie, ODB: odborná literatura, PUB: publicistika) a v mluveném korpusu ORAL. Pád je označen číslem odpovídajícím tradičnímu řazení pádů v české jazykovědě (1 nominativ, 2 genitiv, 3 dativ, 4 akuzativ, 5 vokativ, 6 lokál, 7 instrumentál). I zde jsou sledované texty nejpodobnější subkorpuse beletrie, vyšším zastoupením akuzativu a vokativu se texty přibližují korpusu mluvené češtiny ORAL. Výše uvedené celkové hodnocení analyzovaných textů tak platí i u tohoto parametru, ovšem s tím, že texty se jeví bližší mluveným, než bylo zřejmé z první zde probírané proměnné.

## 3.4

# Zpracování dat a výsledky výzkumu CPACT

## 3.4.1

# Statistická explorace vztahů mezi osobnostními charakteristikami a textovými parametry

*Jiří Haviger*

Analýza vztahů mezi osobnostními charakteristikami a textovými parametry je v prvním výzkumném modulu, resp. kvótním souboru P200, prováděna primárně pomocí korelačních koeficientů a jejich testování na nulovou hodnotu. Vzhledem k testování velkého množství vztahů bylo nutné přistoupit ke korekcím výsledků jednotlivých testů. Použité metody korekce jsou popsány v této kapitole.

### 3.4.1.1 Popis použitých korekcí

---

Základním kamenem statistické inference a příslušných statistických testů je zobecnitelnost výsledků z náhodně vybraného výzkumného vzorku na celou populaci. Předpokládá se (pro někoho možná paradoxně), že v populaci hledaný stav není (tzv. nulová hypotéza,  $H_0$ ) a ze statistik (např. průměru a směrodatné odchylky) výzkumného vzorku a jeho velikosti se usuzuje, s jakou pravděpodobností bylo možno vybrat z populace s  $H_0$  dosahující takovýchto nebo ještě více extrémních statistik. Pravděpodobnost je popsána tzv. p-hodnotou a ta je porovnána s předem deklarovanou hodnotou  $\alpha$ , nazývanou hladina významnosti (nejčastěji  $\alpha=0,05$ ). Pokud je p-hodnota menší než  $\alpha$ , pak zamítáme platnost  $H_0$  u populace a zjednodušeně tvrdíme, že jsme učinili **objev** o tom, že v populaci hledaný vztah je. Pro objev se také často užívají pojmy **pozitivní výsledek** nebo **statisticky signifikantní vztah** (angl. *discovery*).

Rozhodování o populaci na základě statistických testů může trpět dvěma zásadními typy chyb:

1. chyba prvního druhu, tedy že jsme prezentovali objev, který však v populaci není, tedy **falešně pozitivní výsledek testu**,
2. chyba druhého druhu, tedy že jsme neobjevili vztah, který v populaci je, tedy **falešně negativní výsledek testu**.

Při studiích s větším množstvím testů se výrazně zvyšuje pravděpodobnost prezentace falešně pozitivních výsledků, tedy výsledků, které ukazují na vědecky zajímavý vztah, který však ve skutečnosti nenastal – je přítomen pouze ve výběru, nikoli v populaci. Při hladině významnosti  $\alpha=0,05$  na jednotlivý test, která udržuje chybu prvního druhu v přijatelných mezích, je však již pro 10 testů pravděpodobnost falešné signifikance alespoň jednoho testu 0,401 a při 100 testech se tato pravděpodobnost zvyšuje na 0,994, což je prakticky jistota. Metody korekcí takovýchto chyb se zaměřují na p-hodnoty příslušné k testům, jejich adjustaci a případně na korekci prahové hodnoty  $\alpha$ . Tyto metody jsou ve stručnosti založeny na třech přístupech:

1. korekce směřující k udržení pravděpodobnosti alespoň jednoho falešně pozitivního výsledku (angl. *family wise error rate*, FWER) pod hladinou 0,05,
2. hierarchické modely, které jsou založeny na postupném dělení základní množiny dat,
3. korekce směřující k udržení podmíněné pravděpodobnosti počtu falešně pozitivních výsledků z pozitivních výsledků (angl. *false discovery rate*, FDR) pod hladinou 0,05.

První přístup je vhodný pro situace, kdy je relativně málo testování, nebo kdy každý chybný výsledek vede k závažným důsledkům. Například u mikrobiologických testů nemoci by se neměla objevit žádná chybná diagnóza. Druhý přístup je vhodný, pokud máme výběr segmentovatelný na několik skupin. Například pokud máme skupinu nemocných s několika diagnózami, prověříme vztah nejprve v celé skupině a pokud vyjde signifikantní, pak postupujeme směrem k jednotlivým diagnózám. Třetí přístup je vhodný, pokud máme velké množství testů



a malou míru falešně pozitivních výsledků lze tolerovat. Například při testování kvality výrobků z výrobní linky lze tolerovat malé množství kvalitních výrobků, které neprojdou kontrolou, nebo např. při skenování mozku a testování jeho obrazu na přítomnost nádoru lze tolerovat, že například z detekovaných 10000 pixelů ukazujících na nádor je 300 detekováno chybně.

Pro analýzu vztahů v tomto výzkumu byly použity tři metody korekce, které ve stručnosti popíšeme:

1. **Šidákova korekce falešně pozitivních výsledků (FWER)**, která minimalizuje chybu prvního druhu celého souboru testů (tj. i těch, které nevyšly pozitivně) úpravou prahové hodnoty  $\alpha \rightarrow \alpha\text{-SID}$  pro určení signifikance každého jednotlivého testu (Šidák, 1967).
2. **Benjamini – Hochbergova korekce (FDR)** – stanovení hranice pro p-hodnoty tak, aby byla pravděpodobnost falešně pozitivních výsledků pouze ze souboru pozitivních výsledků kontrolována v požadovaných mezích (Benjamini & Hochberg, 1995).
3. **Storeyova korekce (FDR)** – jiný způsob stanovení hranice pro p-hodnoty tak, aby byla pravděpodobnost falešně pozitivních výsledků pouze ze souboru pozitivních výsledků kontrolována v požadovaných mezích (Storey, 2002).

#### 3.4.1.1.1 Šidákova FWER korekce (SID)

---

Jednotlivé testování je standardně prováděno na hladině  $\alpha = 0,05$ , tedy s přípustnou pětiprocentní možností chybného zamítnutí nulové hypotézy. Pokud je prováděno m testů, je pravděpodobnost této chyby u alespoň jednoho testu (FWER), kde  $\alpha_{\text{FWER}} = (1 - \alpha)^m$ . Pokud chceme kontrolovat FWER např. na hladině  $\alpha = 0,05$ , je nutno upravit  $\alpha \leftarrow \alpha_{\text{SID}} = 1 - (1 - \alpha_{\text{FWER}})^m$ . Připomeňme, že známá Bonferroniho korekce  $\alpha \leftarrow \alpha_{\text{BONF}} = \alpha/m$  je pro hodnoty  $\alpha \approx 0$  linearizovanou verzí Šidákovy korekce a při praktickém použití dává totožné výsledky jako Šidákova korekce. Například pro:  $m = 1000$ ,  $\alpha = 0,05$  je  $\alpha_{\text{SID}} = 5,13 \cdot 10^{-5}$ ,  $\alpha_{\text{BONF}} = 5,00 \cdot 10^{-5}$  a jejich rozdíl je tedy o řád nižší než samotné upravené

hodnoty. Obě metody jsou známy jako velmi konzervativní, tedy že pro zachování nízké chyby prvního druhu výrazně zvyšují chybu druhého druhu, tj. velké množství objevů neodhalí. Populárně řečeno je pak test vůči velké části objevů slepý.

### 3.4.1.1.2 Benjamini Hochberg FDR korekce (BH)

---

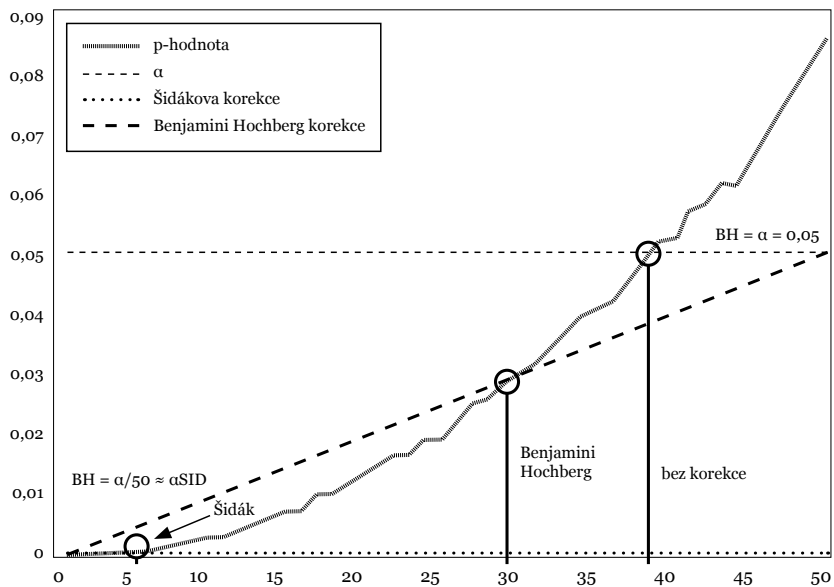
Metody kontroly FDR (*false discovery rate*) vycházejí z předpokladu, že je vhodné udržet chybu chybně prezentovaných objevů (chybu prvního druhu) v daných mezích relativně k počtu objevů kontrolovaných na hladině  $\alpha_{\text{FDR}} = 0,05$ . Metoda BH (Benjamini & Hochberg, 1995) seřadí všechny p-hodnoty vzestupně  $p_1 < \dots < p_m$ , pak stanoví  $k = \max \{ i \mid p_i \leq \alpha_{\text{FDR}} \cdot i/m \}$  a na základě získaného k upraví hodnotu  $\alpha_{\text{FDR}} \leftarrow \alpha_{\text{BH}} = \alpha_{\text{FDR}} \cdot k/m$ . Lineární charakter této korekce je vidět na následujícím demonstrativním obrázku 3.4.1.1.2, na kterém je i porovnání s Šidákovou korekcí a s  $\alpha$  bez korekce. Šikmá přímka začíná v bodě  $\alpha_{\text{BH}} = \alpha/m = \alpha_{\text{BONF}}$ , na nejnižší p-hodnotu je tedy kladeno stejné měřítko jako při Bonferroniho korekci. Končí na hladině  $\alpha=0,05$ , na největší p-hodnotu je tedy kladeno měřítko bez korekce. Protože přímka má směrnici 0,05, je dodržení FDR na hladině 0,05 zřejmé.

### 3.4.1.1.3 Storey FDR korekce (ST)

---

Alternativní způsob kontroly FDR je aproximace hodnoty určující horizontální linii  $y = \pi_0$  v grafu hustoty pravděpodobnosti reprezentující distribuci p-hodnot všech provedených testů. Distribuce p-hodnot splňuje teoreticky dva předpoklady: pro výběry z populace s nulovou hypotézou je rovnoměrná, pro výběry z populace s alternativní hypotézou je klesající a pro p-hodnoty blízké se k jedné se četnosti p-hodnot blíží k nule. Protože ve výzkumech bývají zastoupeny obě varianty (nulová i alternativní hypotéza) odpovídá distribuce p-hodnot takzvanému mix-modelu (Storey, 2002). Horizontální linie  $y = \pi_0$  pak odděluje oblast platnosti nulové a alternativní hypotézy. Na jejím základě je ke každé p-hodnotě  $P_i$  přiřazena příslušná q-hodnota  $q_i$ , která vyjadřuje FDR, pokud by se příslušná p-hodnota nastavila jako  $\alpha_{\text{FDR}} \leftarrow p_i$  (Storey,

Obr. 3.4.1.1.2 – Ilustrační graf 50 p-hodnot seřazených vzestupně; osa x zobrazuje pořadí p-hodnoty v seřazeném souboru, osa y její hodnotu. Konkrétní p-hodnoty jsou vyznačeny čárkovanou linií. Čárkovaně je zobrazena omezující přímka pro oddělení p-hodnot bez korekce (horní vodorovná), s Šidákovou korekcí (tečkovaná dolní vodorovná) a s Benjamini-Hochbergovou korekcí (čárkovaná šikmá)



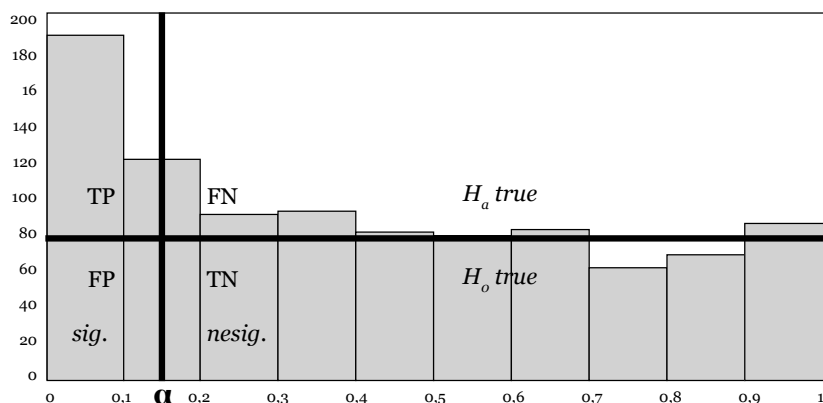
2002). Předpokládanou distribuci p-hodnot ve výběrovém souboru ukazuje obr. 3.4.1.1.3 níže. V grafu jsou znázorněny oblasti pravdivě pozitivní (*true positive* TP), falešně pozitivní (*false positive* FP), pravdivě negativní (*true negative* TN) a falešně negativní (*false negative* FN), v obrázku ilustrativně pro hodnotu  $\alpha = 0,15$ .

Na rozdíl od BH metody akcentuje tato metoda konkrétní charakter distribuce p-hodnot a je vizuálně přehlednější. Pro praktické užití však musí splňovat dvě podmínky:

1. Charakter grafu hustoty musí mít ustálenou pravou stranu, na jejímž základě se odhaduje horizontální linie;

- počet p-hodnot pocházejících z  $H_0$  musí být srovnatelný nebo menší s počtem p-hodnot pocházejících z  $H_A$ , aby horizontální linie nebyla příliš vysoko a alespoň některé q-hodnoty dosáhly hodnot menších než  $\alpha\text{FDR} = 0,05$ .

Obr. 3.4.1.1.3 – Ilustrační distribuce p-hodnot s vyznačením zón objevů (*sig.*), nesignifikantních výsledků (*nesig.*), platnosti nulové hypotézy ( $H_0$  true), oblasti platnosti alternativní hypotézy ( $H_A$  true) a průniků těchto oblastí true positive (TP), true negative (TN), false positive (FP), false negative (FN)



Pro splnění první podmínky je nutné dostatečné množství testů, pro splnění druhé podmínky je vhodné odstranit skupiny testů pocházející zřejmě z  $H_0$  (korektně, podložitelně). Splnění obou podmínek tedy vyžaduje vyvažování mezi dostatečně velkým souborem testů a přiměřeně malým souborem testů pocházejících z  $H_0$ .

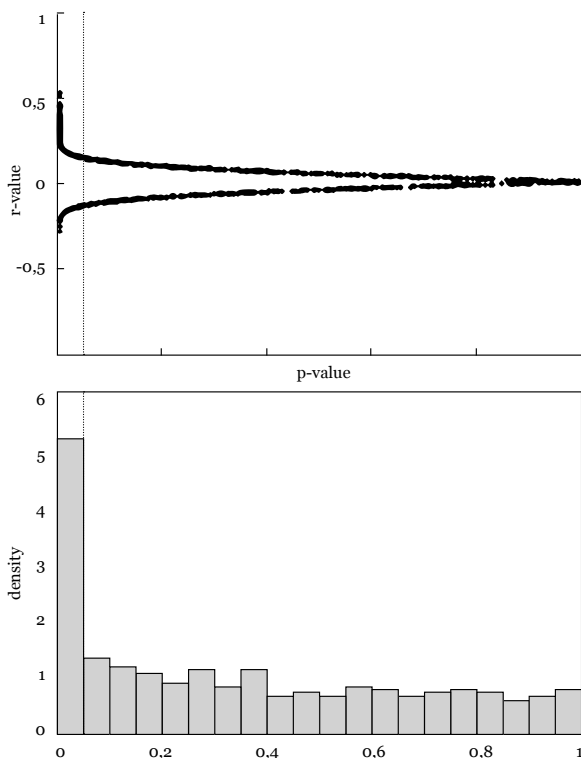
### 3.4.1.2 Provedení základní analýzy

Pro kvótní výběr se prováděla korelační analýza lingvistických parametrů šesti různých textů vzhledem k psychologickým testům a jejich škálám. Významnost korelačních koeficientů roste symetricky se vzdáleností od nuly a příslušné p-hodnoty naopak se vzrůstajícím

vzdáleností korelačního koeficientu od nuly klesají. Vztah mezi hodnotou korelačního koeficientu a distribucí p-hodnot ukazuje následující obr. 3.4.1.2.A.

Základní analýza proběhla pomocí Storeyovy FDR korekce pro všechny vztahy mezi textovými parametry a osobnostními charakteristikami, a to pro každý text zvlášť. Psychologická baterie obsahovala 11 testů ve verzi S a 9 testů ve verzi O, celkově bylo operováno s 85 proměnnými. V lingvistické části bylo využito 120 proměnných. Celkově tedy bylo analyzováno 10 200 vztahů pro každý text, použití korekcí je tedy nezbytné.

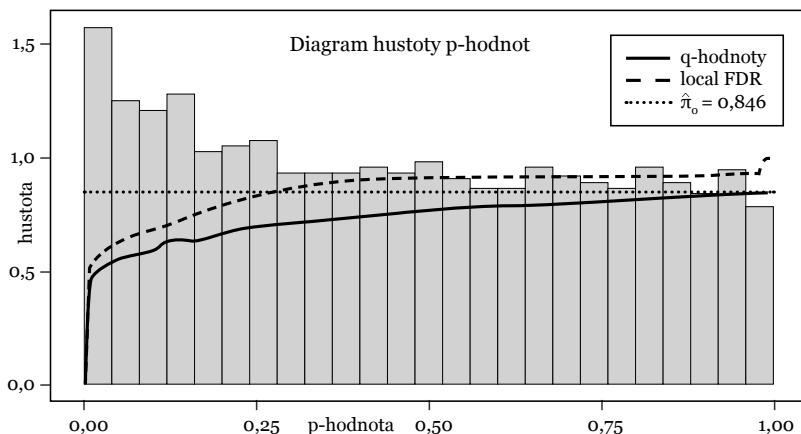
*Obr. 3.4.1.2.A – Ilustrační obrázek vztahu korelačních koeficientů k distribuci p-hodnot. Legenda: r value = hodnota korelačního koeficientu, density = hustota, p-value = p-hodnota*



### 3.4.1.2.1 Distribuce p-hodnot motivačního dopisu (TXT1)

Distribuce p-hodnot u písemné žádosti o práci má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,846$  lze říci, že 84,6 % výsledků pochází z populace s nulovou hypotézou a 15,4 % z populace s alternativní hypotézou. Levá část grafu ukazuje na pomalý klesající trend, jehož důsledkem je, že i při provedení korekce bude relativně velká chyba druhého druhu.

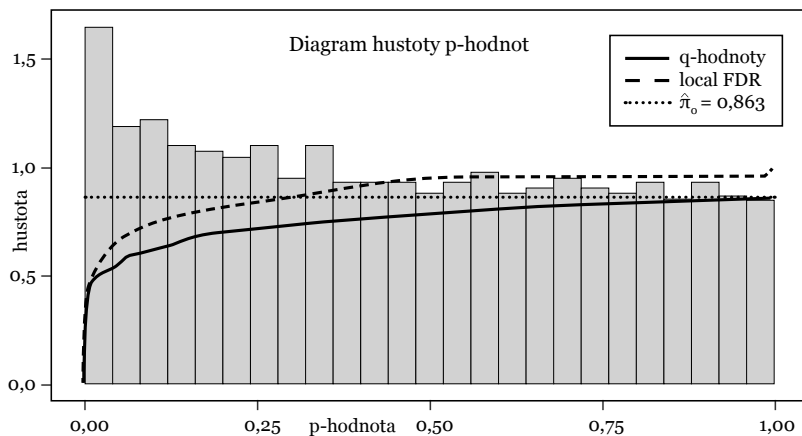
Obr. 3.4.1.2.1 – Vztahy mezi osobnostními charakteristikami a textovými parametry motivačního dopisu: distribuce p-hodnot (P200)



### 3.4.1.2.2 Distribuce p-hodnot dopisu z dovolené (TXT2)

Graf u dopisu z dovolené má podobný tvar jako u žádosti o práci, proto i komentáře jsou podobné. Distribuce p-hodnot má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,863$  lze říci, že 86,3 % výsledků pochází z populace s nulovou hypotézou a 13,7 % z populace s alternativní hypotézou. Levá část grafu ukazuje na pomalý klesající trend, jehož důsledkem je, že i při provedení korekce bude relativně velká chyba druhého druhu.

Obr. 3.4.1.2.2 – Vztahy mezi osobnostními charakteristikami a textovými parametry dopisu z dovolené: distribuce p-hodnot (P200)



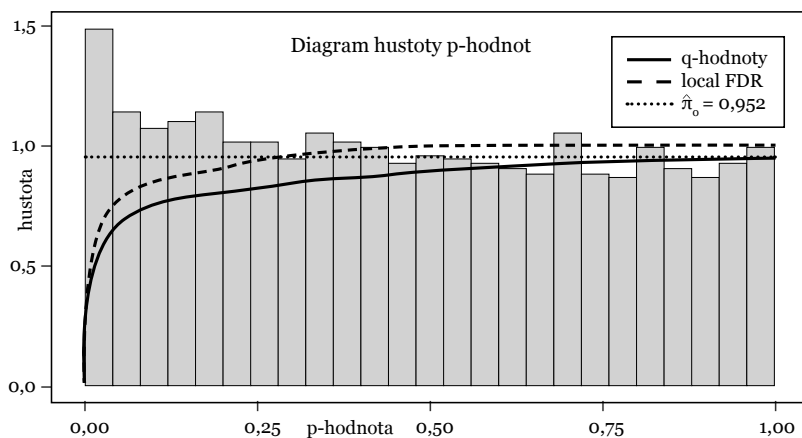
### 3.4.1.2.3 Distribuce p-hodnot stížnosti (TXT3)

Distribuce p-hodnot u stížnosti má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,952$  lze říci, že 95,2 % výsledků pochází z populace s nulovou hypotézou a 4,8 % z populace s alternativní hypotézou. Obr. 3.4.1.2.3 ukazuje na nutnost výrazně redukovat počet testů pocházejících z  $H_0$ , protože bez redukce je takovéto rozdělení nepoužitelné.

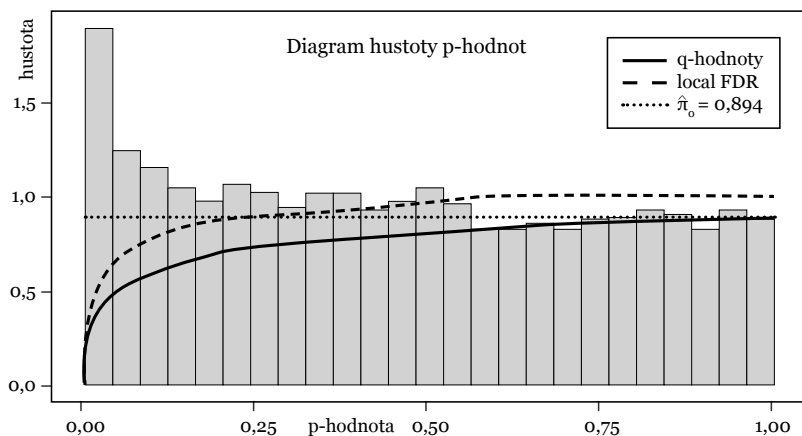
### 3.4.1.2.4 Distribuce p-hodnot dopisu s omluvou (TXT4)

Distribuce p-hodnot u omluvy má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,894$  lze říci, že 89,4 % výsledků pochází z populace s nulovou hypotézou a 10,6 % z populace s alternativní hypotézou. Levá část obr. 3.4.1.2.4 ukazuje na rychlý klesající trend, jehož důsledkem je, že při provedení korekce bude chyba druhého druhu nižší než v případě TXT1.

Obr. 3.4.1.2.3 – Vztahy mezi osobnostními charakteristikami a textovými parametry stížnosti: distribuce p-hodnot (P200)



Obr. 3.4.1.2.4 – Vztahy mezi osobnostními charakteristikami a textovými parametry dopisu s omluvou: distribuce p-hodnot (P200)

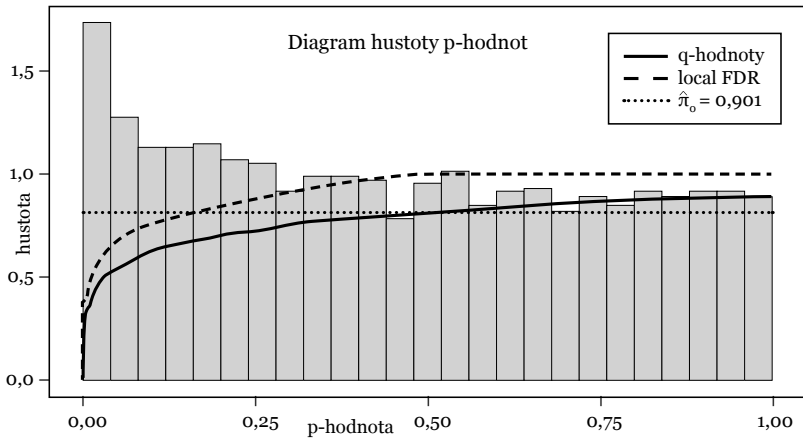




### 3.4.1.2.5 Distribuce p-hodnot motivačního pohovoru (TXT5)

Distribuce p-hodnot u motivačního pohovoru má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,901$  lze říci, že 90,1 % výsledků pochází z populace s nulovou hypotézou a 9,9 % z populace s alternativní hypotézou. Levá část grafu ukazuje na mírný klesající trend, jehož důsledkem je, že i při provedení korekce bude relativně velká chyba druhého druhu.

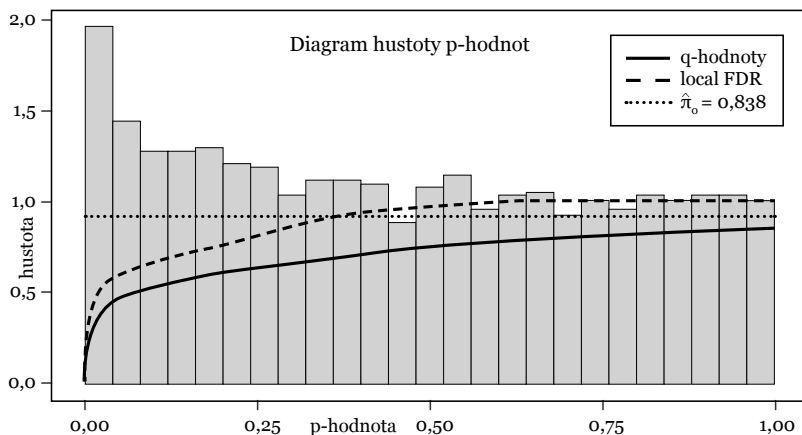
Obr. 3.4.1.2.5 – Vztahy mezi osobnostními charakteristikami a textovými parametry motivačního pohovoru (TXT5): distribuce p-hodnot (P200)



### 3.4.1.2.6 Distribuce p-hodnot vyprávění o příjemném zážitku (TXT6)

Distribuce p-hodnot u ústního vyprávění má splněnou podmínku ploché pravé části, je tedy vhodná pro odhad horizontální linie  $y = \pi_0$ . Z hodnoty  $\pi_0 = 0,83,8$  lze říci, že 83,8 % výsledků pochází z populace s nulovou hypotézou a 11,2 % z populace s alternativní hypotézou. Levá část grafu ukazuje na mírný klesající trend, jehož důsledkem je, že i při provedení korekce bude relativně velká chyba druhého druhu.

Obr. 3.4.1.2.6 – Vztahy mezi osobnostními charakteristikami a textovými parametry Vyprávění (TXT6): distribuce p-hodnot (P200)



### 3.4.1.3 Výsledky první fáze – Storeyova korekce FDR

Ukazuje se, že **hledat vztahy mezi textovými parametry a osobnostními charakteristikami autora je na základě výše uvedených výsledků smysluplné, protože v každém z textů pochází přiměřená část výsledků (5–15 %) z alternativních hypotéz.** Celkově výsledky obsahují velké množství korelací pocházejících z populace s  $H_0$  a proto přímé využití Storeyovy metody FDR ani jiných metod FDR není vhodné.

Výsledky Storeyovy korekce FDR pro jednotlivé testy přináší následující tabulky 3.4.1.3.A a 3.4.1.3.B, porovnávající jednotlivé psychologické testy se všemi textovými parametry. V buňkách jsou odhady na základě distribuce p-hodnot, zda daný osobnostní rys reprezentovaný psychologickým testem má vztah k parametrům příslušného textu.

Psychologické testy ve verzi **sebeposouzení (S)** obsahují významné vztahy se všemi texty. Slabé výsledky ukazují na testy BOBR, SPSBS, SHQ a SMS. Ve verzi **posouzení druhou osobou (O)** se významnější vztahy ukazují pouze BFI, IAS a SMS, u ostatních testů jen ojedinele.

Tab. 3.4.1.3.A – Odhad objevů při sebeposouzení (S): distribuce p-hodnot (P200)

<b>Test</b>	<b>TXT1</b>	<b>TXT2</b>	<b>TXT3</b>	<b>TXT4</b>	<b>TXT5</b>	<b>TXT6</b>
BFI	N	D	D	D	D	D
BOBR	N	N	N	N	N	D
DASS	D	D	D	D	N	N
IAS	D	D	D	D	D	D
MMG	D	D	D	D	D	N
PSSI	N	D	N	D	D	D
SHQ	N	N	N	D	N	D
SMS	N	N	N	D	D	N
SPSBS	N	D	N	N	N	N
SSI	N	D	N	D	N	D
STAIX	N	D	N	D	D	D

*Legenda: D (discovery) = potenciál pro objevy,  
N (nondiscovery) = bez objevů.*

Tab. 3.4.1.3.B – Odhad objevů při posouzení druhou osobou (O): distribuce p-hodnot (P200)

<b>Test</b>	<b>TXT1</b>	<b>TXT2</b>	<b>TXT3</b>	<b>TXT4</b>	<b>TXT5</b>	<b>TXT6</b>
BFI	D	N	N	D	D	D
BOBR	D	N	N	N	N	N
DASS	N	N	N	N	N	N
IAS	D	D	N	N	N	D
MMG	N	N	N	N	N	N
PSSI	D	N	N	N	N	N
SHQ	D	N	N	N	N	N
SMS	D	N	D	N	N	D
SPSBS	N	N	N	N	N	N

*Legenda: D (discovery) = potenciál pro objevy,  
N (nondiscovery) = bez objevů.*

Z výsledků plyne, že některé osobnostní charakteristiky se do textových parametrů promítají a jiné ne. U těch prvních pak má smysl pokračovat s dalšími analýzami. Protože Storeyova metoda FDR pro jednotlivé psychologické testy selhala z důvodů malého počtu testů, byla od této úrovně používána Benjamini-Hochbergova korekce FDR pro minimalizaci rizika falešných objevů.

### 3.4.1.4 Výsledky druhé fáze – Benjamini-Hochbergova korekce FDR

Pro jednotlivé psychologické testy byla provedena BH korekce, získané výsledky najdete v příloze 3.4.1.4. Přehled výsledků po této korekci shrnují tabulky 3.4.1.4.A a 3.4.1.4.B, ve které jsou počty korelací, které jsou statisticky signifikantní po BH korekci. Kompletní dokumentaci je možné nalézt v příloze 3.4.1.4.

*Tab. 3.4.1.4.A – Počty objevů při sebeposouzení (S): korelace s Benjamini-Hochbergovou korekcí (P200)*

Test	TXT1	TXT2	TXT3	TXT4	TXT5	TXT6
BFI	0	4	0	11	10	10
BOBR	0	0	0	0	0	0
DASS	0	3	33	1	0	0
IAS	0	0	0	1	0	5
MMG	23	0	1	0	0	0
PSSI	0	0	0	0	0	0
SHQ	0	0	0	0	0	0
SMS	0	0	0	3	0	0
SPSBS	0	0	0	0	0	0
SSI	0	0	0	0	1	5
STAIX	1	0	6	9	0	1

Tab. 3.4.1.4.B – Počty objevů při posouzení druhou osobou (O):  
korelace s Benjamini-Hochbergovou korekcí (P200)

Test	TXT1	TXT2	TXT3	TXT4	TXT5	TXT6
BFI	0	0	0	6	0	3
BOBR	0	0	0	0	0	0
DASS	0	0	0	0	0	0
IAS	0	1	0	0	0	0
PSSI	0	0	0	0	0	0
SHQ	0	0	0	0	1	0
SMS	2	0	0	0	0	0
SSI	0	0	7	0	0	0
STAIX	0	0	0	0	0	0

Porovnáním odhadovaných objevů na základě tvaru distribuce s uskutečněnými objevy je zřejmé, že některé odhady byly optimistické a po BH korekci se je nepodařilo potvrdit. Jak již bylo zmíněno, tak většina grafů distribuce p-hodnot má pomalu klesající tvar, takže větší množství testů pocházejících z alternativních hypotéz nepůjde odhalit. Tento odhad potvrzuje následující tabulka č. 3.4.1.4.C.

Tab. 3.4.1.4.C – Celkový souhrn počtu objevů pro jednotlivé texty

	TXT1	TXT2	TXT3	TXT4	TXT5	TXT6
Počet analyzovaných korelací	10200	10200	10200	10200	10200	10200
Odhad $\pi_0$	0,846	0,863	0,952	0,894	0,901	0,838
Odhad počtu alternativ dle $\pi_0$	1571	1397	490	1081	1010	1142
Počet objevů po BH korekci	26	8	47	31	12	24
Precision (konst.)	0,950	0,950	0,950	0,950	0,950	0,950
Recall	0,017	0,006	0,096	0,029	0,012	0,021

Pro doplnění tab. 3.4.4.C uvádíme, že v textu TXT1 jsme analyzovali 10200 korelací, odhad podle  $\pi_0$  ukazuje na 1571 alternativ, tedy vědecky zajímavých vztahů. Z celkového analyzovaného počtu korelací se jedná o 15,4 %. Vzhledem k pomalému klesání distribuce se z těchto vztahů podařilo spolehlivě detekovat 26, což je 1,7 % z alternativ. Hodnota *precision* ukazuje na to, že z 26 prezentovaných výsledků je  $0,95 \cdot 26 = 24,7$  spolehlivých, jinými slovy lze očekávat, že mezi nalezenými vztahy může být nejvýše jeden (přesněji 1,3) falešně pozitivní výsledek. Hodnota *recall* ukazuje na skutečnost, že z 1571 vědecky zajímavých vztahů jsme spolehlivě odhalili 1,7 %. V dalších textech je hodnota *precision* stejná (metoda BH zajišťuje stejnou hodnotu) a hodnota *recall* se pohybuje v rozmezí od 0,6 % do 9,6 %.

### 3.4.1.5 Závěr a shrnutí

---

V první části kapitoly jsme představili vybrané metody korekce falešně pozitivních výsledků. Ve druhé části jsme ukázali aplikaci těchto postupů na naše data a na závěr jsme získané výsledky sumarizovali v tabulce 3.4.1.4.C. Na jejím základě můžeme konstatovat:

1. Po provedení výpočtů a korekcí bylo detekováno celkem **148 prokazatelných vztahů** mezi osobnostními charakteristikami respondentů a parametry jejich textů.
2. Pravděpodobnost, že je mezi prezentovanými vztahy nějaký falešně pozitivní, je díky BH korekci **velmi malá**.
3. Prezentované výsledky jsou na základě odhadů z distribuce p-hodnot jen velmi **malou částí vědecky zajímavých vztahů** mezi osobnostními charakteristikami respondentů a jejich projevy v psaném a mluveném textu.

Pokud tedy budeme parafrázovat Járu Cimrmana: „Nevíme skoro nic, zato to víme přesně“ (Cimrman & Smoljak, 1992).

## 3.4.2 Zpracování výsledků dotazníkových variant sebeposouzení a posouzení druhou osobou

*Dalibor Kučera, Jiří Haviger*

V kapitole „Problematika dotazníkového sebeposouzení a posouzení druhým v psychologickém výzkumu“ (2.3) byla podrobně popsána otázka využití dotazníků při psychologickém testování, a to z pohledu možného zkreslování výsledků respondenty, mj. pod vlivem role a perspektivy, kterou posuzovatel k posuzovanému zaujímá. Klíčovým tématem byla asymetrie, ke které často dochází mezi různými variantami posouzení. Tuto asymetrii, již sledovanou v rámci dat CPACT, zde uvádíme do praktického metodologického rámce a rozpracováváme tři hypotézy, které zazněly v odkazované teoretické kapitole.

### 3.4.2.1 Cíle studie a hypotézy

---

Cílem této kapitoly je prezentovat výsledky vztahující se k asymetrii výsledků různých variant dotazníkového posouzení, zejména ve vztahu k teorii „pozorovatelnosti“ a „evaluativnosti“ osobnostních charakteristik, která byla podrobněji popsána v kapitole 2.3. Pro krátké shrnutí: zatímco **pozorovatelnost** (*observability*) popisuje sledovanou charakteristiku z hlediska toho, do jaké míry je její zaregistrování vnějším pozorovatelem snadné (např. vnějšíkové charakteristiky, tj. společenskost, komunikativnost, dominance apod., oproti obtížněji viditelným charakteristikám, např. úzkostnosti či sebeúctě), **evaluativnost** (*evaluativeness*) definuje nenáročnost jejího celkového zhodnocení, tedy jakési „uchopení“ posuzovatelem (např. vysoce evaluativní inteligence, kreativita, atraktivita, společenský status, oproti méně evaluativním charakteristikám jako např. závislost, paranoia či dominance) (viz Andersen, Glassman, & Gold, 1998; Mehl, 2006a; Vazire, 2010).

Díky designu výzkumu CPACT, ve kterém byla většina dat z dotazníků získávána jak ve variantě sebeposouzení, tak ve variantě posouzení druhou (blízkou) osobou, jsme získali poměrně rozsáhlý soubor dat, na kterém bylo možné sledovat řadu parametrů a výstupů. V naší studii jsme se pak zaměřili na posouzení konkrétních vybraných charakteristik:

- Vysoce pozorovatelné a vysoce evaluativní: sociální vyjadřování (test SSI, škála SE), extraverte (test BFI, škála E) a svědomitost (test BFI, škála S).
- Málo pozorovatelné a málo evaluativní: neuroticismus (test BFI, škála N), emoční citlivost (test SSI, škála ES) a přecitlivělost na výrazné chování druhých (test SMS, škála SEBO).

**První hypotézou** je teze, že hodnocení v charakteristikách, které jsou obtížně pozorovatelné a evaluativní (zde N, ES a SEBO), bude mezi oběma variantami posouzení (sebe- a posouzení druhým) zřejmě a signifikantní rozdílnost (viz Funder, 2001; Funder, 2012).

**Druhou hypotézou** je předpoklad, že u škál s vysokou pozorovatelností a evaluativností (zde SE, E a S) bude míra shody obou variant posouzení vyšší oproti ostatním uvedeným škálám (N, ES a SEBO) (viz např. Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008; Connolly, Kavanagh, & Viswesvaran, 2007).

**Třetí hypotézou** je předpoklad, že u posuzovatelů, kteří deklarují, že znají posuzovaného déle, bude zjištěna vyšší shoda jejich hodnocení s výsledky sebeposouzení dané osoby. Tato hypotéza se vztahuje k vlivu délky vztahu dvou osob na míru shody v posouzení. Sledováno je, zda posuzovatelé, kteří deklarují delší dobu vzájemného vztahu s posuzovaným, dosahují také vyšší míry shody v hodnocení osobnostních charakteristik (tj. shoda posouzení se sebeposouzením; viz např. Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008 a Colvin & Funder, 1991). Zde budou sledovány dimenze v modelu Big Five (resp. výsledky dotazníku BFI), konkrétně extraverte (E), neuroticismus (N), otevřenost (O), svědomitost (C) a přívětivost (A) (podrobněji viz níže).



### 3.4.2.2 Soubor a metody

---

Parametry výzkumu a jeho design jsou podrobně popsány v kapitole 3.2. Zde pro kompletnost uvádíme jejich stručné shrnutí.

Výzkumu CPACT se účastnilo 100 dvojic osob starších 15 let ( $N=200$ ), získaných kvótním výběrem, které se do výzkumu dobrovolně přihlásily pomocí elektronického rozhraní a dostavily se na jednodenní výzkumné sezení. Podmínkou účasti bylo, aby se obě osoby ve dvojici dobře znaly a dokázaly se vzájemně upřímně a s nadhledem popsat (např. jako příbuzní nebo dobří přátelé). Vzorek zahrnuje  $n(m)=100$  mužů,  $n(f)=100$  žen. Věk účastníků je zachycen ve čtyřech kvótách;  $n(<25)=26$ ,  $n(25-34)=34$ ,  $n(35-55)=67$ ,  $n(>65)=73$ . Rozložení vzhledem k aktuálně dosaženému vzdělání je  $n(ZŠ)=36$ ,  $n(SŠ)=128$ ,  $n(VŠ)=36$ . Vzhledem k uvedeným kategoriím je vzorek srovnatelný s charakteristikami populace České republiky (ČSÚ, 2015).

V rámci této kapitoly budou prezentovány výsledky těchto vybraných testů, resp. škál (jejich kompletní popis viz kap. 3.3.1):

#### **BFI-44: Big Five Inventory** (Pětifaktorový inventář osobnosti)

BFI je pětifaktorovým osobnostním inventářem, dotazníkem měřícím pět obecných dimenzí osobnosti; extraverci (E), neuroticismus (emoční labilita; N), otevřenost ke zkušenosti (O), svědomitost (S) a přívětivost (P).

Jak uvádí Hřebíčková et al. (2016), extraverte (E; extravertní vs. introvertní) je škálou postihující zejména míru společenskosti, energičnosti a optimismu. Neuroticismus (N; neurotický vs. vyrovnaný) popisuje rozdíly mezi lidmi v prožívání negativních emocí (jakými jsou např. strach a rozpaky). Otevřenost ke zkušenosti (O; liberální vs. konzervativní) se váže k zájmu o nové prožitky a dojmy, k upřednostňování rozmanitosti, fantazii, nezávislému úsudku a estetickému vnímání. Svědomitost (S; svědomitý vs. nedbalý) je druhem sebekontroly, který se vztahuje k aktivnímu plánování, organizování a realizaci cílů. Přívětivost (P; přátelský vs. sobecký) je charakterizována laskavým a vlídným chováním k ostatním a ochotou pomoci.

## **SSI: Social Skills Inventory** (Inventář sociálních dovedností)

Dotazník SSI obsahuje šest škál, měřících komunikační dovednosti na dvou úrovních – emoční (neverbální) a sociální (verbální). Vyjadřování (*expressivity*), vnímavost (*sensitivity*) a kontrola (*control*) jsou hodnoceny na obou úrovních: Vyjadřování zahrnuje schopnost jedince sdělovat informace druhým osobám, vnímavost se vztahuje ke schopnosti přijímat a interpretovat informace od druhých a kontrola vyjadřuje schopnost regulovat a řídit komunikační procesy (Riggio & Carney, 2007).

První sledovanou škálou je sociální vyjadřování (SE). Ta poukazuje na verbální komunikační dovednosti člověka v mezilidském kontaktu. Vysoký skór je typický pro zdatné řečníky, často velmi družné a společenské, kteří snadno vedou rozhovor na jakékoli téma a dokážou tímto tématem nadchnout ostatní. Druhou škálou je pak emoční vnímavost (ES), která souvisí se schopností přijímat a interpretovat neverbální sdělení druhých, i v podobě náznaků. Vysoký skór souvisí i s vyšší empatií a citovou ovlivnitelností (*ibid.*).

## **SMS: Self-Monitoring Scale** (Škála sebemonitoringu)

Test je zaměřen na sledování vůle a schopností jedince měnit způsob své prezentace. Škála SMS-SEBO, přecitlivělost na výrazné chování druhých (*sensitivity to expressive behavior of others*), je charakterizována schopností vhodně interpretovat chování ostatních lidí, často intuitivně a na základě nonverbálních vodítek (Snyder, 1974; Lennox & Wolfe, 1984).

## **DBV: Dotazník blízkosti vztahu**, resp. škála **Délka vztahu**

Dotazník DBV byl vyvinut přímo pro potřeby výzkumu CPACT s cílem zjistit míru blízkosti vztahu mezi účastníky výzkumu ve dvojici. V této studii využíváme pouze jednu položku dotazníku, konkrétně Délku vztahu (DV). Ta je definována otázkou „Jak dlouho trvá Váš blízký a osobní vztah s výzkumným partnerem? (uveďte prosím délku blízkého vztahu, nikoliv jen známosti)“ se čtyřmi možnostmi odpovědí: 0 (méně než 2 roky), 1 (2–7 let), 3 (8–17 let) a 4 (více než 17 let).

Z hlediska deskripce 200 odpovědí, se kterými ve studii pracujeme, je zřejmé, že účastníci vykazovali spíše delší, několikaletou délku vztahu ( $M=1,9$ ;  $Mdn=2,5$ ;  $SD = 1,2$ ).

### 3.4.2.3 Výsledky

Získaná data byla popsána pomocí deskriptivních statistik (viz tab. 3.4.2.3.A). Pro sjednocení rozsahů škál byly všechny hodnoty transformovány na rozsah 0–1 dle vztahu  $x_{-01} = (x - \min) / (\max - \min)$ , kde minimum a maximum značí teoretický rozsah škály (vycházející z upravených dat). Z těchto dat pak byla pro porovnání sebesouzení (S) a posouzení druhou osobou (O) pro každou škálu testu vypočtena rozdílová proměnná „škála\_OminusS\_01“ s teoretickým rozsahem -1–1. U všech rozdílových škál byla testována normalita testem Kolmogorov-Smirnov (*Lilliefors Corrected*; K-S), viz tab. 3.4.2.3.B.

Vzhledem k tomu, že data nevykazovala normální rozložení, byl následně použit příslušný jednovýběrový test na hodnotu 0 (*One-Sample Wilcoxon Signed Rank Test*, WSRT). Výsledky prezentují tabulky 3.4.2.3.C a 3.4.2.3.D.

V rámci testování druhé a třetí hypotézy pak byla vytvořena také proměnná nabývající pouze kladných hodnot,  $x_{shoda\_01} = \text{abs}(x_{OminusS\_01})$ , viz tab. 3.4.2.3.E. Vztah mezi mírou shody a proměnnou DBV\_DV (délka vztahu) byl měřen neparametrickým korelačním koeficientem, viz tab. 3.4.2.3.F.

Tabulka 3.4.2.3.A – Deskriptivní statistiky pro všechny sledované škály (verze škál O označena „\_O“) ( $n=200$ )

Škála	N	M	SD	min	max
BFI_E	200	26,13	6,51	9,00	39,00
BFI_N	200	23,46	6,62	8,00	40,00
BFI_S	200	31,40	5,97	14,00	44,00
SSI_ES	200	47,10	7,34	27,00	66,00
SSI_SE	200	42,39	9,23	17,00	67,00

SMS_SEBO	200	20,16	3,85	6,00	30,00
BFI_E_O	200	28,37	6,10	9,00	40,00
BFI_N_O	200	21,91	6,31	8,00	39,00
BFI_S_O	200	33,49	5,66	17,00	44,00
SSI_ES_O	200	47,02	8,26	26,00	67,00
SSI_SE_O	200	44,60	10,22	17,00	71,00
SMS_SEBO_O	200	18,28	4,48	5,00	30,00

*Tabulka 3.4.2.3.B – Deskriptivní statistiky pro rozdílovou škálu OminusS\_01 a výsledky testu normality K-S (signifikance) (n=200)*

Škála	N	M	SD	min	max	sig.
BFI_E_OminusS_01	200	0,07	0,16	-0,59	0,50	0,002
BFI_N_OminusS_01	200	-0,05	0,21	-0,50	0,69	0,041
BFI_S_OminusS_01	200	0,06	0,16	-0,42	0,53	0,027
SSI_ES_OminusS_01	200	0,00	0,15	-0,37	0,35	0,013
SSI_SE_OminusS_01	200	0,04	0,15	-0,37	0,45	0,025
SMS_SEBO_OminusS_01	200	-0,06	0,18	-0,50	0,47	0,200

**První hypotézu**, tj. že hodnocení v charakteristikách, které jsou obtížně pozorovatelné a evaluativní (zde neuroticismus, emoční vnímavost a přecitlivělost na výrazné chování druhých), bude mezi oběma variantami posouzení (sebeuposouzení a posouzení druhým) zřejmě a signifikantní rozdílnost, potvrzujeme částečně.

Tyto rozdíly jsou signifikantní pro škály neuroticismus (BFI\_N; sig. = 0.000) a pro přecitlivělost na výrazné chování druhých (SMS\_SEBO; sig. = 0.000), ovšem nikoliv pro škálu emoční vnímavost (SSI\_ES), kde rozdíl signifikantní není (sig. = 0.858).

Doplňme, že všechny uvedené rozdíly vykazovaly záporné hodnoty. Posuzovatelé (O) tedy hodnotili dané osobnostní charakteristiky jako méně výrazné (přítomné), oproti sebeuposouzení (S).

*Tabulka 3.4.2.3.C – Rozdíl v hodnotách S a O u škál N, ES a SEBO; signifikance dle Wilcoxon Signed Rank Test (n=200)*

<b>Škála</b>	<b>N</b>	<b>M</b>	<b>SD</b>	<b>med</b>	<b>iqr</b>	<b>sig.</b>
BFI_N_OminusS_01	200	0,07	0,16	-0,59	0,50	0,002
SSI_ES_OminusS_01	200	-0,05	0,21	-0,50	0,69	0,041
SMS_SEBO_OminusS_01	200	0,06	0,16	-0,42	0,53	0,027

**Druhou hypotézu**, tj. že u škál s vysokou pozorovatelností a evaluativností (zde sociální vyjadřování, extraverze a svědomitost; SE, E, S) bude míra shody obou variant posouzení vyšší, oproti výše uvedeným škálám neuroticismus, emoční vnímavost a přecitlivělost na výrazné chování druhých (N, ES, SEBO), nepotvrzujeme.

Důvodem je opět výsledek ve škále testu SSI, tentokrát ES, která oproti hypotéze vykazuje nejvyšší míru shody mezi sebesposouzením a posouzením druhým, přestože ji není možné považovat za vysoce evaluativní a pozorovatelnou. Tabulka 3.4.2.3.D ukazuje shody hodnocení (již nezávisle na znaménku), seřazené vzestupně dle průměru. Zde je patrné, že nejvyšší shody i nejnižšího rozptylu zde dosahují obě proměnné SSI.

*Tabulka 3.4.2.3.D – Řazení proměnných dle míry shody (nižší hodnota průměru = vyšší shoda) (n=200)*

<b>Škála</b>	<b>N</b>	<b>M</b>	<b>SD</b>	<b>min</b>	<b>max</b>
SSI_ES_shoda_01	200	0,120	0,087	0,000	0,367
SSI_SE_shoda_01*	200	0,121	0,091	0,000	0,450
BFI_S_shoda_01*	200	0,135	0,107	0,000	0,528
BFI_E_shoda_01*	200	0,140	0,102	0,000	0,594
SMS_SEBO_shoda_01	200	0,148	0,115	0,000	0,500
BFI_N_shoda_01	200	0,171	0,132	0,000	0,688

*\* Zvýraznění škál s vysokou pozorovatelností a evaluativností*

Pro doplnění uvádíme, že z hlediska vyjádření rozdílu mezi sebeposouzením a posouzením druhým byly ve sledovaných škálách SE, E, S získané hodnoty vždy signifikantní (SSI\_SE, sig.=0.000; BFI\_E, sig. = 0.000; BFI\_S, sig.=0,000) (viz tab. 3.4.2.3.E). Zároveň tyto rozdíly vykazují kladné hodnoty, tj. posuzovatelé (O) hodnotili dané osobnostní charakteristiky jako více výrazné (přítomné), oproti sebeposouzení (S).

*Tabulka 3.4.2.3.E – Rozdíly v hodnotách S a O u škál E, S a SE; signifikance dle Wilcoxon Signed Rank Test (n=200)*

<b>Škála</b>	<b>N</b>	<b>M</b>	<b>SD</b>	<b>med</b>	<b>iqr</b>	<b>sig.</b>
BFI_E_OminusS_01	200	0,070	0,158	0,094	0,188	0,000
BFI_S_OminusS_01	200	0,058	0,162	0,056	0,195	0,000
SSI_SE_OminusS_01	200	0,037	0,147	0,050	0,183	0,000

**Třetí hypotézou** je předpoklad, že u posuzovatelů, kteří znají posuzovaného děle, bude vyšší míra shody jejich hodnocení s výsledky sebeposouzení tohoto člověka. Tuto hypotézu nepotvrzujeme.

Korelační koeficienty jsou dle očekávání záporné, tj. dle směru korelace délka vztahu zvyšuje pravděpodobnost shody v posouzení (proměnná „\_shoda\_01“ ukazuje pro hodnoty blízké nule vysokou míru shody), avšak ve všech vztazích nesignifikantní.

*Tabulka 3.4.2.3.F – Korelační vztah mezi délkou vztahu a shodou v posouzení (n=200)*

<b>Proměnná</b>	<b>Rs (rho)</b>	<b>sig.</b>
BFI_E_shoda_01	-0,016	0,822
BFI_N_shoda_01	-0,123	0,084
BFI_S_shoda_01	-0,023	0,743
BFI_O_shoda_01	-0,006	0,933
BFI_P_shoda_01	-0,124	0,081
SSI_ES_shoda_01	-0,106	0,136
SSI_SE_shoda_01	-0,081	0,252
SMS_SEBO_shoda_01	-0,059	0,409

### 3.4.2.4 Diskuse a závěr

---

Představená studie se zaměřila na aplikaci několika základních statistických postupů při zpracování a interpretaci dat ze dvou variant testů: dotazníkového sebeposouzení a posouzení druhou osobou. Výchozí hodnotu nám v tomto směru poskytlo vyjádření shody v posouzení (do jaké míry se sebeposuzovatelé shodují s posuzovateli, kteří je hodnotí ze své perspektivy), tj. hodnoty, která v proměnné „OminusS“ nabývá záporných i kladných hodnot (v případě kladné hodnoty posuzovatelé udávají vyšší skóre v konkrétní škále, oproti sebeposouzení) a v proměnné „X\_shoda\_01“ nabývá kladných hodnot (vyjádření rozdílu bez ohledu na to, která z variant posouzení udává vyšší skóre ve škále). Naším cílem pak bylo ověřit hypotézy týkající se očekávané míry této shody v posouzení konkrétních škál, a to z hlediska jejich pozorovatelnosti a evaluativnosti, ale také z hlediska míry shody v posouzení v souvislosti s deklarovanou délkou vztahu.

Výsledky studie nepotvrdily žádnou z hypotéz, resp. je nepotvrdily v celém rozsahu. Jak v případě první, tak v případě druhé hypotézy byl příčinou výsledek škály testu SSI. U první hypotézy škála ES (Emoční vnímavost) nevykázala dostatečnou signifikanci při vyjádření rozdílnosti obou variant posouzení (S a O), v rámci druhé hypotézy pak stejná škála vykázala nejvyšší shodu v obou variantách posouzení, přestože tato škála nemůže být v daném vymezení považována za vysoce pozorovatelnou a evaluativní. Rozložení ostatních testových škál však už stanovenému předpokladu druhé hypotézy odpovídá (navíc jsou u těchto škál rozdíly mezi hodnocením S a O signifikantní).

Nabízí se několik interpretací zjištěného rozporu. První interpretací může být to, že se jedná o nepřímý důsledek počtu položek v testu – SSI měl totiž v rovině jedné škály největší počet položek (15) oproti ostatním testům (škála BFI zahrnovala 8-10 položek, škála SMS-SEBO měla pouze 6 položek) a projevila se zde tzv. regrese k průměru (Bland & Altman, 1994). Další interpretací může být to, že konstruktové vymezení škály emoční vnímavost (ES) zcela neodpovídá naší interpretaci a škála ve skutečnosti vystihuje odlišný konstrukt, resp. osobnostní charakteristiku. Předpoklad, který studii předcházela, tj. že je ES málo

pozorovatelnou a evaluativní charakteristikou, tak může být lichý. Při rozboru dané eventuality bylo zjištěno, že definice škály z testového manuálu (Riggio & Carney, 2007) našemu předpokladu odpovídá, nicméně některé položky této škály odkazují spíše k manifestnímu projevu než k intrapersonálním procesům, které s nízkou pozorovatelností a evaluativností souvisejí (viz např. položka 56: „Nemám rád, když se mi lidé svěřují se svými problémy“, 62: „Při sledování smutného filmu se někdy rozpláču“, 68: „Snadno dokážu utěšit objetím či pohlazením někoho, kdo je zoufalý“, 74: „Dokážu dlouze pozorovat druhé lidi“, 80: „Často mi někdo řekne, že jsem citlivý a chápající člověk“ a 86: „Když se moji přátelé rozčílí, často mě vyhledají, abych jim pomohl uklidnit se“; *ibid*). Tento postřeh samozřejmě nemusí znamenat nízkou konstruktovou validitu testu SSI; ten je primárně zacílen na interpersonální kontext a v tomto ohledu je manifestní projev poměrně logickou komponentou charakteristiky. Nabízí se ale úvaha, zda volba samotného testu nebyla pro účely studie, resp. z hlediska naší charakterizace škály a práce s ní, spíše nevhodná.

Další interpretací může být také to, že vyšší pozorovatelnost a evaluativnost charakteristiky nemusí automaticky znamenat vyšší míru shody mezi posouzeními. Dobrým příkladem je například rys svědomitosti (z pětifaktorového modelu osobnosti), u kterého často dochází k asymetrii posouzení, i přes relativní transparentnost tohoto rysu, a to i v případě, kdy je posuzovatel rysu dobře obeznámen s posuzovaným (např. jako jemu velmi blízký člověk; viz např. Hřebíčková, 2003). Podobné úvahy podporují i výzkumy zaměřené na hodnocení atraktivity, kdy posuzovatelé obvykle skórují výše než sebe-posuzovatelé (tj. sebe-posuzovatel se vidí jako méně atraktivní; viz např. Vazire, 2006). Z uvedeného vyplývá, že na asymetrii hodnocení se nepochybně podílí velké množství faktorů, které zatím nejsou dostatečně popsány, případně fungují jen v určitých kontextech, nikoliv obecně. Zcela pak pomíjíme otázku, že shoda mezi posuzovateli také vůbec nemusí znamenat skutečnou přítomnost posuzované charakteristiky, resp. že se mohou oba posuzovatelé mýlit, případně že mohou být jejich hodnocení (opět) závislá na určitém kontextu či situaci (viz např. Funder & West, 1993).



Třetí hypotéza, spočívající v předpokladu, že deklarovaná délka vztahu souvisí s mírou shody v posouzení S a O (tj. pozitivně koreluje), rovněž nebyla potvrzena. Příčinou byla nedostatečná signifikance výsledků korelačních koeficientů. K tomuto stavu se rovněž nabízí řada vysvětlení. Prvním, spíše technickým, může být to, že v našem souboru nebyly přítomny osoby, které by deklarovaly jiný než blízký vztah k posuzovanému. Chybělo tedy zastoupení posuzovatelů, kteří by posuzovaného znali jen málo, povrchně a krátce, což mohlo zdeformovat výchozí data. Dalším důvodem, proč nebyla třetí hypotéza potvrzena, mohlo být také to, že oba posuzovatelé (kteří se vzájemně popisovali) deklarovali odlišnou délku vzájemné známosti, tj. neshodovali se v ní (např. jeden udal na škále „méně než 2 roky“ a druhý „2–7 let“), a sledovaná proměnná DV tak byla nesprávně stanovená. V tomto případě však můžeme konstatovat, že shoda v délce vztahu byla víceméně pravidlem (resp. totožnou odpověď poskytlo 87 % procent posuzovatelů). Z hlediska dalších interpretací se nabízí i vysvětlení, které se týká samotných principů posouzení; totiž že délka vztahu je pouze jednou z komponent definujících to, že se dva lidé znají a jsou schopni se komplexně popsat. A jiné komponenty (např. emoční charakter vztahu či jeho intenzita) mohou být v tomto směru lepšími prediktory než prostý čas známosti (viz např. Kenny, 2004).

V navazujících studiích, věnujících se shodě či asymetrii posouzení, by bylo nepochybně vhodné představený výzkumný design obohatit o několik dalších parametrů. Prvním takovým je zvýšit počet respondentů, především pak rozšířit skupinu „vnějších“ hodnotitelů (posuzovatelů) a pracovat s různými úrovněmi vztahů, které k popisované osobě zaujímají (např. životní partner, blízcí přátelé či příbuzní, známí a cizí pozorovatelé). Tento aspekt se vztahuje k tzv. *other-other agreement*, resp. konsenzu posuzovatelů (Funder, 2012). Dalo by se uvažovat také o administraci další varianty dotazníků sebeposouzení, ve které by se (sebe)posuzovatel pokoušel vystihnout své charakteristiky (simulovanou) optikou svého okolí. Jiným perspektivním parametrem by mohlo být vytvoření takové dotaznickové baterie, která by efektivněji pracovala s řazením osobnostních charakteristik dle předpokládané míry jejich intrapersonality a interpersonality, či z hlediska zde pre-

zentované pozorovatelnosti a evaluativnosti. Díky tomu by bylo možné přesněji vystihnout to, které osobnostní charakteristiky jsou v tomto směru spíše konsenzuální a které mají naopak sklon k asymetrickému hodnocení. V souvislosti s vyhodnocováním testů by bylo také možné zjišťovat, zda lidé určitých charakteristik nevykazují specifické tendence v hodnocení – ať už v roli posuzujícího, nebo posuzovaného. Posledním návrhem, který by mohl přinést cenné výsledky, je výzkumné zajištění dodatečných referenčních dat, vůči kterým by se daly výpovědi (jakýchkoliv) respondentů bezprostředně vztahovat – např. objektivizované deskriptory chování posuzovaných osob, které by byly relevantní ve vztahu ke sledovaným charakteristikám (např. dochvilnost či včasné plnění povinností u rysu svědomitosti). Tento model by umožnil triangulaci dat a zjistit, do jaké míry je shoda posouzení relevantní k přítomnosti daného rysu.

### 3.4.2.5 Shrnutí

---

Na základě výsledků této studie můžeme konstatovat následující:

1. V hodnocení osobnostních charakteristik vykazuje škála neuroticismus (test BFI) a přecitlivělost na výrazné chování druhých (test SMS) **signifikantní odlišnost** mezi výsledky sebeposouzení a posouzení druhou osobou.
2. Osobnostní charakteristiky přecitlivělost na výrazné chování druhých a neuroticismus vykazují **nižší míru shody** v posouzeních (tj. vyšší asymetrii) oproti charakteristikám sociální vyjadřování (test SSI), svědomitost a extraverze (test BFI).
3. Osobnostní charakteristika emoční citlivost (test SSI) vyjadřuje **nejvyšší míru shody** mezi oběma variantami posouzení.
4. Nebyl nalezen vztah mezi shodou v posouzeních (sebeuposouzení a posouzení druhou osobou) a deklarovanou délkou vztahu mezi posuzovatelem a posuzovaným.

### 3.4.3

## Rozsah registrové variability textů

*Václav Cvrček, Zuzana Komrsková, David Lukeš*

Cílem této kapitoly je porovnat registrovou variabilitu textů shromážděných v rámci projektu CPACT s reprezentativní populací psaných i mluvených textů v češtině (pozn. kapitola obsahově vychází z připravovaného článku Cvrček et al., k němuž byly doplněny informace o mluvených textech CPACT). Texty sesbírané v projektu CPACT byly popsány v oddílu 3.2 této publikace, přičemž pro účely našeho srovnání je podstatné, že jejich charakteristiku zde uvedenou můžeme označit za „vnětextovou“ – nepopisuje přímo jazykové prostředky, které se v textech hojně užívají, ale je primárně založena na popisu situačního kontextu, v němž byly texty produkovány (scenáře), příp. na základě konvenčně ustálené textové formy (např. dopis či rozhovor). Ačkoli tyto charakteristiky máme konvenčně spojeny s určitými jazykovými prostředky, které se v nich běžně užívají, nejedná se o popis použitých jazykových prostředků jako takových. O takovou deskripci usiluje popis registru, ke kterému chce tato kapitola dospět a který je v opozici vůči předchozímu založen na charakteristikách „vnitrotextových“. Opozice vnitrotextové a vnětextové charakteristiky (někdy také ztotožňována s distinkcí mezi registrem a žánrem; viz Biber, 1995) by se dala přirovnat k rozdílu mezi těstem a formou; složení těsta a jeho výsledná chuť může být nezávislá na tom, do jaké formy ho nalijeme, ačkoli jsou určité konvence, které se (z dobrých důvodů) sluší zachovávat (bábovkové těsto se povětšinou uplatňuje v bábovkové formě atp.). Pokud bychom chtěli zvolit příklad, který je bližší sféře komunikace, můžeme odkázat k všeobecně známým dílům české či světové beletrie, které nikoli ojediněle poskytují doklady rozporu formy a použitých vyjadřovacích prostředků: např. knihu K. Čapka *Válka s mloky* je z hlediska vnětextové charakteristiky možné označit za román, její jazykové zpracování však vykazuje výrazné znaky publicistiky, v důsledku čehož je někdy označována za román-fejeton.

Elicitované texty CPACT budeme porovnávat s korpusem Koditex (Zasina et al., 2017), který byl vytvořen v rámci projektu ČNK za účelem reprezentace variability psaných a mluvených textů. Tento korpus byl využit jako datová základna pro vytvoření multidimenzionálního modelu registrové variability (Cvrček et al., 2018a, 2018b); o podrobnostech složení korpusu pojednává následující oddíl, k analýze a vytvoření modelu viz 3.4.3.2.

### 3.4.3.1 Korpus Koditex

---

Celková velikost korpusu Koditex, který byl sestaven za účelem vytvoření obecného modelu registrové variability českých textů, činí 10,9 mil. pozic (včetně interpunkce), resp. 9 039 137 slov (tokenů bez interpunkce). Korpus je tvořen 3 428 textovými vzorky, v nichž najdeme 719,7 tis. vět. Z hlediska inventáře obsahuje 509,8 tis. různých slovních tvarů a 205,6 tis. lemmat.

Korpus Koditex byl sestavován s primárním zřetelem k tomu, aby pokrýval co nejširší spektrum textů, tj. aby obsahoval texty co nejružnějších charakteristik od co největšího počtu autorů/mluvčích. Korpus netvoří celé texty, ale pouze jejich vzorky v rozsahu 1000 až 5000 tokenů, které jsou na nejvyšší úrovni kategorizace rozřazeny do 3 modů (*mode*) – mluveného, psaného, webového, dělených dále do 8 divizí (*division*) a 45 tříd (*class*). Každou třídu tvoří přibližně 200 000 slov (cca 240 000 tokenů s interpunkcí). Průměrná délka textového vzorku je 2746 slov (kromě interpunkce), směrodatná odchylka činí 749. Detailnější popis korpusu viz <https://wiki.korpus.cz/doku.php/en:cnk:koditex> a v tab. 3.4.3.1.A–C.

Tab. 3.4.3.1.A – Mluvená komunikace (spo) v korpusu Koditex

SPO divize	Třída	Tokeny	Vzorky
<b>int</b> (interaktivní)	<b>bru</b> (nepřipravené veřejné/vysílané rozhovory)	221 812	90
	<b>eli</b> (formální rozhovor)	201 690	82
	<b>inf</b> (neformální rozhovor)	208 565	86
<b>nin</b> (neinteraktivní)	<b>wbs</b> (připravený/čtený projev)	213 201	71

Tab. 3.4.3.1.B – Internetová komunikace (web) v korpusu Koditex

WEB divize	Třída	Tokeny	Vzorky
<b>mul</b> (mnohospměrná komunikace)	<b>dis</b> (internetové diskuse)	197 948	87
	<b>fb</b> (facebookové statusy)	199 418	91
	<b>for</b> (webová fora)	200 104	85
<b>uni</b> (jednosměrná komunikace)	<b>blo</b> (blogy)	204 356	74
	<b>wik</b> (wikipedie)	201 691	84

Tab. 3.4.3.1.C – Psaná komunikace (wri) v korpusu Koditex

WRI divize	Supertřída	Třída	Tokeny	Vz.
<b>fic</b> (beletrie)	<b>nov</b> (romány)	<b>crm</b> (detektivky)	190 026	68
		<b>fan</b> (fantasy)	189 432	69
		<b>gen</b> (bez bližšího určení)	193 667	67
		<b>lov</b> (milostné)	189 893	70
		<b>scf</b> (sci-fi)	188 703	68
		<b>col</b> (povídky)	195 595	70
		<b>scr</b> (scénáře a dramata)	182 689	76
		<b>ver</b> (poezie a písně)	205 837	76

<b>nfc</b> (oborová literatura)	<b>pop</b> (populárně naučná)	<b>fts</b> (formální a tech. vědy)	207 607	68	
		<b>hum</b> (humanitní vědy)	204 837	74	
		<b>nat</b> (přírodní vědy)	204 751	71	
		<b>ssc</b> (společenské vědy)	203 698	68	
	<b>pro</b> (profesní literatura)	<b>fts</b> (formální a tech. vědy)	210 010	71	
		<b>hum</b> (humanitní vědy)	207 916	69	
		<b>nat</b> (přírodní vědy)	209 580	70	
		<b>ssc</b> (společenské vědy)	209 385	72	
	<b>sci</b> (vědecká literatura)	<b>fts</b> (formální a tech. vědy)	202 932	67	
		<b>hum</b> (humanitní vědy)	204 300	71	
		<b>nat</b> (přírodní vědy)	206 716	72	
		<b>ssc</b> (společenské vědy)	205 358	67	
	<b>nmg</b> (noviny a časopisy)	<b>lei</b> (volnočasová publicistika)	<b>adm</b> (administrativa)	203 542	82
			<b>enc</b> (encyklopedie)	203 957	73
			<b>mem</b> (auto-/biografie)	203 390	71
<b>hou</b> (bydlení, zahrada, hobby)			207 499	68	
<b>int</b> (zajímavosti ze světa)			209 232	69	
<b>lif</b> (životní styl)			203 124	72	
<b>mix</b> (víkendové přílohy)			205 310	75	
<b>sct</b> (bulvár)			201 417	73	
<b>new</b> (tradiční publicistika)		<b>spo</b> (sport)	199 238	70	
		<b>com</b> (komentáře)	205 372	68	
		<b>cul</b> (kultura)	205 690	68	
		<b>eco</b> (ekonomika)	211 481	70	
		<b>fre</b> (volnočasové aktivity)	208 532	71	
		<b>pol</b> (politika)	206 893	70	
		<b>rep</b> (reportáže)	206 377	70	
<b>pri</b> (soukromá komunikace)	<b>cor</b> (dopisy)	96 366	68		

### 3.4.3.1.1 Složení korpusu Koditex

---

Při sestavování korpusu byl rovněž brán ohled na složení textů z hlediska časového (korpus obsahuje pouze texty vzniklé po roce 1989, většina ovšem pochází z doby po roce 2000). Prioritu dostaly texty původem české, pouze v třídách, v nichž nebyl k dispozici dostatek originálních textů (milostné romány, detektivky atp.), byly použity vzorky z překladových textů.

Korpus byl kromě metadat obsahujících standardní bibliografickou informaci a žánrovou (vnětextovou) klasifikaci anotován několika způsoby: nástroji pro morfologickou analýzu (Jelínek, 2008; Petkevič, 2014; Spoustová, Hajič, Votrubec, Krbec, & Květoň, 2007; Straková, Straka, & Hajič, 2014), pro rozpoznávání tzv. pojmenovaných entit (*named-entity recognition*, Straková, Straka, & Hajič, 2013), pro označení frazémů (Hnátková, 2002), které umožnily komplexní hledání jazykových rysů v textech. Právě identifikace těchto rysů je v základu multidimenzionální analýzy.

### 3.4.3.2 Multidimenzionální analýza (MDA)

---

MDA představuje kvantitativní přístup k výzkumu jazykové variability, který jako první představil ve svých pracích D. Biber (1988, 1995). Cílem MDA není určení textových prototypů a vymezení jejich charakteristik, ale odhalení dimenzí variability na základě využívání jazykových prostředků (dále rysů), u nichž lze funkční variabilitu předpokládat. Na vymezení každé dimenze se může podílet mnoho rysů z různých rovin (od velmi obecných, např. *syntaktika*, až po specifické, např. *infinitiv končící na -ci*). Konkrétní podoba textu je dána průsečíkem těchto dimenzí. Dimenze pak odráží různé faktory (např. *situční ne/zakotvenost*, *téma, ne/přípravenost*), které se na výsledné podobě textu podílejí.

Postup vytváření MDA modelu je následující: výchozím bodem je sestavení korpusu (v našem případě korpusu Koditex, viz výše) a seznamu rysů, které se na registrové (funkční) variabilitě mohou podílet. Tento seznam byl v případě modelu pro češtinu vytvořen na základě

inspirace Biberovým modelem (Biber, 1988) a na základě excerpcie lingvistické literatury zabývající se stylovým rozrůzněním českých textů či variabilitou obecně (např. Cvrček et al., 2010; Čermák, 2014; Hoffmannová et al., 2016; Karlík et al., 1995; Kodýtek, 2008; Čechová, Krčmová & Minářová, 2008; Miller & Weinert, 1998; Mistrík, 1989; Petr et al., 1986). Shromážděné rysy, jichž bylo pro češtinu nakonec použito 122, je následně třeba operacionalizovat (tj. převést do podoby, která umožní jejich kvantifikaci v jednotlivých vzorcích textů korpusu). Tímto způsobem získáme vstupní data pro statistické vyhodnocení, které má podobu rozsáhlé tabulky, jejíž řádky tvoří jednotlivé vzorky textů v korpusu, ve sloupcích jsou uvedeny hodnoty jednotlivých rysů naměřené v daném textovém vzorku. Hodnotou se ve většině případů myslí relativní frekvence, tj. frekvence vztažená k celkové délce vzorku. V případech, kdy rys není operacionalizován jako frekvence, ale např. jako počet typů (např. počet různých předložek), jsme zvolili specifický způsob relativizace pomocí zTTR (Cvrček & Chlumská, 2015). Ke statistickému vyhodnocení dat se užívá explorační faktorová analýza (Hendl, 2004, s. 505), která umožňuje snížit počet dimenzí v tabulce (počet sloupců, tj. rysů) na několik faktorů (abstraktních numerických deskriptorů nebo také latentních proměnných) při zachování maximální možné variability. Zjednodušením tabulky ze 122 rysů na model s 8 faktory/dimenzemi se stává celý dataset snáze interpretovatelný.

Vícedimenzionální model tedy předpokládá, že texty seskupené do registrů variují podél několika dimenzí, přičemž každá z nich sleduje jiná témata, funkce a situační okolnosti vzniku či cíle textu. Model MDA zahrnuje pro češtinu 8 dimenzí (Cvrček et al., 2018a). Interpretace těchto dimenzí se zakládá na *loadingu* (č. termín zátěž pro jeho omezené rozšíření v této studii nepoužíváme) lingvistických rysů (tj. korelaci mezi frekvencí rysů s dimenzí) a skóru jednotlivých textů, který je umísťuje do prostoru dané dimenze. Jedná se o následující dimenze (pro snadnější orientaci každou dimenzi uvozuje shrnující popis jejích extrémů):

**1. Dynamický (+) vs. statický (-):** První dimenze představuje rozdíl mezi verbálním (dynamickým) a nominálním (statickým) vyjádřením. Dynamický pól tvoří rysy spojené se slovesy: frekvence sloves



jakožto slovního druhu, zejména verba finita v minulém čase, v indikativu obecně, v 1. osobě, slovesa negovaná a vyjadřující mluvení a myšlení, dále frekvence synsémantik, zvláště zájmen pro 1. a 3. osobu i zájmen obecně, spojek a různých tříd adverbii (časová, zájmenná, sloužící subordinaci). Pól statického vyjadřování tvoří substantiva a adjektiva v různých funkcích a podobách (přívlastky, sekvence, různé hodnoty pádu, verbální subst. a adj.), sekundární předložky, dlouhá slova obecně, pasivní tvary sloves a vyšší hodnota tematické koncentrace (Popescu, Best, & Altmann, 2007, s. 68). Texty situované k pozitivnímu pólu nejsou pouze narativní (tj. romány, povídky), ale i reflexivní (např. soukromá korespondence, internetová fóra). U negativního pólu se shlukují informačně nabitě texty z oblasti administrativy, přírodních, technických a formálních věd a encyklopedická hesla.

**2. Spontánní (+) vs. připravený (-):** Spontánnost se jazykově projevuje např. užíváním kontaktních výrazů, výplňkových slov, zájmen, sémanticky vyprázdněných a neurčitých adverbii a nestandardních morfologických tvarů (např. *mejdlo, novej, polívka*). Naopak připravenost je spojena s vyšším výskytem a větší rozmanitostí předložek, vět s interogativními a vztažnými adverbii, lexikální bohatostí (zTTR)<sup>1</sup> množstvím substantiv a obecně delších slov. U pozitivního pólu se vyskytují všechny registry mluveného jazyka (monolog a dialog v soukromém i televizním prostředí). Zajímavé je, že předem sepsané mluvené projevy leží u negativního pólu společně s administrativou, ekonomickými zprávami a např. články z Wikipedie.

**3. Vyšší (+) vs. nižší (-) stupeň koheze:** Třetí dimenze představuje rozdíl v míře výskytu spojovacích prostředků. Pozitivní pól tedy tvoří rysy, které text propojují po formální stránce, např. korelativa (typ *...tím, že...*), vztažné věty se zájmenem *který*, zájmena přívlastňovací a s tím související široký inventář zájmen obecně, dále jmenné přísudky se substantivem (např. *otec byl hrdina*), inventář spojek, komplementace slovesa obsahovými větami (např. *myslel, že...*). Vyšší

---

(1) zTTR je míra inspirovaná z-skórem, která umožňuje srovnávat velikost inventáře typů (např. různých slov) nezávisle na délce textu. zTTR je založeno na porovnání TTR (type-token ratio) s distribucí referenčních hodnot TTR shromážděných ze souboru textů stejné délky (Cvrček & Chlumská, 2015).

kohezi vykazují předem napsané mluvené projevy, vědecké texty z oblasti společenských věd a diskusní pořady, nižší naopak encyklopedie, neformální dialogy, Wikipedie a hobby magazíny.

**4. Polytematický (+) vs. monotematický (-):** Čtvrtá dimenze nejvíce akcentuje rozdíly v bohatosti lexikonu. S pozitivním pólem koreluje širě inventáře lemmat, inventář bigramů, frekvence toponym, inventář předložek a zájmen, zatímco s negativním pólem se pojí tematická koncentrace textu (Popescu et al., 2007), Yulův koeficient (vyjadřující repetitivnost; Oakes, 1998, s. 204), verbální substantiva, pasivum, abstrakta atd. Mezi nejvíce polytematické se řadí novinové či časopisecké rubriky složené zpravidla z krátkých, tematicky různorodých zpráv z kulturního a společenského života. Jako nejvíce monotematické se jeví administrativní a vědecké texty.

**5. Vyšší (+) vs. nižší (-) míra explicitní adresnosti:** Rysy, které nejvíce skórují v páté dimenzi, jsou tyto: otázky, slovesa ve 2. osobě sg. i pl., budoucí čas, vokativ, imperativ a věty s vykřičníkem. Tyto rysy jsou nejvíce zastoupeny v textech napodobujících nebo explicitně uvádějících dialog, tj. např. ve scénářích, poezii, detektivních prózách, příbězích o lásce či v povídkách.

**6. Obecný (+) vs. konkrétní (-):** Kladný pól této dimenze koreluje se sémanticky vyprázdněnými slovy a synsémantiky, zatímco záporný pól tvoří rysy, které mají spojitost s konkrétními údaji (propria, číslovky, časové údaje). Texty u obou pólů jsou na první pohled různorodou směsicí: encyklopedie, básně, volnočasové časopisy se shromažďují u kladného pólu, zatímco sportovní časopisy, scénáře, ekonomické zprávy, bulvár, domácí zpravodajství u pólu záporného. Vztáhneme-li rysy formující záporný pól k textům u něj shromážděným, interpretujeme tuto oblast jako zakotvení do specifického kontextu tvořeného časoprostorem se snahou o aktuálnost. Opačný pól pak vyjadřuje obecnost a nadčasovost.

**7. Prospektivní (+) vs. retrospektivní (-):** Tato dimenze rozděluje texty podle toho, zda se vztahují k budoucnosti a přítomnosti (prospektivní pól), nebo k minulosti (retrospektivní pól). Interpretace se zakládá na korelaci pozitivního pólu s časem budoucím a přítom-

ným oproti negativního pólu s časem minulým. S pozitivním pólem jsou asociovány elicitované dialogy a internetová fóra, s negativním beletrie, v níž se retrospekce odráží v naraci.

**8. Postojovost (+) vs. faktualnost (-):** Pozitivní pól koreluje s rysy a tedy i funkcemi asociovanými zejména s částicemi: č. oslabujícími význam – „downtoners“ (*jen, pouze, spíš, snad, téměř* apod.), zesilujícími význam – „intenzifikátory“ (*ještě, vůbec, samozřejmě, především, dokonce* apod.), omezujícími počet či míru – „restrikto-ry“ (*aspoň, hlavně, jedině, jen* apod.), členícími text (*mj., případně, popř., předně* atp.), dále i s adverbii a spojovacími výrazy. Přítomnost takových rysů indikuje snahu vyjádřit subjektivní názor a evaluaci. U pozitivního pólu se shlukují internetové diskuse a fóra, soukromá korespondence, elicitovaná řeč a blogy. Negativní pól je tvořen pouze jedním rysem – koordinací – a shromažďuje administrativní texty, Wikipedii, scénáře a bulvární noviny.

Pro doplnění je třeba uvést, že počet faktorů/dimenzí v modelu je jedním z jeho vstupních parametrů a představuje zásadní metodologickou otázku, kterou jsme podrobněji rozebrali v předchozí studii (Cvrček et al., 2018a). Pořadí zvolených 8 faktorů je závislé na faktorovací metodě a neodráží poměr vysvětlené variability. Některé dimenze vysvětlují více variability v celkovém souboru, a jsou proto z hlediska popisu modelu důležitější než jiné. V našem modelu hrají nejvýznamnější roli faktory 1 a 2, které pokrývají 21,5 %, resp. 14,2 % variability, zbylé faktory vysvětlují výrazně méně. Celkově model vysvětluje 56 % variability, což je srovnatelné s obdobnými modely v jiných jazycích: např. Biber (1995, s. 121) udává 52 % pro svůj model angličtiny.

### 3.4.3.3 Srovnání modelu MDA s daty CPACT

---

Faktory identifikované pomocí faktorové analýzy vymezují prostor o 8 dimenzích, který vzhledem k reprezentativnosti korpusu Koditex lze považovat za dostatečně obsáhlý na to, aby pokrýval většinu registrované variability češtiny. Takto vzniklý model lze pomocí statistických metod použít na srovnání libovolného textu (nebo korpusu textů) s celou populací a zjistit tak jeho polohu ve vymezeném prostoru. Proces

srovnání dat z projektu CPACT s MDA modelem pro češtinu probíhal v následujících krocích:

1. Identifikace a kvantifikace lingvistických rysů v datech CPACT: použili jsme stejný soubor 122 rysů se stejnou operacionalizací (s výjimkou méně důkladných ručních kontrol) jako pro MDA. Výsledkem je opět tabulka s jednotlivými texty CPACT v řádcích a hodnotami rysů ve sloupcích.
2. Na základě modelu vytvořeného pro Koditex jsme vypočítali faktorové skóry (*factor scores*) pro data CPACT.
3. Interpretace pozic textů CPACT (seskupených podle scénáře) ve všech dimenzích.

Před samotnou interpretací jednotlivých scénářů je nutno uvést ještě dvě metodologické poznámky. První se týká větší disperze textů CPACT v jednotlivých dimenzích. Faktorové skóry pro texty CPACT nejsou přímo srovnatelné s faktorovými skóry pro Koditex, ale dimenze vytvořené na základě korpusu Koditex pro srovnání použity být mohou. Také mohou být prostředkem vzájemného porovnání jednotlivých textů CPACT.

Druhá poznámka se týká problému, že pro výpočet faktorových skóre neexistuje jednoznačný algoritmus a ze své podstaty může tato úloha vést k několika různým, a přesto vyhovujícím řešením (Grice, 2001, s. 431). Obecně není možné říci, zda faktorové skóry v originálním modelu MDA (a stejně tak v datech CPACT) reprezentují nejadekvátnější projekci textů do prostoru definovaného modelem, nicméně hodnoty „indexů neurčitosti“ (Grice, 2001, s. 435), které slouží jako heuristika spolehlivosti projekce, jsou v našem případě příznivé, takže na výsledcích můžeme stavět.

### **3.4.3.4 Srovnání modelu MDA s daty CPACT**

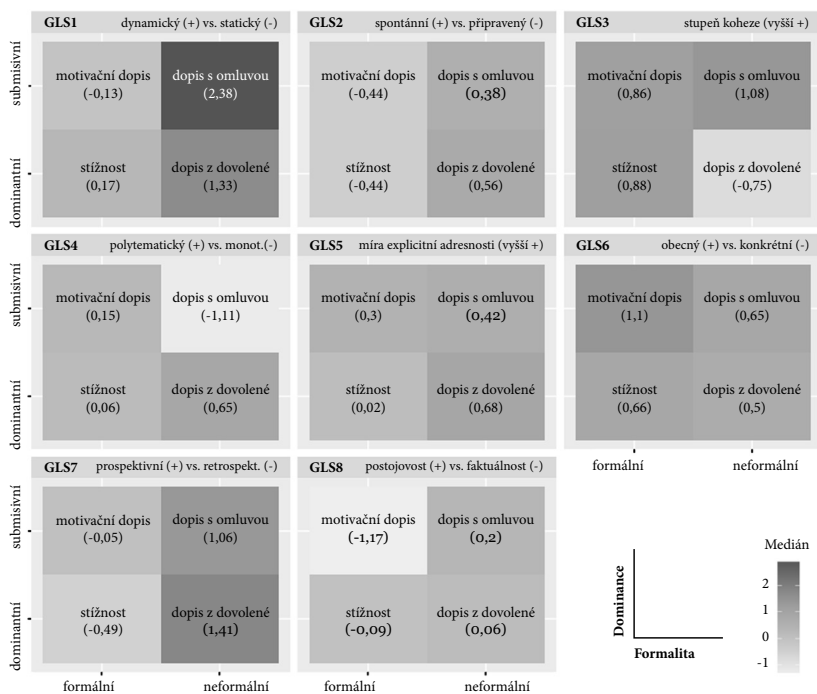
---

Data CPACT představují 4 psané a 2 mluvené útvary, které byly vybrány podle kritéria formálnosti (formální vs. neformální) a pozice pisatele/mluvčího vůči adresátovi (dominantní vs. submisivní). Oba mluvené útvary mají svůj psaný ekvivalent: vyprávění o příjem-

ném zážitku je protějškem dopisu z dovolené, motivační pohovor tvoří pendant motivačního dopisu. Tím se liší od korpusu Koditex, pro nějž byl výběr textů lišících se pouze kanálem (tj. psaným nebo mluveným) jen jedním z kritérií při výběru textů. Další odlišností je skutečnost, že všechny psané texty CPACT spadají na základě vnětextové klasifikace mezi dopisy; psaná část korpusu Koditex je v tomto ohledu bohatší, neboť zahrnuje např. romány, vědecké články, scénáře, básně.

Již na základě výše uvedeného obecného přehledu si lze představit řadu odlišností mezi oběma typy dat. Odlišnost, resp. specifčnost dat CPACT je ostatně prezentována v rovině jednotlivých dimenzí. Jako nejdůležitější se pro srovnání ukázaly dimenze, které podchycují největší množství variability, tj. dimenze 1, 2, 8, a dále dimenze 3 a 7. Naopak dimenze 4, 5 a 6 téměř neodlišují jednotlivé texty CPACT mezi sebou (viz obr. 3.4.3.4). U čtvrté dimenze, která podchycuje rozdíl mezi polytematickými a monotematickými texty, se výrazněji projevuje dopis s omluvou; jeho odlišnost tkví v příklonu k monotematicčnosti, v podobnosti s formalizovanými texty z administrativy či technických věd. Malé rozdíly mezi psanými daty CPACT v páté dimenzi (vyšší vs. nižší míra explicitní adresnosti) jsou dány jednotnou vnětextovou charakteristikou psaných textů. Ve všech dopisech je vždy přítomen adresát, mají tedy zhruba stejnou úroveň dialogičnosti. Odlišnosti mezi nimi jsou patrné pouze v míře explicitnosti tohoto dialogu. Inertnost vůči šesté dimenzi, akcentující obecné vs. konkrétní, souvisí se zadaným scénářem: v dopisech se v souladu s ním objevují popisy konkrétních událostí a míst. Jediný podstatnější rozdíl je u motivačního dopisu, který víc inklinuje k obecnému (ale i v tom je žánrem smíšeným – nižší výskyt propríi apod.).

Obr. 3.4.3.4 – Srovnání psaných textů CPACT v jednotlivých dimenzích (GLS1–GLS8). Intenzita barvy polí označuje medián jednotlivých skupin textů (čím vyšší intenzita, tím vyšší medián).



### 3.4.3.4.1 Motivační dopis

Největší specifická vykazuje tento útvar v osmé dimenzi, která jej určuje jako faktualní. Pisatel se v motivačním dopise zdržuje explicitního sdělování svých postojů, ponechává v textu fakta bez hodnocení (např. *Cvičením se zabývám 6 let, před 4 lety jsem složil zkoušky a stal jsem se certifikovaným instruktorem fitness.* 67R\_TXT1).

Motivační dopis se nejvíce blíží druhému formálnímu psanému textu – stížnosti. Tato blízkost je patrná v první, druhé, třetí a sedmé dimenzi. Oba formální útvary preferují statické vyjádření jisté skuteč-

nosti z minulosti, která je – byť fiktivní – příčinou, proč pisatel tvoří tento text. Také obsahují velmi málo výrazů implikujících interaktivitu (např. vokativ, kontaktné výrazy typu *hele, víš*). Větší péče věnovaná formálnějšímu žánru, která je běžně asociovaná s připraveností, souvisí s vyšší mírou koheze: vztahy mezi výrazy nominální povahy jsou specifikovány pomocí předložek místo prostého kladení slov vedle sebe.

Motivační dopis se od motivačního pohovoru nejvíce liší v prvních dvou dimenzích (viz dále). Tento rozdíl lze chápat jako projev rozdílného komunikačního kanálu: u mluvení není možné změnit již řečené, mluvčí je vystaven časovému tlaku a volbu jazykových prostředků podmiňuje také jeho schopnost adaptability v dané situaci, zatímco u psaní je možné text zpětně editovat a časový tlak není tak velký. V porovnání s korpusem Koditex se nejvíce podobá webovým textům jednosměrného (medialogického) charakteru (WEB-UNI), mezi něž jsou zařazeny blogy a články z Wikipedie.

#### 3.4.3.4.2 Dopis z dovolené

---

MDA ukazuje, že scénář zahrnující kontext dovolené (tj. dopis z dovolené a vyprávění o příjemném zážitku) připouští v porovnání s ostatními scénáři projektu CPACT nejvíce variability. Dopis z dovolené se také nejvíce podobá soukromé korespondenci zahrnuté do korpusu Koditex, ačkoli ta nebyla tvořena v kontrolovaném prostředí a omezena na tematiku dovolené (navíc se často jedná o reakci na dopis současného adresáta).

Od ostatních psaných útvarů CPACT se tento dopis nejvíce odlišuje ve čtvrté dimenzi, která ho charakterizuje jako útvar polytematický. Polytematičnost se projevuje zejména v použité slovní zásobě, jejíž volba závisí na kontextu dovolené představované v dopise (dovolená u moře, na chalupě, v zahraničí apod.). Významně se útvar dále projevuje ve třetí dimenzi, v níž je v porovnání s ostatními nejméně kohezí, tj. využívá úzký repertoár spojek pro různé vztahy (tj. absence sémanticky vyhraněných spojek) nebo věty přičleňuje prostou juxtapozicí. Odlišnost tohoto útvaru se ukazuje i v sedmé dimenzi: dopis z dovolené je silně prospektivní, což souvisí s instrukcí pozvat známého či kamaráda na

místo dovolené, ukázat mu denní program pisatele, do něhož se může zapojit, a činnosti, kterým se adresát může na daném místě věnovat. Persvaze je patrná nejen v budoucím čase (tj. co všechno bude adresát dělat), ale i v přítomnosti imperativu.

Z pohledu první a druhé dimenze (spontánnost) má dopis z dovolené blízko k dopisu s omluvou. Prostřednictvím první dimenze je charakterizován jako dynamický. Pisatelé dávají přednost vyjádření toho, co se na dovolené děje a jakým činnostem se věnují, pomocí sloves (např. *Každý den mám co dělat, většinou jedu na výlet a během dne se vždy někde zastavím na pláži vykoupat.* 82R\_TXT2) namísto jmen. Druhá dimenze jej označuje za spontánní, obsahující množství jazykových prostředků k vyjádření interpersonální interakce.

Dopis z dovolené odlišuje od vyprávění o příjemném zážitku nejvýrazněji druhá dimenze, zaměřená na podchycení rozdílu mezi připravenými a spontánními texty. Tento rozdíl interpretujeme stejně jako u motivačního dopisu vs. motivačního pohovoru.

#### 3.4.3.4.3 Stížnost

---

Stížnost má z ostatních útvarů CPACT nejbližší ke druhému formálnímu útvaru – motivačnímu dopisu, s nímž sdílí charakteristiky zejména ve druhé dimenzi. Oba útvary se jeví jako připravené, užívající repertoár specializovaných předložek a množství vedlejších vět (uvozených adverbii nebo výrazem *jenž*).

Největší odlišnosti od ostatních útvarů nabývá v sedmé dimenzi, ve které se jeví jako nejvíce retrospektivní. Každý pisatel totiž reflektuje skutečnost, na kterou si stěžuje, pomocí minulého času (např. *Soukromě jsem hovořil i s ostatními sousedy a rovněž oni mají výše uvedených skutečností tzv. „plné zuby“, již jsme s manželkou rovněž na nájemníky bytu č. 9 volali Policii ČR, bohužel po slovní domluvě Policie ČR, policie odjela a večírek vesele pokračoval dál.* 5038\_TXT3).

V první a třetí dimenzi nabývá stížnost velké podobnosti s motivačním dopisem, tj. jedná se o útvary zaměřené spíše na statický popis (s využitím nominálních vyjadřovacích prostředků), které vykazují



vyšší stupeň koheze oproti dopisu z dovolené nebo s omluvou. Naopak v osmé dimenzi se stížnost od motivačního dopisu liší nejvíce: nejví se ani jako útvar výrazně postojový, ani faktuální.

Podobně jako motivační dopis má stížnost vzhledem ke korpusu Koditex nejbliže k blogům a článkům z Wikipedie, avšak přímý ekvivalent nemá.

#### 3.4.3.4.4 Dopis s omluvou

---

Tento útvar je oproti ostatním nejvíce vyhraněný v určité dimenzi (viz první dimenze na obr. 3.4.3.4). Ve srovnání s ostatními útvary CPACT je nejdynamičtější, tj. využívá častěji a ve větší míře verbálních rysů (např. minulý čas, verba finita, zájmena 3. osoby). Také při srovnání s korpusem Koditex, v němž je nejdynamičtější beletrie, se dopis s omluvou jeví jako dynamičtější útvar, což svědčí o jeho vyhraněnosti, kterou v beletrii obsahující vždy jak popis (statický prvek), tak vyprávění (dynamický prvek) čekat v takové míře nelze.

Další dimenze, v nichž se dopis s omluvou odlišuje od ostatních, jsou dimenze 4 a 7. Ty klasifikují útvar jako silně monotematický, orientovaný spíše na budoucnost (např. jak se bude pisatel nadále chovat, aby se jednání, za něž se omlouvá, už neopakovalo) nebo popisující děj aktuální, platný v kterékoli časové rovině.

Ve třetí dimenzi se tento útvar podobá stížnosti a motivačnímu dopisu. Snaha o formální provázanost se projevuje častým užíváním korelativ typu *proto*, *že*, vedlejších vět uvozených zájmenem *kerý* a přivlastňovacích zájmen, které explicitně provazují subjekt pisatele s již řečeným. Explicitní uvádění vztahů mezi zmiňovanými subjekty je typické pro texty vykazující prvky stereotypního či formalizovaného vyjadřování.

Ve druhé a páté dimenzi se dopis s omluvou podobá dopisu z dovolené. Oba útvary jsou neformální, což tyto dvě dimenze potvrzují: neformálnost bývá spojena se spontánností vyjadřování, pozorností věnovanou spíše obsahu než formě, což se projevuje v přiblížení dopisu mluvenému vyprávění (zde vyprávění o příjemném zážitku). Zároveň

se jedná o útvary založené na interakci pisatele a adresáta: slouží nejen ke zprostředkování informací, ale i k udržení kontaktu a posílení vztahu mezi účastníky.

#### 3.4.3.4.5 Motivační pohovor

---

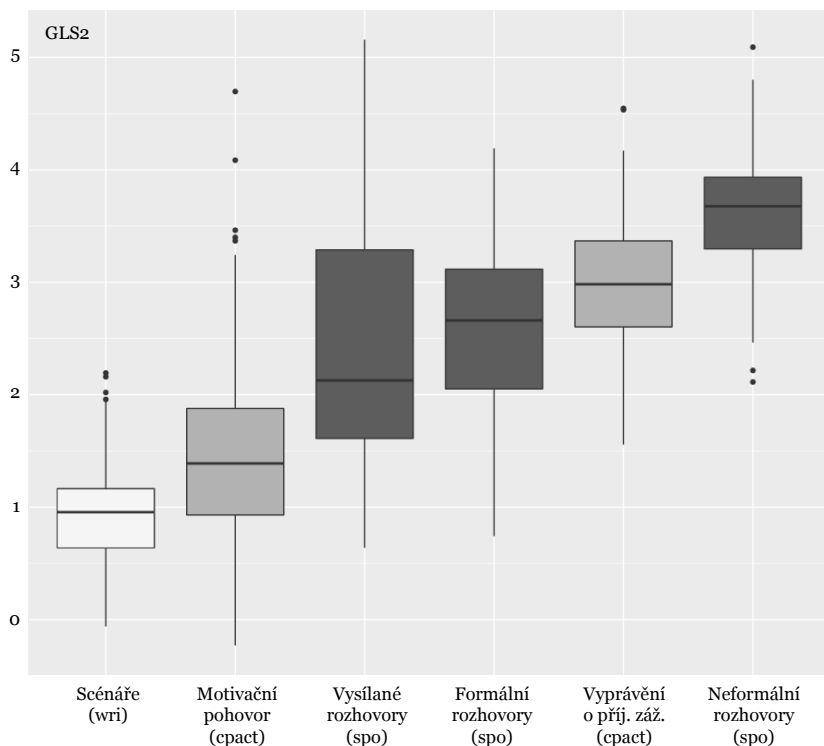
Zástupce mluveného formálního útvaru, ve kterém je zkoumaný mluvčí v submisivní pozici, se nejvíce projevuje v sedmé dimenzi (prospektivní vs. restrospektivní). V motivačním pohovoru se propojuje prospektivita (jaký mluvčí bude jako zaměstnanec) s aktuálností (jaký mluvčí je), např. *Pokud na to budu mít prostor a čas, tak dokážu zvednout výkonnost a jak říkám, motivovat ten tým. To zase jako můžu mluvit z vlastní zkušenosti, že dokážu, že se o to můžu opřít, že dokážu motivovat tým k lepším výsledkům a obrátit lidi, který jsou v nějakým prostě negativním smýšlení k tomu pozitivnímu, aby z té práce měli radost a tak.* (172P\_TXT5).

Také z pohledu druhé dimenze je motivační pohovor specifický útvar. Společně s vyprávěním o příjemném zážitku se vyděluje od psaných útvarů CPACT směrem k pólu spontánnosti, ale na rozdíl od něj zůstává na pomezí mezi mluvenými a psanými útvary (obr. 3.4.3.4.5.A), což je dáno vyšší formálností.

V první a páté dimenzi se motivační rozhovor chová stejně jako vyprávění o příjemném zážitku (viz obr. 3.4.3.4.5.B). Oba mluvené útvary CPACT jsou dynamické: mluvčí volí ke svému vyjádření spíše slovesa než jména, což je tendence, kterou můžeme pozorovat obecně u všech mluvených komunikátů. Příklon k pólu explicitně neadresných textů (monologů) v páté dimenzi znamená, že v motivačním pohovoru spíše nenacházíme prostředky explicitní kontaktovosti či interaktivity (obdobný trend nacházíme i v datech korpusu Koditex: texty s explicitními znaky rozhovoru, např. scénáře a dramata, jsou hodnoceny jako víc dialogické než běžné neformální rozhovory, které takovou potřebu signalizace nemají). Kdo s kým hovoří je dáno a není to potřeba zdůrazňovat např. vokativem či zájmeny pro 2. osobu.

Tento útvar sice je podle scénáře pendantem motivačního dopi-

Obr. 3.4.3.4.5.A – Řazení textů z projektu CPACT a z korpusu Koditex podle druhé dimenze (GLS2; spontánnost) – pouze textové kategorie, které se umístily na pozitivním extrému této dimenze.

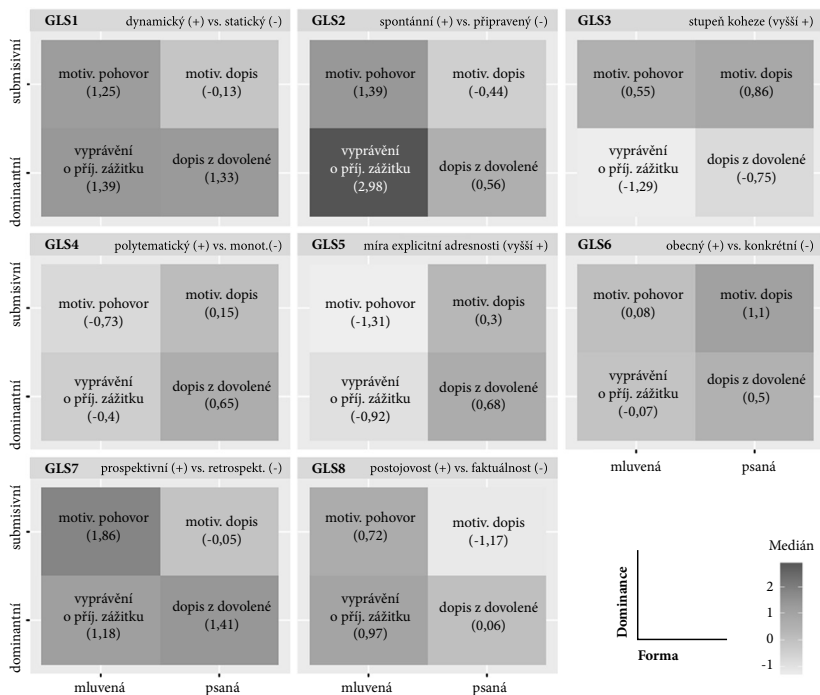


su, ale oba útvary se v jednotlivých dimenzích jeví spíše jako rozdílné (viz obr. 3.4.3.4.5.B). Motivační pohovor také nemá žádný ekvivalent v korpusu Koditex, což potvrzuje i obr. 3.4.3.4.5.A.

#### 3.4.3.4.6 Vyprávění o příjemném zážitku

Podobně jako dopis z dovolené je i vyprávění o ní zástupcem neformálního útvaru, v němž jeho autor vystupuje v dominantní pozici vůči adresátovi. Tyto dva útvary se ve scénáři liší především zprostředkujícím kanálem a u vyprávění vyšší retrospektivitou.

Obr. 3.4.3.4.5.B – Mluvené texty CPACT spolu s jejich tematicky nejblížejšími psanými koreláty v jednotlivých dimenzích (GLS1–GLS8). Intenzita barvy polí označuje medián jednotlivých skupin textů (čím vyšší intenzita, tím vyšší medián).



Nejvýrazněji se vyprávění o příjemném zážitku projevuje ve druhé dimenzi, v níž dochází k výraznému odlišení mluvených a psaných útvarů CPACT. V této dimenzi se vyprávění o příjemném zážitku překrývá s neformálními mluvenými dialogy (SPO-INT) z korpusu Koditex, v němž není tematika rozhovorů nijak regulována. Ačkoli bylo vyprávění o příjemném zážitku nahráváno v kontrolovaném prostředí a mluvčí se předem neznali (navíc si vykali), podařilo se výzkumníkům navodit podobnou atmosféru, jaká panovala při nahrávání v soukromém prostředí mezi komunikanty, kteří se dobře znají. Spontánnost se projevuje zejména v užívání kontaktních prostředků, výplňkových

slov, expresivních částic, dále v častém zmiňování ukazovacích zájmen a adverbii nebo v užívání nestandardních morfologických tvarů (např. *Tam je to, jo, tak tam se, tam je to zase trochu náročnější už, protože tam už jakoby samozřejmě tady byly jako chatky, společenská místnost, že jo všechno jakoby vytápěný, kuchyň zařízená, todle no. 70P\_TXT6*).

Také ve třetí dimenzi se vyprávění o příjemném zážitku výrazně odlišuje od ostatních útvarů CPACT: jedná se o nejméně kohezní útvary. Mluvíci propojují informace juxtaopozičně nebo pomocí spojovacích výrazů s obecným významem (např. spojky *a, když, tak*). Ze psaných útvarů CPACT se mu v tomto nejvíce podobá dopis z dovolené a z korpusu Koditex mluvené neformální dialogy (SPO-INT).

Podle první dimenze se tento útvary zařazuje po bok motivačního pohovoru a dopisu z dovolené mezi útvary dynamické, užívající spíše verbální prostředky místo nominálních (viz obr. 3.4.3.4.5.B). Podobnosti s motivačním dopisem dosahuje vyprávění z dovolené ještě v páté dimenzi, v níž se oba útvary vyskytují u pólu asociovaného s monologičností (viz poznámka k této dimenzi výše).

### 3.4.3.5 Shrnutí

---

Srovnání dat CPACT s multidimenzionálním modelem registrové variability vytvořeným na datech korpusu Koditex, reprezentujícím celou populaci českých psaných i mluvených textů, dává příležitost ke zhodnocení toho, do jaké míry naplnily podmínky sběru (scénáře, prostředí atp.) požadavky tvůrců experimentu. Na základě výše popsané analýzy můžeme konstatovat, že elicitované texty CPACT nabývají očekávaných charakteristik, tedy že se respondenti skutečně snažili ve svých odpovědích přiblížit požadovanému registru. Ačkoli je rozptýl textů CPACT vyšší, což může být způsobeno mj. i jejich průměrně menší délkou, jejich pozice na jednotlivých dimenzích odpovídá očekávaným charakteristikám scénářů.

Odlišnosti mezi jednotlivými scénáři jsou tedy výsledkem různých komunikačních situací, do nichž byli účastníci výzkumu posta-

veni. To dává možnost zkoumat vliv těchto situací na verbální reakce respondentů s ohledem na jejich osobnostní charakteristiku (osoby s rozdílnou osobnostní charakteristikou se vyrovnávají s různými situacemi různě).

Pokud naopak rozdíl mezi scénáři nenacházíme, což je případ pouze formálních psaných scénářů (motivační dopis a stížnost), pak texty za věrohodný zdroj informací o tom, jak se pisatelé s různou osobnostní charakteristikou vypořádávají s různými komunikačními úkoly, považovat nelze. Ať už je registrová shoda mezi texty těchto scénářů artefaktem elicitace nebo jednoduše obecně platnou vlastností těchto dvou (a možná i dalších?) typů formálních dopisů, nesplňují výchozí výzkumný předpoklad, že reprezentují odlišné situační stimuly, a nelze je tedy tímto způsobem spolehlivě využít.

Z těchto důvodů soudíme, že je obecně vhodné elicitované texty validovat porovnáním s obecným modelem registrové variability daného jazyka ještě před započítáním operací směřujících k usouvztažnění lingvistických proměnných s psychologickými.

## 3.4.4

# Odraz osobnosti v textu

## – rysy modelu Big Five

*Jana M. Havigerová, Jiří Haviger, Justýna Franková*

Cílem studie popsané v této kapitole je explorovat souvislosti mezi rysy osobnosti dle modelu Big Five a textovými parametry, tj. morfo-syntaktickými a částečně sémantickými charakteristikami (českého) jazyka dospělé populace. Studie je realizována na klinicky intaktní populaci českých dospělých jedinců, rodilých mluvčích v českém jazyce, tj. souboru P200.

### 3.4.4.1 Teoretická východiska

#### 3.4.4.1.1 Osobnost

---

Osobnost bývá obecně vymezována jako soustava vlastností charakterizujících celistvou individualitu konkrétní osoby. Na osobnost je v psychologii pohlíženo z mnoha různých úhlů pohledu (blíže viz např. Blatný a kol., 2006, Říčan, 2010). Se značnou mírou zjednodušení lze rozlišit dva hlavní proudy v psychologii osobnosti podle cíle, k němuž směřují, a to psychologii osobnosti zaměřenou na studium osobnosti z hlediska **architektury** lidské psychiky (pojetí je analogické zkoumání anatomie lidského těla) směřující k popisu a analýze členění lidské psychiky na jednotlivé složky a jejich vzájemné vztahy a na studium osobnosti z hlediska **individuality** (jedinečnosti, osobitosti) lidské psychiky směřující především k popisu a vysvětlení individuálních rozdílů mezi lidmi (volně dle Říčan, 2010, s. 13-14).

Výzkum CPACT je zaměřen na popis, analýzu a eventuálně predikci vztahu mezi osobností a textem, jinak řečeno cílem výzkumu je odhalování a předpovídání individuálních rozdílů mezi lidmi na základě charakteristik jimi vyprodukovaného textu. Se stanoveným cílem nej-

lépe korespondují koncepce psychologie osobnosti zaměřené na studium osobnosti z hlediska individuality, které poskytují kvalitní podklad k popisu a vzájemnému srovnávání individualit mezi sebou. Většina popisných systémů teorie osobnosti z této kategorie vzešla primárně indukce z empiricky realizovaných výzkumů, zpravidla uskutečňovaných na pozadí extenzivního kvantitativního výzkumu metodou víceškálových dotazníků. Silné metodologické zázemí je tak další z hodnot, které tyto teorie přinášejí pro výzkum, a je dalším motivem pro volbu těchto přístupů k osobnosti jako výchozích pro výzkum CPACT. Ze všech popisných systémů teorie osobnosti jsme pro hlavní studii zvolili pětifaktorový model (*Five factor model*, BFF) tradičně i v českém kontextu označovaný jako Big Five (BF), model Velké pětky (např. Hřebíčková, 2011).

#### 3.4.4.1.2 Pětifaktorový model osobnosti (model Big Five)

---

Pětifaktorový model osobnosti patří mezi vysoce respektované modely osobnosti vycházející z tzv. **rysového přístupu k osobnosti**. Pojem rys je v rámci tohoto přístupu používán pro denotaci určitých vzorců chování, resp. dispozic chovat se určitým způsobem, pro které platí, že jsou (relativně) konzistentní, perzistentní a stabilní v čase, a které lze reprezentovat lineárně (Blatný, 2006, s. 14). Model Big Five vychází z **faktorového přístupu**, který rysy osobnosti odvozuje na základě statistických postupů umožňujících identifikaci struktury dat (původně metodou exploratorní faktorové analýzy, odtud název faktorový přístup) a je orientován ryze **nomoteticky**, tj. popisuje obecnou strukturu rysů (strukturu společnou všem lidem), individualitu konkrétního jedince pak lze popsat kombinací umístění na osách popisujících jednotlivé rysy. Připomeňme, že test vychází z klasické bipolární rysové teorie, která odkazuje na fakt, že rozložení jedinců v populaci v rámci jednotlivých rysů odpovídá Gaussově křivce (vylučují vysoký počet silně vyhraněných osobností v populaci), přičemž u rysu extraverte se nevyhraněným typům dostalo dokonce samostatné označení ambivert, centrovert či omnivert (Löhken, 2014).

Personologický model osobnosti Big Five prezentuje hierarchické



uspořádání rysů do pěti základních dimenzí (hierarchický, protože daná pětice označuje vrchol pyramidy popisu osobnosti). Každá z pěti dimenzí v tomto modelu reprezentuje rys osobnosti, podstatný pro její popis. Pětifaktorový model vypracovávala postupně celá řada psychologů (Tupes, Christal, Digman, Goldberg, Russell, Karol aj.), v našem prostředí je spojován především s autory McCrae a Costa (1984) a jejich NEO dotazníkem a Martinou Hřebíčkovou (autorka českých verzí dotazníků pro měření Big Five rysů osobnosti a výzkumů využívajících tyto dotazníky). V roce 1963 publikoval Norman studii, ve které poprvé použil názvy pěti faktorů, jak je používáme dodnes. Označení Big Five zavedl Lewis Goldberg a bývá též uváděn pod označením OCEAN, které vzniklo jako akronym (počáteční písmena jednotlivých rysů Velké pětky v angličtině).

Nejnámější dotazník autorů Costy a McCrae měřící pět rysů vznikl dvoufázově. Autoři nejprve na základě shlukové analýzy Cattelova 16 PF dotazníku (který měří 12 rysů osobnosti) identifikovali tři shluky primárních faktorů, které označili neuroticismus nebo též emocionální stabilita (*neuroticism*), extravertze nebo též živost (*extraversion*), a otevřenost zkušenosti (*openness to experience*). Ve druhé fázi pak autoři vycházeli z klasického lexikálního přístupu, který předpokládá, že nejzásadnější individuální vlastnosti jsou kódovány ve slovech používaných pro popis individua a vytvořili seznam pojmů popisujících osobnost. Na základě faktorové analýzy získaných popisů doplnili další dva obecné rysy osobnosti – přívětivost (*agreeableness*) a svědomitost (*conscientiousness*). Následně oba výsledky sloučili a vytvořili metodu NEO pro měření těchto pěti dimenzí osobnosti (v českém prostředí například NEO-PI-R, BFI aj.) a zkoumali je v závislosti na věku, kultuře nebo v souvislosti s jinými psychologickými koncepty (Hřebíčková, 2011). Tento trend ve výzkumu osobnosti prostřednictvím dotazníků ze série NEO v současnosti kulminuje, neboť vědecká komunita existenci pětifaktorové struktury osobnosti již plně akceptovala a současné výzkumy se tak odklánějí od zkoumání struktury osobnostních charakteristik právě k ověřování jejich souvislostí s jinými psychologickými či transoborovými souvislostmi (Hřebíčková a kol., 2016). K tomuto cíli směřuje i prezentovaná studie projektu CPACT.

Jednotlivé dimenze osobnosti dle modelu Big Five jsou v této studii pojímány v souladu s popisem uvedeným v manuálu pětifaktorového osobnostního dotazníku (NEO inventář, Hřebíčková, 2004, BFI, Hřebíčková et al., 2016).

**Neuroticismus** (*neuroticism*): neurotický (+) vs. vyrovnaný (-). Škála, která zjišťuje rozdíly mezi lidmi v prožívání negativních emocí jako je strach nebo rozpaky. Vysoké skóre na této škále představuje emocionální labilitu – neuroticismus. S tím se pojí častější prožívání negativních emocí, jako jsou strach a smutek, nejistota, obavy, úzkosti nebo rozpaky. Psychická rovnováha vysoce skórujících osob se snadno naruší a obtížněji se vyrovnávají s každodenními problémy. Naopak osoby skórující nízko jsou klidné a vyrovnané a i stresující situace snášíji podstatně lépe.

**Extraverze** (*extraversion*): extrovertní (+) vs. introvertní (-). Osoby skórující vysoko v dimenzi extraverze jsou zpravidla ti, kteří mají rádi společnost druhých lidí. Projevují se v ní směle a nenuceně. Působí energicky, vesele a optimisticky. Osoby skórující nízko, tedy introverti, jsou ve společnosti zdrženliví, chovají se více nezávisle a samostatně. Nízká extraverze nemusí nutně znamenat, že jsou jedinci sociálně nejistí nebo nepřátelští, ale spíše dávají přednost samotě a nezávislosti.

**Otevřenost ke zkušenosti** (*openness to experience*): liberální (+) vs. konzervativní (-). Škála otevřenosti se váže k zájmu o nové prožitky a dojmy, k upřednostňování rozmanitosti, fantazii, nezávislému úsudku a estetickému vnímání. Osoby skórující vysoko mají živější představivost, jsou ochotnější experimentovat a zkoušet nové způsoby jednání. Jsou také otevřenější k novým myšlenkám a hodnotám a kritickému nahlížení na platné normy. Osoby s nízkým skóre mají spíše tendence chovat se konvenčně a přejímat názory autorit. Před novým a neznámým dávají spíše přednost známému a osvědčenému.

**Přívětivost** (*agreeableness*): přátelský (+) vs. sobecký (-). Osoby skórující vysoko na škále přívětivost jsou v manuálu popisovány jako

lidé chovající se k druhým laskavě a vlídně, a kteří jsou vždy ochotní pomáhat druhým. Dávají přednost spolupráci a ostatním důvěřují. V nežádoucí rovině se mohou ale také chovat servilně a ponížene. Osoby s nízkým skórem neberou zájmy druhých lidí tak na zřetel a mohou se k nim chovat až hrubě nebo neomaleně. Spíše než ke spolupráci mají sklony k soutěživosti.

**Svědomitost** (*conscientiousness*): svědomitý (+) vs. nedbalý (-). Svědomitost je vymezena jako druh sebekontroly vztahující se k aktivnímu plánování, organizování a realizaci cílů. Osoby skórující vysoko jsou cílevědomé, disciplinované, systematické, spolehlivé a vytrvalé. V negativní rovině mohou být také ale pedantské nebo workoholické. Osoby s nízkým skóre projevují menší zaujetí pro plnění cílů. Mohou tak být popisovány jako nestálé, nedbalé a nespolehlivé.

#### 3.4.4.1.4

#### Osobnost a text – studie vycházející z pětifaktorového modelu

---

Vztah mezi lingvistickými parametry textu a Big Five osobnostními rysy byly v minulosti prokázány řadou studií. Jednou z prvních a rozhodně nejčastěji citovanou studií, která studovala lingvistické koreláty osobnosti, je studie autorů Pennebaker a Kinga (1999), kteří už v devadesátých letech minulého století analyzovali eseje 841 studentů texaské univerzity sebrané v letech 1992 – 1997. Faktorovou analýzou autoři extrahovali čtyři faktory (děláním rozdílů, sociální minulost, racionalizace a bezprostřednost), které následně korelovali s dimenzemi osobnosti dle modelu Velké pětky. **Neuroticismus** pozitivně koreloval s faktorem „bezprostřednost“ (charakterizován větším počtem slov 1. os. singuláru, podmiňovacím způsobem a menším počtem členů a slov delších než 6 písmen) a slovy vyjadřujícími negativní emoce, negativně koreloval se slovy vyjadřujícími pozitivní emoce. **Extraverze** pozitivně korelovala s faktorem „sociální reference“ a s výrazy vyjadřujícími pozitivní emoce, negativně korelovala s faktorem „děláním rozdílů“ (charakterizován kategoriemi vylučující slova jako např. *but*, *without*, *except*, váhavá slova jako např. *perhaps*, *maybe*, negace jako např. *not*, *never* a nízkým zastoupením kategorií zahrnující slova typu

*and, with*, příčina jako např. *because, reason*), s výskytem členů a se slovy vyjadřujícími negativní emoce. **Otevřenost**, na rozdíl od neuroticismu, korelovala negativně s bezprostředností (autenticitou), dále pak s užíváním první osoby singuláru, přítomného času a kategorie slov označujících příčinu (*because, reason*), pozitivně s otevřeností koreloval faktor „sociální minulost“ (charakterizován používáním minulého času, přítomného času a sociálních referencí, malým zastoupením pozitivních emocí), výskyt členů, kategorie delších slov, vylučujících slov (*but, without*), váhavých slov a vhléd. **Přívětivost**, stejně jako neuroticismus, pozitivně korelovala s faktorem „bezprostřednost“ (více první osoby singuláru a méně členů), opačně než neurotické pracovaly přívětivé osoby s emočními výrazy – v textech projevovaly více pozitivní emoce, méně negativní emoce. **Svědomitost**, stejně jako extraverteze, negativně korelovala s faktorem „děláním rozdílů“. Svědomitější osoby, shodně s extravertními, užívaly méně vylučujících slov, negace, příčiny a negativně citově zabarvených slov, a naopak více užívaly pozitivně emočních slov. Svědomitost negativně korelovala s výrazy využívajícími podmiňovací způsob (*would, should, could*). Všechny naměřené korelace měly jen malou sílu vztahu, absolutní hodnota  $r$  se pohybovala v rozmezí od 0,07 do 0,16.

Faktorovou analýzu v kombinaci s nástrojem LIWC (*Linguistic Inquiry and Word Count*) použil Pennebaker ve výzkumu realizovaném společně s Chungem (Chung & Pennebaker, 2008) téměř o deset let později při analýze esejí od více než tisíce univerzitních studentů. V esejích měli studenti za úkol popsat sebe sama. Na základě faktorové analýzy použitých přídavných jmen bylo extrahováno 7 faktorů a ty pak byly korelovány se škálami BFI. Skóre na škále neuroticismus pozitivně korelovalo s faktory „hodnocení“ (přídavná jména *ošklivý, tlustý, krásný, atraktivní*) a „negativita“ (*šilený, ublížený, špatný*). Extraverteze korelovala dokonce s pěti ze sedmi extrahovaných faktorů a to pozitivně s faktorem „sebeakceptace“ (*báječný, milující, ztracený*), s faktorem „zapadání“ (v originále *fitting in*, který obsahoval přídavná jména jako: *zajímavý, zábavný, crazy, cool*) a s faktorem „zralost“ (*vyspělý, úspěšný, laskavý*), negativně koreloval s faktory „společenskost“ (*tichý, stydlivý, rezervovaný*) a hodnocení. Přívětivost vykazovala těsný

pozitivní vztah pouze s jedním faktorem „společensko-  
st“. Škála svědomitost korelovala rovněž pouze s jedním faktorem označovaným jako zralost. U osobnostní škály otevřenost ke zkušenosti nebyl zjištěn vztah k žádnému z faktorů získaných na základě adjektiv použitých při sebe-  
popisu.

Kwantes, Derbentseva, Lam, Vartanian a Marmurek (2016) v kanadské studii použili k analýze textů metodu LSA (latentní sémantická analýza), která je založená na faktorové analýze sémantického prostoru. Každý z participujících studentů (N=87) popsal v celkem pěti textech své pocity a chování v situacích podle pěti scénářů, které byly připraveny tak, aby se každý z nich vztahoval k jednomu z pěti rysů Big Five. Pomocí LSA byly vytvořeny rýsové vektory pro každou dimenzi. Pro každý text participantů pak byl na základě podobnosti s každým z pěti rýsových vektorů vyhodnocen vektor textu. U tří dimenzí – **extraverze**, **neuroticismu** a **otevřenosti zkušenosti**, byl zjištěn vztah mezi hodnotou skóre na dané škále BFI a tím, do jaké míry se sémantický obsah jeho textu blížil rýsovému vektoru: čím výraznější byl daný rys u konkrétní osoby, tím více obsahoval narativních pojmů s daným rysem souvisejících.

V korejské studii (Lee, Kim, Seo, & Chung, 2007) byly rovněž analyzovány texty vysokoškolských studentů. Korejští studenti produkovali po dobu 20 minut text pomocí techniky proudu vědomí. Následně byly pomocí korejské verze LIWC zjišťovány vztahy mezi lingvistickými proměnnými, textovými parametry, a úrovněmi dimenzí osobnosti. Vyšší **extraverze** souvisela s vyšším zastoupením sloves, **přívětivost** zase s vyšším zastoupením kategorie „očekávání“ a naopak menším počtem z kategorie „zvolání“. **Svědomitější** jedinci více užívali slov z kategorií „kognitivní slova“ a „sebejistota“ a méně vulgarismů. Jejich texty vykazovaly v průměru menší počet vět. Úroveň **neuroticismu** pozitivně korelovala s počtem slov a negativně s vlastními jmény, emočními slovy a výrazy z kategorie „hněv“. Lidé více **otevření zkušenosti** napsali v průměru větší počet vět, používali více osobních zájmen a méně vlastních jmen, příslovcí a slov v sémantické kategorii „spánek“.

Zatímco Pennebaker a jeho kolegové sbírali texty stovek studentů, kteří nad psaním textů pro analýzu strávili nejméně jednou nejméně dvacet minut psaním, Yarkoni (2010) se rozhodl zanalyzovat texty na osobních stránkách uživatelů, tzv. blogů. Yarkoni se svým designem snažil překonat omezení klasických designů, jakým je vznik textu v laboratorních podmínkách a omezené množství dat – jak počtem osob ve výzkumu, tak množstvím slov od jednoho subjektu, od čehož si sliboval větší ekologickou validitu. Navíc se rozhodl zkoumat osobnostní rysy nejen na úrovni pěti velkých dimenzí osobnosti, ale i jejich dílčích rysů. Kromě množství morfologických a sémantických kategorií navíc zjišťoval i korelace jednotlivých slov (nelze nezmínit slova, která nejsilněji korelovala se sledovanými rysy, neboť dobře ilustrují těsný vztah lexikonu k osobnostním rysům: *neuroticismus + awful*, *extraverze + bar*, *otevřenost + folk*, *přívětivost + wonderful* a *svědomitost + completed*). Výsledné korelace replikovaly mnoho předchozích zjištění o vztazích mezi lingvistickými proměnnými (textovými parametry) a dimenzemi osobnosti (např. poznatky z Pennebaker & King, 1999), čímž byla potvrzena zobecnitelnost těchto vztahů i na on-line prostředí.

Nejen on-line prostředím a přirozeným původem textů se od předchozích liší studie výzkumníků Qiu, Lin, Ramsay a Yang (2012), kteří zkoumali příspěvky na sociální síti Twitter, jejichž délka byla v roce 2012 ještě omezena na pouhých 140 znaků. Analyzovány byly vždy příspěvky za jeden měsíc. Zúčastnění uživatelé pocházeli z několika zemí (USA, Singapur, Velká Británie, Kanada, Indie, Austrálie, Španělsko atd.), byli zahrnuti pouze uživatelé píšící „tweets“ v angličtině (nemusela být jejich rodný jazyk). I v těchto datech z krátkých textů byla potvrzena existence množství vztahů mezi osobností a textem. **Neuroticismus** pozitivně koreloval s 1. os. jednotného čísla, spojkami a negací. **Extraverze** pozitivně korelovala s kategorií pro slova související se sociálními ději (např. *mate*, *talk*, *they*), afektivními procesy (např. *happy*, *cried*) a pozitivními emocemi. Opačné, tedy negativní korelace, byly zjištěny mezi afektivními procesy, pozitivními emocemi a **otevřeností**. **Přívětivost** se vyznačovala menším užíváním negace, slov s vylučujícím významem a slov z oblasti sexuality. Korelační koeficienty okolo absolutních hodnot 0,17–0,28 svědčí pro malou, v ně-

kterých případech střední sílu vztahu. Žádné vztahy nebyly zjištěny mezi lingvistickými proměnnými a **svědomitostí**.

Na krátké texty z přirozeného prostředí se zaměřil také nevadský sociální psycholog Holtgraves (2011). Ten požádal účastníky studie, aby mu vypsali 20 posledních odeslaných zpráv. Holtgraves spočítal korelace jak pro celý soubor, tak zvlášť pro muže a ženy. Ukázalo se, že vztahy mezi osobností a textem jsou v mnoha případech významně genderově podmíněné – pohlaví bylo identifikováno jako významná intervenující proměnná. Například **extraverze** negativně korelovala se slovy vyjadřujícími negativní emoce pouze u mužů, **neuroticismus** se slovy vyjadřujícími negativní emoce zejména u žen, **přívětivost** korelovala s negativními emocemi negativně jak pro celý soubor, tak i u obou pohlaví zvlášť.

Výzkumnou strategií zcela odlišnou od všech dosud zmíněných výzkumů zvolili ruští výzkumníci Stepikhov a Loukina (2014), kteří realizovali klasický experiment. Účastníci studie dostali přepisy spontánních monologů bez interpunkce nebo jiných vodítek, ve kterých měli vyznačit hranice vět. Kromě toho také vyplnili osobnostní dotazníky. Výsledky naznačují, že osoby více **otevřené zkušenosti** preferují delší věty a více **neurotické** osoby zase kratší.

Mnohé další, výše nejmenované studie potvrzují vztahy mezi jednotlivými dimenzemi osobnosti dle modelu Big Five a dílčími jazykovými fenomény. Ve zkratce uveďme ještě některé z nich. Pro rys **neuroticismus** bylo prokázáno, že vyšší míra emoční lability se shoduje s větším využitím singulárních zájmen v první osobě, negativně emočně laděnými slovy a menším používáním pozitivně emočně laděných slov (Gill et al., 2009, Oberlander & Gill, 2006) a ojedinělým používáním zájmen ve třetí osobě (Oberlander & Gill, 2006). Osoby výše skórující na škále neuroticismu používají konkrétnější slova a slova, která mají obecně vyšší frekvenci v korpusu (Gill & Oberlander, 2003). Naopak emočně stabilní jedinci používají bohatší slovní zásobu (Dewaele & Furnham, 1999). Studie Jakovljevové a Milina (2017) realizovaná v srbštině (n = 114 účastníků ve věku 15–65 let), tedy jazyku příbuzného češtině, odhalila shodu mezi negativní valencí a nižší lexikální bohatostí lexikonu u osob s vyšší mírou neuroticismu. Gill a kol. (2006) shledal

významnou pozitivní korelaci mezi rysem **svědomitost** a délkou slov. **Otevřenost** ke zkušenosti je charakterizována preferencí pro delší slova, vyšším užíváním zájmen v první osobě a naopak nižším užíváním zájmen ve druhé osobě (Mehl, Gosling, & Pennebaker, 2006, Qiu, Lin, Ramsey, & Yang, 2012). Lidé výše skórující na škále otevřenosti používají více členů (Mairesse et al., 2007; Mehl, Gossling & Pennebaker, 2006). **Extraverze** se dle výzkumů shoduje s častějším užíváním zájmen (Gill, Nowson, & Oberlander, 2009, Mairesse, Walker, Mehl, & Moore, 2007), méně častým užíváním členů a méně častým užíváním negace (Mairesse a kol., 2007). Texty extravertů vykazují nižší poměr typů ku tokenům (tzv. *type/token ratio*, Mairesse & Walker, 2011), introverti používají bohatší slovní zásobu (Dewaele & Furnham, 1999; Gill & Oberlander, 2002; Scherer, 1979). Rys extraverze vykazuje významně těsný pozitivní vztah s kontextuálností (převaha zájmen, sloves, příslovčí a citoslovčí), negativní s formálností textu (Heylighen & Dewaele, 2002). Pro úplnost doplňme, že se zdá, že čím složitější je úloha a čím vyšší je úroveň úzkosti, tím jednodušší je rozlišení mezi introvertem a extravertem na základě kvantitativních charakteristik jejich textu (Dewaele & Furnham, 1999; Mairesse & Walker, 2010). Iacobelli, Gill, Nowson a Oberlander (2009) ve shodě s dalšími autory dále ukazují, že rysy osobnosti mají vztah nejen k morfologickým a sémantickým jevům, ale též k syntaktickým jevům, konkrétně k užívání interpunkčních znamének (*punctuation marks*).

Ačkoliv jsou dosavadní výzkumy prováděné na různých textech (psané, mluvené, formální, neformální, autentické, podmíněné atp.) a na různých jazykových rovinách, lze shrnout, že studie se shodují v tom, že osobnost měřená prostřednictvím pětifaktorového modelu se prokazatelně promítá do produkovaných textů, přičemž jednotlivé nalézané vztahy mezi osobnostními rysy a dílčími charakteristikami textu vykazují spíše mírně těsné vazby (průkazné korelace jen ojedinele překračují hodnotu  $r = 0,20$ ) a že některé vztahy jsou opakovaně potvrzovány napříč rozličnými výzkumnými designy a jazykovými základnami. Přehled potvrzených vztahů je uveden v tabulce 3.4.4.1.4. Tabulka prezentuje pouze ty lingvistické charakteristiky textů, které v prostudované literatuře vykazovaly korelaci s jednotlivými rysy Big Five na



$p < 0,05$  hladině průkaznosti a které odkazují na takové lingvistické charakteristiky textů, které se vyskytují také v českém jazyce (proto neuvádíme například členy) a které jsou měřeny našimi analytickými lingvistickými nástroji (neměříme například jiné než emoční sémantické kategorie). Protože neexistují analogické studie v českém jazyce (vyjma našich pilotáží), vycházíme výhradně z výzkumů realizovaných na odlišné jazykové bázi. Jsme si vědomi, že mechanické srovnání analogických jazykových jevů jiných jazyků (většina citovaných výzkumů je realizována na textech v angličtině, jen sporadicky v jiných jazycích) s těmiž jevy v češtině není možné, neboť tytéž lingvistické prvky mohou v českém jazyce podléhat rozdílným sémantickým, syntaktickým, morfologickým či pragmatickým pravidlům či odlišným principům, způsobům, zvyklostem a kontextům v jejich užívání a tudíž mohou odkazovat na jiné psycholingvistické souvislosti či znemožnit jejich exploraci (například, jak uvádí Zimová, 1988, přítomnost osobního zájmena jako exoforického identifikátoru v češtině není nezbytná, neboť určitý tvar slovesa v přísudku svou koncovkou gramatickou osobu podmětu dostatečně signalizuje, například věta *Jsem student*; zatímco v angličtině je přítomnost podmětu ve větě nezbytná). Z téhož důvodu nestanovujeme hypotézy; výzkum CPACT je koncipován a prezentován v explorativním modu. Vztahy odhalené na českém jazykovém podkladu a v různých žánrech budeme porovnávat s dosavadním stavem poznání získaným na cizojazyčné bázi. Vztahy nově explorované na českých textech mohou poskytnout bázi pro tvorbu nových hypotéz a pro další zkoumání psycholingvistických souvislostí mezi osobností a textem.

*Tab. 3.4.4.1.4 – Rysy (osobnostní charakteristiky) modelu Big Five\* a jejich lingvistické koreláty v produkovaných textech*

<b>R*</b>	<b>Textová proměnná</b>	<b>Zdroj</b>
	počet slov (+)	Lee, Kim, Seo, & Chung, 2007
	spojky (+)	Qiu, Lin, Ramsay, & Yang, 2012
<b>N</b>	interpunkce (+)	Oberlander & Gill, 2006
	1. osoba singuláru (+)	Pennebaker & King, 1999; Yarkoni, 2010; Qiu et al., 2012; Gill et al., 2009; Oberlander & Gill, 2006

	negace (+)	Yarkoni, 2010; Qiu, Lin, Ramsay & Yang, 2012
	slova s vyšší frekvencí v korpusu (+)	Gill & Oberlander, 2003
	slova vyjadřující negativní emoce (+)	Pennebaker & King, 1999; Yarkoni, 2010; Holtgraves, 2011; Jakovljevoj & Milin, 2017
	počet slov na větu (-)	Stepikhov & Loukina, 2014
<b>N</b>	podstatná jména (-)	Oberlander & Gill, 2006
	příslovce (-)	Oberlander & Gill, 2006
	předložky (-)	Holtgraves, 2011
	zájmena ve 3. osobě (-)	Oberlander & Gill, 2006
	2. osoba (-)	Yarkoni, 2010
	bohatství slovní zásoby (-)	Dewaele & Furnham, 1999; Jakovljevoj & Milin, 2017
	slova vyjadřující pozitivní emoce (-)	Pennebaker & King, 1999; Gill et al., 2009; Oberlander & Gill, 2006
	slova vyjadřující emoce (-)	Lee a kol., 2007; Gill a kol., 2009; Oberlander & Gill, 2006
	zahrnující, spojující výrazy (+)	Pennebaker & King, 1999
	<hr/>	
	slovesa (+)	Lee, Kim, Seo, & Chung, 2007
	přídavná jména (+)	Oberlander & Gill, 2006
	spojky (+)	Oberlander & Gill, 2006
	osobní zájmena (+)	Holtgraves, 2011; Gill, Nowson & Oberlander, 2009, Mairesse a kol., 2007
<b>E</b>	kontextuálnost (součet zájmen, sloves, příslovčí a citoslovčí) (+)	Heylighen & Dewaele, 2002
	slova vyjadřující pozitivní emoce (+)	Pennebaker & King, 1999; Qiu a kol., 2012; Oberlander & Gill, 2006
	1. osoba singuláru (+)	Holtgraves, 2011
	1. osoba plurálu (+)	Yarkoni, 2010
	2. osoba (+)	Yarkoni, 2010

	poměr typů ku tokenům (-)	Mairesse & Walker, 2011
	bohatství slovní zásoby (-)	Dewaele & Furnham, 1999; Gill & Oberlander, 2002; Scherer, 1979
<b>E</b>	číslovky (-)	Yarkoni, 2010
	negace (-)	Pennebaker & King, 1999; Mairesse a kol., 2007
	slova vyjadřující negativní emoce (-)	Pennebaker & King, 1999; Holtgraves, 2011
	délka slov (+)	Pennebaker & King, 1999; Gill a kol., 2006
	počet slov na text (+)	Lee, Kim, Seo, & Chung, 2007
	počet slov na větu (+)	Stepikhov & Loukina, 2014
	předložky (+)	Yarkoni, 2010
	osobní zájmena (+)	Lee, Kim, Seo, & Chung, 2007
	osobní zájmena v 1. osobě (+)	Mehl, Gosling, & Pennebaker, 2006; Qiu, Lin, Ramsey, & Yang, 2012
	osobní zájmena ve 2. osobě (-)	Mehl, Gosling, & Pennebaker, 2006; Qiu, Lin, Ramsey, & Yang, 2012
<b>O</b>	1. osoba singuláru (-)	Pennebaker & King, 1999; Yarkoni, 2010;
	2. osoba (-)	Yarkoni, 2010
	přítomný čas (-)	Pennebaker & King, 1999; Yarkoni, 2010
	minulý čas (-)	Yarkoni, 2010
	zájmena (-)	Yarkoni, 2010
	příslovce (-)	Lee, Kim, Seo, & Chung, 2007
	negace (-)	Yarkoni, 2010; Iacobelli a kol., 2009
	slova vyjadřující negativní emoce (-)	Yarkoni, 2010
	1. osoba singuláru (+)	Pennebaker & King, 1999
<b>P</b>	1. osoba plurálu (+)	Yarkoni, 2010

	slova vyjadřující pozitivní emoce (+)	Pennebaker & King, 1999
<b>P</b>	zájmena (+)	Yarkoni, 2010
	slova vyjadřující negativní emoce (-)	Pennebaker & King, 1999; Holtgraves, 2011
	negace (-)	Qiu et al., 2012
	délka slov (+)	Gill a kol., 2006
	slova vyjadřující pozitivní emoce (+)	Pennebaker & King, 1999
	slova vyjadřující negativní emoce (-)	Pennebaker & King, 1999; Holtgraves, 2011
<b>S</b>	počet vět (-)	Lee, Kim, Seo, & Chung, 2007
	negace (-)	Pennebaker & King, 1999; Yarkoni, 2010
	podmiňovací způsob (-)	Pennebaker & King, 1999
	vulgarismy (-)	Lee, Kim, Seo, & Chung, 2007

\* *N* = *neuroticismus* (*neuroticism*), *E* = *extraverze* (*extraversion*), *O* = *otevřenost ke zkušenosti* (*openness to experience*), *P* = *přívětivost* (*agreeableness*), *S* = *svědomitost* (*conscientiousness*)

### 3.4.4.2 Praktická část studie

#### 3.4.4.2.1 Výzkumný soubor a výzkumná data

Výzkumný soubor: P200, který sestává z 200 respondentů; metoda získání výzkumného souboru: kvótní výběr; kritéria pro zařazení: psychiatricky intaktní dospělá populace rodilých mluvčích v českém jazyce (blíže viz kap. 3.2.1.1).

Lingvistická data: analyzováno bylo celkem  $N_{\text{txt}} = 1200$  textů, každý z 200 respondentů poskytl 4 psané a 2 mluvené texty produkováné na základě konkrétních scénářů v kontrolovaném prostředí (viz kap. 3.2.1.4). Každý jednotlivý text je popsán prostřednictvím kvantifikovatelných lingvistických proměnných – textových parametrů; do analýz vstupovaly parametry popsané v kapitole 3.3.2, celkem

120 proměnných pro každý text. Výpočty jsou provedeny a výsledky prezentovány pro každý text (scénář) zvlášť.

Psychologická data: výsledky získané metodou sebeposouzení v dotazníku BFI, analyzovány jsou výsledné skóry, kterých respondenti dosáhli v pěti škálách dotazníku BFI (blíže viz kap. 3.3.1).

#### 3.4.4.2.2 Postup zpracování dat a získané výsledky

Vztahy mezi psychologickými rysy a textovými parametry byly testovány výpočtem korelačních koeficientů: Spearmanovo  $\rho$ . Pro každý text bylo vypočteno 600 korelací (5 škál reprezentujících osobnostní rysy krát 120 textových parametrů), tj. celkem bylo testováno 3600 korelačních vztahů. Statisticky průkaznou těsnost vztahu (tj. takové korelace, které dosahovaly před FDR korekcí hodnoty  $p < 0,05$ ) vykazovalo průměrně 8,9 procenta ze všech testovaných vztahů. Aby byla minimalizována chyba I. druhu, byla provedena Benjaminiova-Hochbergova (BH) korekce p-hodnot statisticky průkazných výsledků (Benjamini & Hochberg, 1995). Připomeňme, že cílem BH korekce je dosáhnout stavu, kdy pravděpodobnost, že mezi prezentovanými výsledky je falešně pozitivní výsledek je menší nebo rovna 5 procentům (blíže viz kapitola 3.4.1).

Statisticky průkaznou těsnost vztahu po BH korekci dosáhlo v průměru 1 % ze všech testovaných vztahů. Počet (absolutní i relativní četnost) statisticky průkazných vztahů před a po BH korekci je uveden v tabulce 3.4.4.2.2.A.

*Tab. 3.4.4.2.2.A – Statisticky průkazné korelace před a po Benjamini-Hochberg (BH) FDR korekcí: absolutní a relativní četnosti*

Text	Počet statisticky průkazných vztahů před BH korekcí		Počet statisticky průkazných vztahů po BH korekci		Počet testovaných korelačních vztahů
	n ( $p < 0,05$ )		n ( $p_{BH} < 0,05$ )		
TXT1	31	5,2 %	0	0,0 %	600
TXT2	46	7,7 %	4	0,7 %	600

TXT3	43	7,2 %	0	0,0 %	600
TXT4	64	10,7 %	11	1,8 %	600
TXT5	76	12,7 %	10	1,7 %	600
TXT6	62	10,3 %	10	1,7 %	600
<b>total</b>	<b>322</b>	<b>8,9 %</b>	<b>35</b>	<b>1,0 %</b>	<b>3600</b>

Konkrétní vztahy, jejichž p-hodnoty dosahovaly i po BH korekci požadovaných hodnot, jsou prezentovány v následujících tabulkách 3.4.4.2.2.B–E pro jednotlivé osobnostní rysy. Uvedeny jsou průkazné vztahy mezi rysy osobnosti a jednotlivými lingvistickými charakteristikami (relativizovanými vůči počtu slov či danému lingvistickému jevu), resp. lingvistickými indexy, které kombinují více lingvistických jevů.

## Extraverze

Pro osobnostní rys extraverte byly detekovány čtyři signifikantní vztahy. Všechny vztahy byly shledány pouze u mluvených textů. U páteho textu (motivační pohovor) výsledky ukazují na těsný vztah rysu extraverte se zvýšeným podílem zájmen (ku počtu všech slov, ku počtu podstatných jmen, například *já, ty, námi*), ať už bez specifikace (u pracovního pohovoru), tak specifikovaných na osobní zájmena v 1. osobě singuláru (například *já, mi, mojí* u rozhovoru o příjemném zážitku). U rozhovoru o příjemném zážitku se se dále projevil těsný vztah extraverte s výskytem přičestí minulého (aktivní participium tvořené koncovkami *-l, -la, -lo, -li, -ly, -la*, včetně tvarů s přívěsným *-s*), které znamená obvykle výsledek ukončeného děje (např. *Přišel jsem*). Vztah je ve všech případech pozitivní – **čím více extraverte, tím větší podíl zájmen či přičestí minulého v mluvené řeči.**

Tab. 3.4.4.2.2.B – Vztah textových parametrů a rysu extraverte po provedení BH korekce

Text	Text. parametr	$\rho$	$p_{BH}$	Popis text. parametru
TXT5	POSP_per_W	0,248	0,047	počet zájmen / počet slov
TXT5	Pronom_index	0,243	0,047	počet zájmen / počet podstatných jmen

TXT6	PO1_per_W	0,235	0,045	počet osobních zájmen v 1. osobě / počet slov
TXT6	SbPp_A_perW	0,275	0,006	počet sloves minulého činného příčestí / počet slov

## Neuroticismus

Pro osobnostní rys neuroticismus bylo identifikováno pět signifikantních vztahů. Čtyři vztahy odkazují s rostoucím neuroticismem na zvýšenou expozici osobních zájmen v 1. osobě (obecně, např. *já, my, námi* a v singuláru, např. *já, mě, mi, mou*) a vysoký poměr sloves ku počtu přídavných jmen (tzv. *trager index*) a naopak snížený počet neurčitých číslovek (např. *mnoho, málo, tolik, několik, kdovíkolik*) v Dopise s omluvou. V motivačním pohovoru pak promluvy vykazovaly těsný negativní vztah mezi mírou neuroticismu a bohatostí slovní zásoby (čím vyšší neuroticismus, tím nižší poměr lemmat ku odmocnině z počtu slov, neboli tím menší slovní zásoba). Zobecněně řečeno, **čím vyšší neuroticismus, tím více zájmen v 1. osobě, sloves, méně neurčitých číslovek a menší bohatost slovní zásoby.**

Tab. 3.4.4.2.2.C – Vztah textových parametrů a rysu neuroticismus po provedení BH korekce

Text	Text. parametr	$\rho$	$p_{BH}$	Popis text. parametru
TXT4	PO1_per_W	0,258	0,033	počet osobních zájmen v 1. osobě / počet slov
TXT4	PO1S_per_W	0,250	0,040	počet os. zájmen v 1. osobě singuláru / počet slov
TXT4	Trager_index	0,240	0,043	počet sloves / počet přídavných jmen
TXT4	SbPa_A_per_W	-0,23	0,046	počet neurčitých číslovek / počet slov
TXT5	Lem_per_Sqrt_W	-0,241	0,047	počet lemmat / odmocnina z počtu slov

## Otevřenost zkušenosti

Pro osobnostní rys otevřenost bylo identifikováno osm signifikantních vztahů. Všechny vztahy byly shledány pouze u mluvených textů. U obou mluvených textů jsou přitom identifikovány prakticky totožné vztahy: **s rostoucí otevřeností roste bohatost slovní zásoby, klesá množství interpunkce** (v poměru k tokenům) **a počet emočních slov a frazémů, u kterých nelze zjistit emoční příznak** (protože jsou neutrální nebo umožňují oba póly). U motivačního pohovoru s rostoucí otevřeností klesá výskyt negativních slov (všech slovních druhů, sloves i sloves v 1. osobě, např. *nechci, neděláme*).

Tab. 3.4.4.2.2.D – Vztah textových parametrů a rysu otevřenost ke zkušenosti po provedení BH korekce

Text	Text. parametr	$\rho$	$p_{BH}$	Popis text. parametru
TXT5	Lem_per_Sqrt_W	0,342	0,000	počet lemmat / odmocnina z počtu slov
TXT5	POSZ_per_Tok	-0,238	0,047	počet interpunkce / počet tokenů
TXT5	Negativity_index	-0,234	0,047	počet negativních / (negativních + afirmativních)
TXT5	Vneg_per_W	-0,238	0,047	počet negativních sloves / počet slov
TXT5	Vneg1S_per_W	-0,248	0,047	počet negativních sloves v 1. osobě / počet slov
TXT5	Em2_per_W	-0,234	0,047	počet emočních slov (širší výběr) – neurčeno / počet slov
TXT5	PhE_per_W	-0,241	0,047	emoční náboj frazému – neurčeno / počet slov
TXT6	Lem_per_Sqrt_W	0,245	0,033	bohatost slovní zásoby: počet lemmat / odmocnina z počtu slov
TXT6	POSZ_per_Tok	-0,300	0,002	počet interpunkce / počet tokenů
TXT6	Emo_per_W	-0,302	0,002	počet emočních slov – neurčeno / počet slov



TXT6	Em2_per_W	-0,314	0,002	počet emočních slov (širší výběr) – neurčeno / počet slov
TXT6	PhE_per_W	-0,302	0,002	počet frazémů (neurčený emoční náboj) / počet slov

## Svědomitost

Pro osobnostní rys svědomitost bylo identifikováno třináct signifikantních vztahů. Všechny vztahy byly shledány pouze u neformálních textů, jak psaných (dopis z dovolené, omluva), tak mluvených (rozhovor o příjemném zážitku). Psaný text dopis z dovolené odkazuje na **těsný záporný vztah mezi rysem svědomitost a počtem hovorových tvarů a výrazů obecné češtiny** (s rostoucí svědomitostí klesá počet výrazů s koncovkami obecné češtiny, např. *dobrej*, a obecných výrazů, např. *dovča*), se svědomitostí **klesá počet výrazů v negativním tvaru** (např. *nedělej*, *nechutný*). V dopise s omluvou byly detekovány kladné korelační vztahy, které vedou k závěru, že s rostoucí svědomitostí **roste počet podstatných jmen, přídavných jmen, neurčitých číslovek** (např. *mnoho*, *málo*, *tolik*, *kdovíkolik*) a **přivlastňovacích zájmen** (např. *můj*, *mého*, *naši*) v poměru k počtu slov a klesá poměr sloves k přídavným a podstatným jménům, neboli **klesá aktivita a roste deskriptivita textu** (což potvrzuje i průkazný vztah svědomitosti s dimenzí vyprávění–popis). V rozhovoru o příjemném zážitku se ukázal jako **průkazně těsný vztah mezi svědomitostí a výskytem 2. osoby** (u všech slov, u sloves i v poměru k ostatním osobám, např. *děláš*, *prosíte*).

Tab. 3.4.4.2.2.E – Vztah textových parametrů a rysu svědomitost po provedení BH korekce

Text	Text. parametr	$\rho$	$p_{BH}$	Popis text. parametru
TXT2	Var6_per_W	-0,293	0,007	počet hovorových tvarů / počet slov
TXT2	ClqC_per_W	-0,252	0,050	počet obecně českých výrazů / počet slov

TXT2	Negativity_index	-0,251	0,050	počet negativních / (negativních + afirmativních)
TXT4	POSN_per_W	0,234	0,046	počet podstatných jmen / počet slov
TXT4	POSA_per_W	0,310	0,002	počet přídavných jmen / počet slov
TXT4	SbPa_A_per_W	0,314	0,002	počet neurčitých číslovek / počet slov
TXT4	PP1_per_W	0,248	0,040	počet přívlastňovacích zájmen / počet slov
TXT4	Trager_index	-0,308	0,002	počet sloves / počet přídavných jmen
TXT4	Readines_to_action	-0,241	0,043	počet sloves / počet podstatných jmen
TXT4	GLS1	-0,239	0,043	dimenze vyprávění – popis
TXT6	@2P_per_W	0,283	0,005	počet slov v 2. osobě plurálu / počet slov
TXT6	Per2_index	0,277	0,006	počet slov ve 2. osobě / (slova v 1. + 2. + 3. osobě)
TXT6	V2_per_W	0,237	0,045	počet sloves ve 2. osobě / počet slov

## Přívětivost

Pro rys přívětivost bylo nalezeno 40 signifikantních individuálních vztahů, po Benjamini-Hochbergově korekci však **žádný z nalezených vztahů nedosáhl adjustované hodnoty  $p_{BH} < 0,05$ .**

### 3.4.4.3 Diskuse a závěr

#### 3.4.4.3.1 Jednotlivé osobnostní rysy

Celkem bylo detekováno 35 spolehlivých korelací (po BH korekci, tj. s akceptovaným rizikem, že nejvýše 5 % z 35 detekovaných korelací, tedy jedna až dvě mohou být falešné objevy) mezi lingvistickými charakteristikami textu a rysy osobnosti dle modelu Big Five. Většina

detekovaných vztahů je v souladu s nálezy ostatních autorů, kteří realizovali výzkum na jiných jazycích. Objevily se i některé nové vztahy, bude otázkou dalších výzkumů porovnat je s výskytem u jiných jazyků.

**Extraverze** se v textech českých mluvčích vyskytovala současně s výskytem osobních zájmen, stejně jako u cizojazyčných studií (Holtgraves, 2011; Gill, Nowson & Oberlander, 2009, Mairesse a kol., 2007), extraverti používají více osobních zájmen v rozhovorech, přičemž v rozhovoru o příjemném zážitku intenzivněji osobní zájmena v 1. osobě singuláru, což koresponduje s výsledky Holtgraves (2011). Zvýšený výskyt minulého aktivního přičestí v rozhovorech extravertů o příjemném zážitku v našem výzkumném souboru nemá v mezinárodních výzkumech obdobu. Je možné, že tento vztah byl nalezen díky specifiku nastaveného komunikačního rámce (rozhovor o ideální dovolené žádný jiný výzkumník nepoužil). Můžeme spekulovat o tom, že extraverti odkazovali na minulou zkušenost, protože jejich společenská a optimistická povaha umožňuje zažívat anebo reprezentovat minulé zkušenosti jako příjemné, zatímco introverti o ideální dovolené zatím jen sní.

**Neuroticismus** se v českých textech projevil nejsilněji v dopisech s omluvou zvýšenou produkcí osobních zájmen v 1. osobě, a to především v singuláru. Právě 1. osoba singuláru se zvýšeně objevovala v textech neurotiků i podle celé řady zahraničních studií (Pennebaker & King, 1999; Yarkoni, 2010; Qiu et al., 2012; Gill et al., 2009; Oberlander & Gill, 2006). V pozadí takového vztahu patrně stojí univerzální transjazykový mechanismus. Možná interpretace je, že tendence hovořit o sobě (1. osoba singuláru) se odráží tendence neurotiků zvýšeně se zabývat vlastní osobou v kontextu narušeného sebepojetí (těsný vztah mezi neuroticismem a narušeným sebepojetím byl opakovaně prokázán, např. Oosterwegel & Wicklund, 2012; Lodi-Smith & DeMarree, 2018). Stejně jako bylo zjištěno v zahraničních studiích (Dewaele & Furnham, 1999; Jakovljevo & Milin, 2017), také v českém jazyce jsme shledali statisticky průkazný vztah mezi neurotismem a bohatstvím slovní zásoby – čím větší míra neuroticismu, tím chudší slovní zásoba v motivačním pohovoru, tedy v řečové produkci ve stresové situaci. Zjištění reflektuje v současnosti již dobře známý mechanismus (např. Hampson, 2014) vlivu neuroticismu na stres a copingové me-

chanismy zvládnání (neuroticismus působí jako mediátor vztahu mezi stresem a mechanismem zvládnání, u neurotiků zpravidla převažuje zvládnání zaměřené na emoce). Zvýšené prožívání negativních emocí charakteristické pro rys neuroticismus by se mohlo projevovat vyšším poměrem použitých slov s negativním emočním nábojem, jako se ukázalo v mnoha zahraničních studiích (Pennebaker & King, 1999; Yarkoni, 2010; Holtgraves, 2011; Jakovljevov a Milin, 2017), v naší studii se však takový vztah jako statisticky průkazný neukázal, a to ani před korekcí. V naší studii se navíc objevil vztah negativní korelace mezi neuroticismem a výskytem neurčitých číslovek v dopise s omluvou.

**Otevřenost zkušenosti** se v textech českých mluvčích vyskytovala v souvislosti s bohatostí slovní zásoby (pozitivní korelace), množstvím interpunkce (negativní korelace) a počtem emočních slov a frazémů, u kterých nelze zjistit emoční náboj (protože jsou neutrální nebo umožňují oba póly). Podobná zjištění nemají v zahraničních výzkumech obdobu. U motivačního pohovoru bylo zjištěno, že s rostoucí otevřeností klesá výskyt negace, což koresponduje se studii Yarkoniho (2010, studie blogerů) a Iacobelliho (2009, studie mluvené i psané řeči), kteří našli totožný vztah. V naší studii se ukázalo, že tento vztah je nejsilněji syčen výskytem negace u sloves v 1. osobě singuláru, tj. otevření jedinci používají k popisu vlastních aktů afirmativní varianty sloves, zatímco uzavření jedinci s nízkým skóre využívají více negativní varianty sloves (*nechci, nepůjdu, nemohl jsem* apod.). Nízkému skóre v dimenzi otevřenost odpovídají jedinci, kteří bývají popisováni jako lidé konvenční, přízemní a v postojích konzervativní, s úzkými zájmy, preferující známé a osvědčené, kteří mívají utlumené emoční reakce). Je možné, že zvýšený výskyt sloves v negativní formě je lingvistický projev zvýšené seberegulace uzavřených jedinců – uzavření se do sebe je rovněž udáváno jako strategie zvládnání, která silně koreluje s nízkým skóre v rysu otevřenost (Pustina & Daňková, 2016).

**Svědomitost** v textech (psaných i mluvených) českých respondentů vykazovala signifikantně těsné vztahy s největším počtem lingvistických prvků. U dopisu z dovolené málo svědomití jedinci používali zvýšeně výrazy hovorové a nespisovné češtiny. V zahraničních studiích dospěli ještě k „radikálnějšímu“ vztahu Lee, Kim, Seo a Chungová

(2007), jejichž korejští respondenti s nízkým skóre svědomitosti užívali ve svých eseích zvýšeně vulgarismy. Oba výsledky mohou být vnímány též jako projev validity, neboť vysoké skóre svědomitosti značí perfekcionismus, který znamená výraznou sebedisciplinu a potřebu být nejlepší. Jedinci tedy vykazují přirozeně i zvýšenou tendenci splnit očekávání a (vztaženo k výzkumu) produkovat text v optimálním jazykovém modu (ve výzkumu je zpravidla implicitně očekáváno produkovat text ve spisovné formě). Tomu odpovídá i detekovaná rostoucí deskriptivita textu v dopise s omluvou (klesá poměr sloves ku podstatným a přídavným jménům). Zajímavý je zvýšený počet neurčitých číslovek v psané omluvě svědomitých jedinců. Podobný vztah v jiných výzkumech nebyl odhalen. V rozhovoru o příjemném zážitku se ukázal jako průkazně těsný vztah mezi svědomitostí a výskytem 2. osoby, což může odkazovat silnou ukotvenost v societě či silnou potřebu sociálního kontaktu svědomitých lidí. Tomu odpovídají výsledky jiných výzkumů – například Balaščíková a Blatný (2003) identifikovali těsný záporný korelační vztah mezi svědomitostí a sociální izolací.

**Přívětivost** popisuje člověka na škále afliace (přátelskost) – kompetice (soutěživost). V českých textech nebyl po korekci detekován žádný dostatečně silný vztah přívětivosti s lingvistickými prvky textu. Zahraniční studie (Gill a kol., 2006; Holtgraves, 2011; Lee, Kim, Seo, & Chung, 2007; Pennebaker & King, 1999; Yarkoni, 2010) některé vztahy odhalily (zvýšený výskyt 1. osoby, zájmen a slov vyjadřujících pozitivní emoce, snížený výskyt slov vyjadřujících negativní emoce a negace jako morfologického jevu). V naší studii byla také identifikována celá řada vztahů, jejich síla však po korekci nesplnila přísně stanovené podmínky pro riziko chyby I. druhu. Hodnoty nekorigovaných korelací naznačují, že vztahy vyskytující se v českých textech směřují k obdobným výsledkům jako uvedené zahraniční studie: v dopise z dovolené korelovala přívětivost s negativitou ( $r = -0,15$ ), ve stížnosti osobní zájmena v 1. osobě ( $r = 0,18$ ), v omluvě přivlastňovací zájmena v 1. osobě ( $r = 0,18$ ) a negativní příslovce ( $r = 0,17$ ), v přijímacím pohovoru první osoba ( $r = 0,18$ ) a navíc 3. osoba záporně ( $r = -0,14$ ). České texty navíc vykazují silnější vztahy přívětivosti k lingvistickým prvkům v přijímacím pohovoru: minulý čas ( $r = 0,14$ ) a budoucí čas ( $r = -0,14$ );

v rozhovoru o příjemném zážitku minulé aktivní přičestí ( $r = 0,18$ ); ve stížnosti silně negativně korelují s přívětivostí afirmativní příslovce ( $r = -0,21$ ). Jedná-li se o specificky české trendy, nebo jen náhodně nalezené vztahy ukáže teprve případný další výzkum.

### 3.4.4.3.2 Implikace k vnětovým charakteristikám

---

V různých sociálních interakcích bývá užíváno několik základních řečových registrů, které se liší specifickým lexikem a vyznačují se fonologickými, morfologickými anebo syntaktickými zvláštnostmi v užívání jazyka (např. deklamační, formální, informativní, familiární a intimní dle Hayes, 2000). Jednotlivé rejstříky se kromě uvedeného liší též **kapacitou** (rozsahem) a **ostrotí hranice** registrů, a to směrem od nejmenšího a jasně vymezeného deklamačního rejstříku po intimní. Například deklamační scénář „český prezidentský projev“ začíná konvencionalizovaným oslovením: *Vážení spoluobčané* a je jen velice omezený okruh jiných přípustných variant (tj. má velmi omezený rejstřík). Na druhou stranu například v intimním scénáři partnerský rozhovor může oslovení životního partnera variovat (tj. má široký rejstřík) od prostého jména a jeho zdrobněliny, přes *miláčka, lásku, zlatíčko, broučka, mamku* k neotřelému *taťouzovi, dobyvateli* a jiným až bizarním oslovením typu *čudlymán* či *čut pet* (výrazy z diskuse na téma „Jak říkáte svým partnerům kromě jména?“ na portálu eMimino, MAFRA, 2011), které jsou součástí jen velmi omezeného počtu idiolektů a přesto bezesporu patří do intimního řečového registru a plnohodnotně naplňují jeho funkce.

Počty nalezených vztahů, které dosáhly požadované hladiny průkaznosti, ukazují, že nejméně vztahů mezi sledovanými rysy osobnosti a lingvistickými charakteristikami produkovaných textů lze nalézt u psaných formálních textů (TXT1 motivační dopis a TXT3 stížnost), kde po FDR korekci nebyl nalezen dokonce žádný statisticky průkazný vztah. Nejvíce se osobnost promítá do mluvených textů. Výsledek poukazuje na výrazný **vliv scénáře** (psaný versus mluvený projev) **na vztah mezi osobností a lingvistickými prvky textu.**

K interpretaci můžeme využít metodu analogie.

**Kapacita řečového registru:** Ve formálních textech je registr komunikačních prostředků v porovnání s neformálními texty chudý, omezený a poskytuje tak komunikujícímu jen výrazně omezený „manévrovací prostor“ pro volbu vyjadřovacích prostředků. Situaci můžeme přirovnat k volbě oděvu na formální událost (například na koncert, svatbu, pohřeb, promoci, křtiny, předávání výročních ocenění, či ve vztahu k naší studii na přijímací pohovor na vysněné zaměstnání či na soudní jednání jako svědek porušování nočního klidu) – v běžném šatníku je v porovnání s každodenní zkušeností podstatně zúžený výběr vhodného oděvu na tyto mimořádné formální události. Zatímco na všední dny vybíráme a kombinujeme z mnoha kousků oděvu včetně různých doplňků, na výjimečné formální události jsme rádi, je-li vůbec z čeho volit. Analogicky je tomu u řečového projevu. Prostor pro projevení individuality je v případě formálních situací výrazně omezen (zpravidla) **nízkým množstvím dostupných** lingvistických stylistických prostředků řečového registru; neformální situace aktivující neformální registry poskytují větší prostor projevení osobnosti skrze idiolekt.

**Ostrost hranic řečového registru:** Ve formálních textech je registr komunikačních prostředků v porovnání s neformálními texty mnohem pevněji ohraničen. Například formální dopis předpokládá pečlivé užití gramatiky a volbu adekvátních spisovných výrazů. Použití hovorových gramatických forem či slangových výrazů ve formálním dopise zřetelně signalizuje překročení stylové hranice registru formálního dopisu, zatímco u neformálního dopisu (z dovolené či s omluvou) je užití takových prostředků obvykle vnímáno jako přípustné. Situaci můžeme opět přirovnat k volbě oděvu na formální událost – většina formálních událostí má jasně vymezený *dress code* (například svatba či přijímací pohovor). Zatímco na dovolenou si můžeme obléknout takřka cokoli, aniž bychom vyvolali pocit něčeho nepatřičného, přijít na svatbu v teplácích či na přijímací pohovor do vysněného zaměstnání v lidovém kroji po babičce by jistě vyvolalo překvapení, údiv, úžas, ohromení či přímo vlnu nevole. To proto, že byly překročeny nepsané hranice normy. Čím jasněji (detailně, výslovně, přímo, stabilně) by byl *dress*

*code* vztahující se ke konkrétní situaci vymezen, tím intenzivnější by byla odezva na jeho porušení. Analogicky je tomu u řečového projevu. Neformální texty odkazující na familiární a intimní řečový registr poskytují rozsáhlejší prostor pro uplatňování větší variability různých lingvistických prostředků, aniž by došlo k porušení hranice registru, jejich hranice je méně zřetelně vymezená a podstatně volnější. Prostor pro projevení individuality je v případě formálních situací výrazně omezen (zpravidla) **úzce vymezenou hranicí** přípustných lingvistických stylistických prostředků řečového registru; neformální situace aktivující neformální registry skýtají uvolněnější podmínky projevení osobnosti skrze idiolekt.

Z psychologického hlediska je vhodné si uvědomit obecný mechanismus působení hranic, norem, pravidel, konsenzu na jedince. Ať už jedinec vyrazil na gothic metal koncert v obleku od Armaniho, jel v obci stotřícítkou nebo napsal žádost o práci ve verších, ve všech případech se jedná o situace překračování (vnímaných, očekávaných, předpokládaných) hranic, norem, pravidel, úzu, konsenzu. Obecně platí, že čím výrazněji je situace vymezena normami, tím silněji působí sociální tlaky na jejich dodržování (a tím silnější jsou sankce v případě jejich nedodržení) a tím intenzivnější musí být pak působení vnitřních a situačních vlivů (kognitivních, intelektových, osobnostních, emočních, motivačních aj.), aby k překročení hranic došlo. Přírozeným důsledkem je, že čím pevněji jsou stanoveny normy (hranice registru), tím méně jedinců zpravidla tyto hranice poruší. **Omezené množství jedinců**, kteří překonají překážku ve formě psaných či nepsaných norem, tj. v našem případě ve formálních textech (žádost, stížnost) překročí hranice jasně ohraničeného formálního řečového registru, pak statisticky znemožňuje, abychom našli dostatečně silné korelační vztahy mezi rysy osobnosti a lingvistickými charakteristikami textů.

Navzdory tomu nelze formální texty jako zdroj dat pro nalezení vztahů mezi osobností a idiolektem jednoduše zavrhnout. Je-li k dispozici normalizovaný soubor lingvistických charakteristik fiktivních žádostí o práci, může být velmi cenné identifikovat jedince, jejichž formální texty hranice stylu překračují, tj. jejichž idiolekt ve formálních textech vykazuje výrazně individuální strukturu (například statisticky



identifikovat tzv. outliers) a **analyzovat** u nich sledované vztahy jinými metodami a postupy, např. formou **případových studií**.

### 3.4.4.3.3 Druh textu – psaný nebo mluvený

---

Je známo (viz kapitola 3.4.3), že existují rozdíly mezi psaným a mluveným textem. Mluvená a psaná řeč byly označovány např. jako dva autonomní módy řeči (Adam, 2009). Konkrétně například Neku-la (1991) uvádí, že text zaznamenávaný na trvalé médium (papír, displej atp.) dovoluje více rozvinout informační jádro, zhustit informaci, ve větší míře využít syntaktických vzorců, zatímco nestálé médium (mluvená řeč) vede k výrazovým stereotypům, opakování apod. McLuhan (1991) dokonce prokazuje vliv psaného textu na organizaci lidské životní zkušenosti a má celospolečenské důsledky. Oba druhy textu z tohoto hlediska tak mohou být cenným zdrojem poznatků a zavrnutí některého z nich by omezilo možnosti bádání a ohrozilo odhalení cenných souvislostí. Ve prospěch sledování textů obou modů hovoří i fakt, že vztahy detekované v naší studii u mluvených a u psaných textů nejsou ve vzájemném rozporu, nýbrž se spíše navzájem doplňují. Umberto Eco (2001) k tomu dodává, že nové technologie naopak mohou vést ke stírání rozdílů mezi mluvenou a psanou řečí, když například umožňují psát needitovanou řeč, a obohacují tak psaní o rys spontánnosti a bezprostřednosti charakteristický pro řeč mluvenou. Dělení na mluvenou a psanou řeč je při tom velmi hrubým sítem. Proto byly v našem výzkumu použity jak psané, tak mluvené scénáře a dále bylo přistoupeno k porovnání scénářů dle tohoto kritéria. Ukazuje se, že **více silnějších vztahů bylo před i po korekci detekováno u mluvených textů** (TXT5 motivační pohovor a TXT6 vyprávění, vždy 10 průkazných vztahů).

### 3.4.4.4 Shrnutí

---

Získané výsledky nás opravňují k formulaci následujících tvrzení o vztazích mezi osobnostními charakteristikami dle pětifaktorového modelu (BigFive) a textovými charakteristikami:

1. **Vztahy mezi osobností a textem jsou (až na výjimky) transjazykové podstaty.** Většina vztahů detekovaných i po korekci jako signifikantních byla potvrzena také nejméně jednou, zpravidla však hned několika dalšími zahraničními studiemi.
2. **Osobnostní rysy se více promítají do textových charakteristik v komunikačních situacích s neformálním rámcem.** V této studii se jedná o dopis z dovolené, vyprávění o příjemném zážitku a dopis s omlouvou. Formální rámec poskytuje prostor pro kvalitativní analýzy outlierů, spíše než pro kvantitativní zobecnění.
3. **Osobnostní rysy se více promítají do textových charakteristik mluvených textů, oproti psaným textům.** Do dalších výzkumů je možné použít jakékoli médium, důležité je detailně popsat komunikační situaci, resp. řečový registr.
4. **Čím větší je kapacita a neurčitější hranice řečového registru daného scénářem, tím více se do něj promítá osobnost jedince.** Ve formálních textech je zvýšené riziko, že omezení dané vlastnostmi řečového registru (omezená kapacita, ostrá hranice) redukuje nebo zcela eliminuje promítání individuálních rysů osobnosti do textu.

## 3.4.5

# Odraz osobnosti v textu

## – projevy deprese

*Jana M. Havigerová, Jiří Haviger, Justýna Franková*

V říjnu 2012, u příležitosti 20. výročí zavedení Mezinárodního dne zdraví, zvolila Světová federace pro duševní zdraví (WFMH, 2012) jako téma roku problematiku deprese, která „byla v roce 2004 třetí nejčastější nemocí na světě a do roku 2030 se přesune na první příčku“. V březnu tohoto roku (WHO, 2018a) vydala Světová zdravotnická organizace prohlášení (WFMH, 2018), ve kterém konstatuje, že deprese již „je celosvětově hlavní zdravotní příčinou dočasné i trvalé pracovní neschopnosti (invalidity)“ (volně přeloženo z: *Depression is the leading cause of disability worldwide, and is a major contributor to the overall global burden of disease*).

Neblahý trend se naplnil dříve, než odborníci očekávali; deprese je v současnosti běžným onemocněním, které aktuálně celosvětově postihuje 300 milionů lidí všech věkových kategorií (častěji ženy), 800 tisíc lidí každoročně v důsledku deprese vlastní rukou umírá a u dospívajících a mladých dospělých ve věku 15-29 je sebevražda dle WHO (2018a) dokonce druhou nejčastější příčinou úmrtí. Jedním dechem Světová zdravotnická organizace (WHO, 2018a) dodává, že existuje účinná psychologická a farmakologická léčba. Účinnost, podobně jako u řady jiných nemocí, zvyšuje časná identifikace – čím dříve je rozvíjející se deprese odhalena, tím větší šance na uzdravení. Nepříznivý trend lze zastavit a obrátit – vedle prevence, osvěty či práce s ohroženou populací (rizikem je například dlouhodobá nezaměstnanost, zážitek těžké ztráty, psychické trauma) též zachycením pre-klinické symptomatiky deprese a zajištěním časně psychologické intervence či léčby lidí ohrožených a lidí v počátečních fázích tohoto onemocnění. Potenciál časného zachycení rizikové populace má mj. i automatizovaná analýza textu. Potencialita přispět k časně identifikaci ohrožených osob je významnou motivací také pro vznik této studie projektu CPACT.

### 3.4.5.1

## Osobnost a text – studie vztahující se k poruchám emotivity se zaměřením na depresi

---

**Afektivní poruchy** (též poruchy nálady) vymezuje mezinárodní klasifikace nemocí (MKN-10) jako poruchy, jejichž základní charakteristikou je porucha afektivity nebo nálady směrem k depresi (současně s úzkostí nebo bez ní) nebo k euforii. Tato změna nálady je obvykle doprovázena změnou v celkové aktivitě. Většina ostatních symptomů je buď sekundární ke změnám nálady a aktivity, nebo může být v tomto kontextu snadno vysvětlena. Afektivní poruchy se dle klasifikace nemocí označují kódem F30-39. Celosvětově se prevalence afektivních poruch pohybuje kolem hodnoty 10,6 % (IQR = 7,6-17,9 %), výskyt je častější u žen než u mužů (muži 5-12 %, ženy 10-25 %) (Kessler et al., 2009), riziko deprese se mění v závislosti na vzdělání (s rostoucím vzděláním klesá) a na věku (kulminuje v rané adolescenci a středním věku, nejohroženější skupina jsou senioři 70+) (Ritchie & Roser, 2018). Mezi afektivními poruchami jednoznačně dominuje výskyt periodické depresivní poruchy (*major depressive disorder*, MDD), jejíž prevalence zaujímá minimálně 50 % všech afektivních poruch (Kessler, 2009). Je třeba připomenout rozdíl mezi pojmy deprese a depresivní porucha.

**Depresivní porucha** je chronické onemocnění charakterizované epizodami závažné poruchy nálady, která výrazně postihuje prožívání a chování jedince. Charakteristický projevem depresivní poruchy je zhoršená nálada, která je neustále přítomna nezávisle na vnějších podnětech. Objevují se pocity beznaděje, sebeobviňování, ztráta zájmu o sebe i okolí, zhoršení koncentrace a menší množství energie. V případě výskytu somatických symptomů jsou časté psychomotorická retardace či agitovanost, poruchy spánku, nechutenství a ztráta libida. U těžké depresivní fáze jsou přítomny suicidální rozlady a pokusy. Negativní projevy této poruchy mohou vést k fatálním důsledkům v podobě dokonaného suicidia. A právě v souvislosti se suicidiem, nejvýraznějším a krajním projevem deprese, byly realizovány první psycholingvistické studie.

Deprese (z lat. *de* – dolů a *primere* – tlačit) neoznačuje pouze periodickou depresivní poruchu, nýbrž zpravidla označuje všechny afektivní poruchy nebo jejich epizody (např. bipolární afektivní poruchy, depresivní fáze, perzistentní afektivní poruchy, resp. sezónní afektivní poruchy, peripartum deprese, premenstruační dysforická porucha, situační či atypická deprese) v rozličných intenzitách (od lehké, *mild*, formy po těžkou) a s rozličnou délkou trvání (krátká, rekurentní), v jejichž popředí stojí tzv. symptom patologicky hlubokého smutku.

Na začátku nového tisíciletí publikovali Stirman s Pennebakerem (2001) studii, ve které zkoumali básně umělců, kteří spáchali **sebevraždu** a porovnávali je s básněmi těch, kteří sebevraždu nespáchali. V každé z těchto skupin bylo 9 básníků, analyzováno bylo celkem 291 básní. U suicidiálních básníků bylo v jejich pozdějším tvořivém období zjištěno častější užívání zájmena v první osobě jednotného čísla a nižší výskyt zájmen v první osobě plurálu. Struktura osobních zájmen v básních suicidiálních autorů by podle Stirmana a Pennebakerera (2001) mohla odkazovat ke klasické, sociologicky (statistikami) odvozené teorii sebevraždy jakožto důsledku poruchy sociální integrace (viz teorie Émila Durkheima, 1951), podle které jsou sebevražedné osoby buď nedostatečně integrovány do společnosti (např. v individualistických společnostech, kde je málo regulativů a příliš mnoho možností), nebo naopak integrovány až příliš (podléhají příliš velké regulaci). V obou extrémech platí, že sebevražda je projevem patologické socializace, proto se obráží v jazyce v odlišném poměru užívání osobních zájmen první osoby singuláru a plurálu.

Průkazné výsledky psycholingvistických studií sledujících výskyt specifických lingvistických charakteristik textů osob s dokonaným suicidiem (viz zmíněná studie Stirmana a Pennebakerera z roku 2001) podnítil celou řadu dalších studií, které analyzovaly analogické souvislosti mezi depresí a textem a které se shodovaly v základní myšlence. Lišily se však částečně v teoretických východiscích, metodologii i interpretačním rámci. Po již zmíněných případových studiích a analýzách osobních dokumentů lidí, kteří spáchali sebevraždu, následovaly studie, které se snažily ověřit a objasnit souvislost většího výskytu osobního zájmena jako projevu sebezaměřené pozornosti a deprese.

Bezprostředně v návaznosti na studii Stirmana a Pennebaker (2001) provedli Stephanie Rudeová, Eva-Marie Gortnerová a Pennebaker v roce 2004 výzkum, ve kterém se zaměřili na tři skupiny respondentů: studenty aktuálně prožívající depresi, studenty se zkušenostmi s depresí v minulosti a skupinu studentů z hlediska deprese intaktních. Rozdělení studentů bylo provedeno na základě výsledku obecně často používané sebesposuzovací škály BDI (*The Beck Depression Inventory*), analyzovány byly eseje. Pennebaker a jeho kolegyně vycházeli ve svých předpokladech ze dvou teorií, které jsou vodítkem i pro naše analýzy.

První z nich je Beckův kognitivní model deprese, ve kterém hrají stěžejní roli negativní obsahy mysli. Beck (1970) zjistil, že osoby s depresí vykazují typické vzorce myšlení:

1. Negativně interpretují veškeré události ve svém životě,
2. mají negativní vztah k sobě samým (nemají se rádi),
3. negativně vnímají svou budoucnost.

Tzv. kognitivní triáda má dle Pennebaker a jeho kolegyně odraz také v nakládání s jazykem. Autoři na základě Beckova modelu deprese očekávali, že v textech esejí studentů se zkušeností s depresí (v minulosti nebo aktuálně prožívanou) se budou častěji vyskytovat slova s negativním významem jako projev negativního myšlení.

Druhou z nich je teorie depresivního sebezaměřujícího kognitivního stylu Toma Pyszczynského a Jeffa Greenberga (1987), kteří zjistili, že u depresivních jedinců se zvýšeně vyskytuje na sebe zaměřená pozornost, která je zpravidla spouštěná hrozbou nebo zkušeností ztráty. Během procesu seberegulace jsou srovnávány současné a požadující stavy tak, aby chování mohlo vést ke snížení případného nesouladu a obnovení ztraceného objektu. Pokud však nejsou takové reakce k dispozici, sebezaměření způsobuje spirálovitý proces ovlivňující kognitivní, emocionální a behaviorální přizpůsobení a vrcholí depresí. Depresivní sebezaměření obsahuje sebekritiku a sebeobviňování, vyhýbá se zaměření na pozitivní události, přispívá k negativnímu sebeobrazu. Depresivní osoba by podle modelu depresivního sebezaměření, a očekávání Pennebaker a jeho kolegyně, měla vést k častějšímu popisu sebe sama, svých aktivit, chování nebo postojů za účelem snížení nesouladu

mezi vnímaným současným a požadujícím stavem, a tudíž v textech častěji mluvit v první osobě nebo užívat přivlastňovací zájmena, nebo i slovesa v první osobě a minulém čase. Oba očekávané lingvistické příznaky deprese (zvýšený výskyt negativních slov dle Beckova modelu i zvýšený počet zájmen v první osobě jednotného čísla jako znak sebezaměřené pozornosti podle modelu Pyszczynského a Greenberga) byly v textech studentů se zkušeností s depresí skutečně verifikovány. Obě teorie jsou nadále nejčastěji uplatňovány jako interpretační rámec nalezených zjištění.

Opakovaná měření depresivních příznaků v souvislosti s lingvistickými charakteristikami provedli Zimmermann, Brockmeyer, Hunn, Schauenburg a Wolf (2017). S účastníky studie, 29 hospitalizovanými pacienty, byl na začátku (tj. v průběhu jejich hospitalizace, nikoli v době příjmu) proveden polostrukturovaný rozhovor, který byl přepsán a analyzován. Následně pacienti vyplnili dotazník zaměřený na sledování aktuálních příznaků deprese, který poté vyplnili i při propouštění a po uplynutí následujících 6 měsíců. Při prvním měření nebyla zjištěna žádná korelace zájmen a aktuálně prožívané míry deprese, nicméně počet všech zájmen a zvláště pak objektových (*me*) a přivlastňovacích (*my*) v první osobě, vykazoval těsný statistický vztah s depresivními příznaky při propouštění a po něm.

Jednu z dalších studií provedly také výzkumnice Pulvermanová, Lorenzová a Mestonová (2015). U žen, které v dětství zažily sexuální zneužívání, zaznamenaly při expresivním psaní v souvislosti s úbytkem depresivních symptomů pokles používání zájmena „I“ a zvýšení používání pozitivně emočních slov. Výskyt lingvistických jevů koreloval také se zvýšením sexuální funkce a sexuálního uspokojení žen. Autorky tak v této studii prokázaly nejen souvislost mezi depresivní symptomatikou a lingvistickými charakteristikami textu, ale také terapeutický účinek expresivního psaní.

Zajímavý experiment zaměřený na projevy deprese v běžném životě provedl Mehl (2006), který dal účastníkům automatické nahrávací zařízení, které nosili během ranních hodin po dva týdny. Příznaky deprese, měřené Beckovým inventářem BDI, pozitivně korelovaly s užíváním zájmen první osoby čísla jednotného a výskytem slov v souvislosti

s kauzální spojitostí sdělovaných obsahů (slova odkazující na logické a chronologické souvislosti, příčinu a následek, např. *protože*).

Bernard, Baddeleyová, Rodriguez a Burke (2016) chtěli prozkoumat, jak jsou projevy deprese v textu ovlivněny **aktuálním afektivním stavem**. Provedli experiment, při kterém účastníky náhodně rozdělili do tří skupin. Každá skupina byla před psaním textu vystavena sérii speciálních obrázků s cílem navození pozitivního, negativního nebo neutrálního afektu. Závěry studie ukazují, že obojí, deprese i dočasně negativní nálady, ovlivňují používání zájmen, avšak deprese ovlivňuje použití zájmen v první osobě, zatímco negativní afekt ovlivňuje použití zájmen ve třetí osobě. Pozitivní i negativní afekt ovlivňují frekvenci užívání emočně negativních slov, přičemž prožívání pozitivního afektu je spojeno se sníženým výskytem, zatímco prožívání negativního afektu se zvýšeným výskytem negativních slov.

Většina dosud publikovaných studií byla realizována na podkladě germánské větve indoevropských jazyků (převládající jsou studie v angličtině, několik výzkumů bylo realizováno na německém a nizozemském jazyce). V roce 2017 vyšla studie analyzující lingvistické charakteristiky textu v kontextu deprese a neuropsychologických souvislostí v rumunštině, tedy na základě indoevropského jazyka románské větve. Trifuová, Nemes, Bodea-Hateganová a Cozmanová (2017) analyzovali přeepsané 5minutové monology osob s depresí o jejich zájmech. Depresivní osoby oproti těm z kontrolní skupiny užívaly více tautologií, opakování, srovnání. Monology depresivních osob měly podobu krátkých, osekáných a také neosobních vět. Depresivní osoby v monologiích užívaly prokazatelně více zájmen v singuláru, a naopak méně zájmen v plurálu ve všech třech osobách. Změny morfologicko-syntakticko-lexikálních jazykových dimenzí byly podle výsledků této studie dále spojeny s neuropsychologickými deficity v oblastech odpovídajících zpracování pracovní paměti, nastavení posunu (*set shift*), strategického plánování, pozornosti a psychomotorické rychlosti.

Fastová a Funder (2010) zkoumali genderové rozdíly v užívání sebe-referenčních osobních zájmen (*já, mi, mě* atp.) v kontextu deprese. Korelace zájme s depresí byla zjištěna u obou pohlaví a u žen byla jen nevýrazně silnější než u mužů. Na sebe-referenční slova u mužů se



zaměřili i Jarrold et al. (2011). S 12 nedepresivními a 14 depresivními muži průměrného věku 70 let provedli polostrukturovaný rozhovor, který analyzovali pomocí LIWC. Větší výskyt sebe-referenčních slov u depresivní skupiny zjistili ale pouze po zadání jediné konkrétní otázky, nikoliv během celého rozhovoru.

Zimmermann, Wolf, Bock, Peham a Benecke (2013) v německo-*ra-*kouské studii zkoumali používání zájmen v kontextu problémů v mezilidských vztazích. Zájmena první osoby singuláru pozitivně korelovala jak s příznaky deprese, tak i s interpersonálními problémy. Naopak používání zájmen první osoby množného čísla souviselo s mezilidskými problémy negativně.

Prakticky všechny studie zabývající se vztahem mezi depresí a lingvistickými charakteristikami textu, tj. textovými parametry, se shodují, že je evidentní **zvýšený výskyt sebe-referenčního osobního zájmena v první osobě v textech depresivních osob**. V roce 2017 byla publikována metaanalýza ověřující vztah mezi zájmenem první osoby v singuláru a depresí s ohledem na publikační zkreslení (Holtzman, 2017). Do metaanalýzy bylo zahrnuto 21 studií, z toho 9 nepublikovaných, a analýzy zahrnovaly celkem  $N=3\ 758$  subjektů. Závěry metaanalýzy byly jednoznačné: vztah mezi depresí a výskytem zájmena v první osobě singuláru (i s ohledem na možné publikační zkreslení) existuje. Síla tohoto vztahu je sice relativně malá (dle Nicholase Holtzmana odpovídá výši korelace  $r=0,13$ , 95 % CI = 0,10 až 0,16), avšak je prokazatelně nezávislá na pohlaví, věku, klinické či neklinické populaci, i na druhu textu, který byl analyzován.

Celá řada výzkumů se snažila zodpovědět otázku, zda se zvýšený výskyt klíčového osobního zájmena „já“ (ve všech jeho podobách, např. *mi*, *mě*, *mne*) a případně dalších ukazatelů deprese v textu (především negativních slov) váže pouze k fenoménu deprese anebo zda je projevem i jiných duševních poruch, a hledaly se případně společné mechanismy stojící za nalézánými vztahy.

Jednou z takových prací je studie Brockmeyera, Zimmermanna, Kulessové a kolegů (2015), kteří provedli výzkum na populaci osob s anorexií. U zkoumaných osob variovala míra komorbidní symptoma-

tiky deprese a anxiety, k nimž byly vztahovány výsledky analýz lingvistických charakteristik textů, resp. textových parametrů. Také v této sledované populaci koreloval výskyt osobních zájmen s příznaky deprese a rovněž úzkosti, avšak pouze v negativně zaměřené úloze („popište nejmudnější den svého života“). V pozitivně laděném textu („nejšťastnější den“) se korelace neprojevila. S podobnou úlohou, ale v podobě mluveného projevu, který byl přepsán a analyzován, srovnávali autoři používání zájmena také u chronicky depresivních osob, nechronicky depresivních osob a zdravých osob. Očekávání vyšší expozice zájmena u chronicky depresivních osob se nepotvrdilo, avšak významný rozdíl byl zjištěn mezi kontrolní skupinou a nechronicky depresivními osobami. Statisticky průkazná korelace byla zjištěna pouze v negativně zaměřeném textu, a to pouze u zdravé populace, a nebyla větší než korelace zájmena s mírou prožívané anxiety. Závěr autorského týmu byl, že nalezená fakta podporují hypotézu, že maladaptivní na sebe zaměřená pozornost (*self-focused attention*, SFA) je relevantní transdiagnostický fenomén, jehož spolehlivým lingvistickým ukazatelem je zvýšené užívání osobního zájmena v singuláru.

Anxieta, neboli úzkost, která ve studii Brockmeyera a kol. (2015) vykazovala shodné lingvistické koreláty textu jako dříve sledovaná deprese, je velice nepříjemný emoční stav, který směs pocitů zahrnující strach, zlé předtuchy a obavy, doprovázený vegetativními příznaky (pocení, bušení srdce, pocit nevolnosti, bolest na hrudi, zkrácené dýchání, chvění rukou). Chronicky se vracející přítomnost nadměrného, nepříjemného a opakujícího se pocitu úzkosti, která má podstatný vliv na každodenní život, bývá diagnostikována jako úzkostná porucha. **Úzkostné poruchy** zahrnují různé stavy, z nichž nejfrekventovanější jsou čtyři: fobické úzkostné stavy, jiné úzkostné poruchy, obsedantně-kompulzivní porucha, reakce na závažný stres a poruchy přizpůsobení. Úzkostné poruchy (v minulosti označované jako neurózy, angl. *anxiety disorders*) patří celosvětově mezi nejčastější duševní onemocnění: postihují minimálně 33,7 % světové populace (Bandelow & Michaelis, 2015). **Komorbidita** úzkostných poruch s depresí je velmi vysoká, například dle Melartina a kolektivu (2002) dosahuje minimálně hodnoty 57 %, Lamers a kolektiv (2011) dokumentuje na souboru několika

tisíc respondentů, že z osob s diagnostikovanou depresivní poruchou mělo 67 % aktuálně a 75 % epizodicky souběžně úzkostnou poruchu a naopak, u osob s diagnostikovanou úzkostnou poruchou mělo 63 % aktuálně a 81 % epizodicky souběžně depresivní poruchu, Zhou a kolektiv (2017) dokumentují hodnotu 71 %.

Analýze textů klinické populace osob s nejčastější formou úzkostné poruchy, se sociální fobií, se věnovala studie autorů Anderson, Goldin, Kuritaová a Gross (2008). V rámci studie byly analyzovány autobiografické vzpomínky osob fobiků a bylo detekováno častější užívání osobních zájmen singuláru ve výzkumné skupině oproti kontrolní intaktní. Výsledek, který koresponduje s výsledky detekovanými v textech osob s depresí, může být buď pouhým projevem komorbidit obou poruch (zmíněné výše), nebo stojí v pozadí obou poruch stejné mechanismy. Autoři studie interpretují nalezený vztah jako projev maladaptivní, na sebe zaměřené pozornosti, tedy stejného mechanismu, přisuzovaného osobám s depresí.

Významnou skupinou psychických poruch jsou **poruchy osobnosti a chování**. V mezinárodní klasifikaci nemocí MKN-10. jsou sdruženy do kategorie F60–69 a jsou popisovány jako „hluboce zakotvené a trvalé vzorce chování ukazující neměnnou odpověď na širokou škálu osobních a sociálních situací. Reprezentují extrémní nebo významné odchylky od způsobů, kterými průměrný člověk v dané kultuře vnímá, myslí, cítí a zvláště má vztahy k druhým. Takové vzorce chování mají tendenci být stabilní a zahrnovat různé oblasti chování a psychologického fungování. Jsou často, i když ne vždycky, sdruženy s různým stupněm subjektivní tísně a narušení sociálního výkonu“ (WHO, 2018b, n.p.). Epidemiologicky zaměřené studie upozorňují na vysokou komorbiditu s výše sledovanou depresí, např. dle Melartina a kolektivu (2002) je deprese diagnostikována souběžně s poruchami osobnosti ve 44 procentech výskytu.

Několik výzkumných skupin zaměřilo své studie na odhalování souvislostí mezi lingvistickými charakteristikami textu, textovými parametry a poruchami osobnosti, přičemž, jak ukážeme vzápětí, výsledky opět vedou především ke dvěma textovým parametrům: osobním zájmenům a emočně nabitým výrazům (pozitivním a negativním slo-

vům). Právě vysoká komorbidita poruch osobnosti s depresí či depresivní symptomatikou může být jedním z možných vysvětlení.

Nizozemští výzkumníci Molendijk, Bamelisová, van Emmerik, Arntz, Haringsmaová a Spinhoven (2010) zkoumali písemné eseje 304 ambulantních psychiatrických pacientů s poruchou osobnosti a 108 osob intaktní dospělé populace. Zjistili, že klinická populace používala v esejích statisticky průkazně více sebe-vztažných slov (*words referring to self*), více negativních slov a méně pozitivních slov. Zjistili ovšem rovněž, že tyto lingvistické změny nejsou specificky propojeny s aktuálně prožívanou depresí (*current depression*), jak očekávali.

Německé výzkumnice Charlotte Rosenbachová a Babete Rennebergová (2015) analyzovaly texty produkované v rámci testu autobiografické paměti (*autobiographic test*, AMT) osobami s hraniční poruchou osobnosti, u kterých byla souběžně diagnostikována bipolární (BPD; N=30) nebo depresivní porucha (N=27) a texty 30 intaktních dospělých. Lingvistická analýza textů ukázala, že ve srovnání s ostatními dvěma skupinami, pacienti s BPD vykazovali vyšší výskyt slov z oblasti sebezaměření (*self-focus*, např. osobní zájmena singuláru), slov souvisejících s hněvem (negativních slov), častěji odkazovali na sociální prostředí a hodnotili vzpomínky na odmítnutí v minulosti jako důležitější pro jejich současný život.

Tým Arntz, Hawkeová, Bamelisová, Spinhoven a Molendijk (2012) zaměřil pozornost na změny v textech psaných v průběhu terapie u osob s histriónskou, narcistickou nebo paranoidní poruchou osobnosti (N=299 osob) a u kontrolní intaktní populace (N=108). Nezávisle na druhu poruchy byl detekován těsný vztah mezi průběhem léčby a frekvencí používání zájmen první osoby, negativních slov a negací (zde pokles) a pozitivních slov (zde vzestup) v textech esejí o životech participantů. Výsledky dle autorů (i v souladu s výsledky jiných studií) poukazují na **transdiagnostický projev negativního myšlenkového stylu** (*negative thinking style*), který se projevuje v esejích v kontextu negativního obsahu a sebezaměření. Autoři uzavírají, že detekované změny poukazují na stěžejní význam obecné negativity (*general negativity*) jako klíčového diagnostického i léčebného faktoru.

Angela Careyová a její tým (Carey et al., 2015) zaměřili pozornost čistě na narcistickou poruchu osobnosti. Tato bývá dle autorů studie laickou i odbornou veřejností často spojována s tzv. „*I-talk*“, sebestředným řečovým projevem, pro který je charakteristické zvýšené užívání sebe-vztažného osobního zájmena v první osobě singuláru (Raskin & Shaw, 1988). V Careyové výzkumu byla analyzována data z textů získaných od rozsáhlého výzkumného souboru (N=4 811), z 5 různých laboratoří, z 6 různých registrů (komunikačních kontextů: sebe-vztažným, osobním, neosobním, veřejným a metodou automatického psaní, tzv. *stream-of-consciousness tasks*), na dvou jazykových bázích (angličtina a němčina), přičemž narcismus byl měřen pěti různými testy. Celkový efekt byl shledán ve výši aproximující k nule (*near-zero effect*,  $r=0,01$ , 95 % CI = 0,02-0,04) a hypotéza byla vyvrácena. Ke stejnému závěru dospěli například i Lisa Fastová a David Funder (2010), kteří navíc ve své studii poukázali na nalezený výrazný rozdíl v řadě morfostylistických a sémantických kategorií v první a druhé polovině analyzovaných interview (Cronbachovo  $\alpha$  se pohybovalo v rozmezí 0,41 až 0,64). Nezodpovězenou otázkou zůstává, proč vztah mezi seberefrenčním osobním zájmenem a narcistickou poruchou osobnosti nebyl potvrzen, zatímco analogický vztah nalézáme u ostatních poruch osobnosti.

Současný stav poznání o vztahu mezi textovými parametry a poruchami lidské psychiky vede k závěru, že podstatnými lingvistickými indikátory sledovaných afektivních poruch, úzkostných poruch a poruch osobnosti (pravděpodobně mimo narcistickou poruchu) korelují se **zvýšeným výskytem seberefrenčních osobních zájmen** (*já, mi, mě, mnou, má, můj, mou, mojí*) a **slov s negativním emočním nábojem**. Mezi sledovanými poruchami existuje vysoká komorbidita (deprese s úzkostnými poruchami vykazuje komorbiditu 60–70 %, deprese s poruchami osobnosti 40–50 %, viz např. Melartin et al., 2002; Lamers et al., 2011; Zhou et al., 2017), shoda textových parametrů se sledovanými poruchami tak může být ovlivněna tímto souvýskytem nebo se může generovat z obdobných mechanismů, stojících v pozadí sledovaných psychických poruch. Pokud se autoři vyjadřují k příčinám, zpravidla je vysvětlují modelem fungujícím napříč

různými sledovanými diagnózami. Jedná se o maladaptivní kognitivní styl spojený s negativně a sebe-středně zaměřenou pozorností, který se do textů osob s depresí, anxiétou či některou z poruch osobnosti promítá ve formě zvýšeného výskytu sebe-referenčních osobních zájmen v první osobě singuláru, zvýšeného výskytu slov s negativním emočním obsahem a dalšími textovými parametry, které jsou závislé na řečovém registru a vyskytují se napříč různými sledovanými jazyky (zde zejm. jazyky indoevropské rodiny). Na základě popsanych zjištění i v tomto výzkumu očekáváme, že osoby s depresemi nebo úzkostmi budou užívat více zájmen v první osobě čísla jednotného a emočně negativních slov, méně naopak slov s pozitivním emočním významem a potenciál vytvořit relativně spolehlivý prediktivní model umožňující odlišit osoby se zvýšeným rizikem deprese z lingvistických charakteristik textu.

### 3.4.5.2 Cíl studie

---

Cílem této studie je analyzovat projevy deprese v psaném textu a jazyce, konkrétně:

1. **verifikovat vztahy** mezi výskytem dříve detekovaných morfo-syntaktických a sémantických lingvistických charakteristik a sledovaných psychických stavů; očekáváme, že osoby s depresí a úzkostí budou produkovat texty se zvýšeným výskytem zájmen v první osobě a negativních slov,
2. **explorovat nové lingvistické charakteristiky**, které vykazují těsný vztah s negativní emotivitou (depresí) a mohou být eventuálně specifické pro český jazyk (nebo slovanskou jazykovou větev),
3. **vytvořit prediktivní model** umožňující detekovat osoby ohrožené rozvojem deprese (s vysokou mírou aktuálně prožívané deprese) v populaci.

Studie je realizována na textech klinicky intaktní populace českých dospělých jedinců (soubor P200) a klinické populace souboru dospělých jedinců s diagnostikovanou poruchou prožívání, u kterých do popředí vstupuje depresivní symptomatika (soubor P20+). Všichni respondenti jsou rodilí mluvčí češtiny.

### 3.4.5.3 Postup zpracování dat

---

Statistická analýza dat je prováděna v programu IBM SPSS Statistics. Ve výpočtech je operováno s neparametrickými metodami, protože textované charakteristiky nevykazují normální rozložení (ověřeno K-S testem). Pohlaví se v dosavadních výzkumech ukázalo jako výrazná intervenující proměnná, proto jsou všechny výpočty prováděny zvlášť pro ženy a pro muže.

Předpokládáme, že klinický soubor dosahuje vyšší míry aktuálně prožívané deprese než kvótní soubor. Pro ověření předpokládaného rozdílu je použit Mann-Whitney U test pro nezávislé výběry ve škále deprese DASS\_D. Vysokých hodnot na škále deprese však mohou dosahovat i jedinci z kvótního souboru (nejsou diagnostikováni, ale jsou ohroženi), proto je pro hledané vztahy klíčovým kritériem míra aktuálně prožívané deprese vyjádřená hodnotou skóre dosaženého na škále DASS\_D, nikoli příslušnost k souboru (kvótní nebo klinický). Vztah mezi mírou deprese a charakteristikami textu je ověřen Spearmanovým korelačním koeficientem  $\rho$ .

Na základě deskriptivních statistik a maxima každé proměnné pro každý ze čtyř textů jsou redukovány ty lingvistické proměnné, které nebyly v textech dostatečně zastoupeny. Zbývající proměnné jsou zahrnuty do dalších výpočtů. Následně je proveden Mann-Whitney U test pro posouzení rozdílů zastoupení těchto kategorií mezi výzkumnou a kontrolní skupinou s cílem zjistit, které ze zbývajících proměnných jsou pro tyto skupiny rozlišující. Do Mann-Whitney U testu jsou kromě morfologických znaků a indexů zahrnuty také konkrétní sémantické kategorie pro výrazy s emočním významem pozitivním a výrazy s emočním významem negativním.

U většiny lingvistických kategorií je pro výpočty použito jejich zastoupení k celkovému počtu slov v textu (zkratka „per W“), pokud daná lingvistická proměnná není určena jinak (jak je tomu u například u interpunkce – POSZ nebo indexů, viz přehled lingvistických proměnných v příloze 3.3.2). To je důležité pro interpretaci tabulek výsledků a průměrů. Porovnávanou a uváděnou hodnotou tak není počet slov v textu

jako takový, ale jejich podíl v celém textu.

Textové parametry, tj. proměnné, které byly Mann-Whitney U testem alespoň u některých textů vyhodnoceny jako rozlišující mezi výzkumnou a kontrolní skupinou, jsou zahrnuty do výpočtu korelací se škálou DASS-D. Korelace zastoupení lingvistických kategorií se skóry na škále DASS-Deprese jsou zjišťovány pomocí Spearmanovy korelace.

Za účelem vytvoření prediktivního modelu rozlišujícího depresivní a nedeprativní osoby pomocí textových parametrů jsou provedeny výpočty logistické regrese metodou *Backward Stepwise Likelihood Ratio* pro obě pohlaví a každý text zvlášť. Do výpočtu modelů jsou zahrnuty textové parametry, které se při zjišťování rozdílů mezi skupinami Mann-Whitney U testem ukázaly jako rozlišující. Do modelů jsou zařazeni pouze respondenti, v jejichž textech byly zastoupeny všechny sledované charakteristiky, proto počet respondentů v jednotlivých modelech variuje. Jako kritérium pro depresivitu a zařazení do příslušné kategorie byla použita hodnota skóru dosaženého na škále DASS-D > 6 (kategorie depresivní).

Pro každý model jsou vypočteny rovněž hodnoty *precision* a *recall*. *Precision*, neboli pozitivně prediktivní hodnota, v tomto případě udává, jaké procento ze všech osob modelem predikovaných jako depresivní bylo skutečně objektivně depresivních i podle našeho kritéria – skóre na škále DASS-D. Počítána byla z hodnot tzv. *true positives* (tp) a *false positives* (fp) (neboli chyba I. druhu) podle vzorce:  $precision = tp / (tp + fp)$ . *Recall*, neboli senzitivita, udává, jaké procento z depresivních osob dokázal model správně predikovat. Tento údaj byl počítán z hodnot *true positives* (tp) a *false negatives* (fn) (neboli chyba II. druhu) vzorcem:  $recall = tp / (tp + fn)$ . Dále byla vypočtena hodnota F, která vyjadřuje harmonický průměr hodnot *recall* a *precision*, podle vzorce:  $F = 2 \times (precision \times recall) / (precision + recall)$ .



### 3.4.5.4 Výsledky

#### 3.4.5.4.1

Rozdíly mezi kvótním a klinickým souborem ve škále deprese

---

Nejprve byly porovnány hodnoty dosažených skóreů na škále deprese v dotazníku DASS-21 mezi kvótním (N=200; P200) a klinickým (N=72; P20+) souborem. Mann-Whitneyův test prokázal statisticky průkazný rozdíl mezi sledovanými soubory na hladině významnosti 0,01. Průměrné dosažené skóre jsou uvedené v tabulce 3.4.5.4.1.

*Tab. 3.4.5.4.1 – Míra kauzálně prožívané deprese: deskriptivní statistiky (N=200, N=76)*

<b>Pohlaví/skór DASS_D</b>	<b>kvótní soubor – průměr</b>	<b>SD</b>	<b>klinický soubor – průměr</b>	<b>SD</b>
muži	4,64	3,94	9,52	4,39
ženy	4,49	3,48	10,09	6,40

#### 3.4.5.4.2

Rozdíly mezi skupinami v lingvistických charakteristikách textů

---

Rozdíly mezi kvótním a klinickým souborem v expozici lingvistických charakteristik byly rovněž posuzovány neparametrickým Mann-Whitneyovým U testem. Výpočty byly provedeny zvlášť pro každý ze 4 textů a pro každé pohlaví. Na základě výsledků U testu byly označeny všechny proměnné, které nedostatečně diferencují mezi skupinami, a tyto byly z dalších výpočtů vyřazeny.

### **Slovní druhy a detailně určené morfologické proměnné**

U tří slovních druhů – přídavných jmen (POSA), sloves (POSV) a předložek (POSR) byl zjištěn rozdílný počet slov (vzhledem k celkovému počtu slov) mezi skupinami, ale pouze u žen a jen u čtvrtého textu se zadáním dopis s omluvou. Hladiny významnosti uvádíme v tabulce níže. Počet zájmen (POSP) se významně lišil jen v prvním textu

(motiví dopis) u žen, a to na hladině významnosti 0,05. Počet příslovcí (POSD) se naopak lišil jen u mužů, a to v prvním a druhém textu (dopis z dovolené), také na hladině významnosti 0,05. Ve všech textech u žen se ukázal jako statisticky významný rozdíl mezi skupinami v počtu interpunkce (POSZ), u mužů byl tento rozdíl významný pouze u druhého textu ( $p < 0,05$ ).

Statisticky významné rozdíly mezi skupinami na hladině významnosti 0,01 ve všech textech jak u mužů, tak u žen byly zjištěny u všech pěti kategorií určující slovní druhy podrobněji. Jedná se o tyto proměnné: slovesa ve tvaru přítomného nebo budoucího času (SbPB), osobní zájmena (SbPP), obyčejné předložky (SbPR), neurčité číslovky (SbPa A) a slovesné tvary minulého aktivního přičestí (SbPp A).

*Tab. 3.4.5.4.2.A – Základní morfosyntaktické textové parametry – rozdíly mezi skupinami: Mann-Whitney U Test (uvedena signifikance) (N=200, N=72)*

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
POSN per W	0,92	0,48	0,18	0,78	0,27	0,31	0,35	0,20
POSA per W	0,31	0,54	0,16	0,90	0,45	0,18	0,81	0,03
POSP per W	0,27	0,12	0,98	0,73	0,04	0,11	0,22	0,30
POSV per W	0,49	0,94	0,13	0,73	0,93	0,09	0,86	0,01
POSD per W	0,05	0,05	0,35	0,69	0,38	0,66	0,53	0,94
POSR per W	0,73	0,89	0,72	0,94	0,67	0,11	0,15	0,02
POSJ per W	0,96	0,90	0,59	0,95	0,29	0,14	0,99	0,62
POSZ per Tok	0,28	0,05	0,38	0,54	0,02	0,04	0,02	0,04
SbPB per W	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SbPP per W	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SbPR per W	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SbPa A per W	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SbPp A per W	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Co se týče samotných průměrů, kategorie POSD (příslovce), v jejichž zastoupení se od sebe liší muži ve výzkumné a kontrolní skupině v prvních dvou textech, vykazovala vyšší průměrné zastoupení u mužů výzkumné skupiny. U žen výzkumné skupiny byl ve čtvrtém textu vyšší průměr kategorií přídavná jména (POSA) a předložky (POSR), v prvním textu zájmena (POSP). V zastoupení sloves (POSV) byl naopak vyšší průměr u kontrolní skupiny. U interpunkce (POSZ) byl u žen zjištěn ve všech textech vyšší výskyt u výzkumné skupiny, u mužů ve druhém textu byl však vyšší průměr zjištěn u skupiny kontrolní.

Průměr kategorií SpPB (slovesa ve tvaru přítomného nebo budoucího času), SbPR (obyčejné předložky) a SpPp A (slovesné tvary minulého aktivního přičestí) byl ve všech textech mužů i žen nižší u výzkumné skupiny. U kategorií SbPP (osobní zájmena) a SbPa A (neurčité číslovky) byl naopak u výzkumné skupiny vyšší.

### **Morfosyntaktické relativizované proměnné**

V následující tabulce 3.4.5.4.2.B je uvedena statistická významnost rozdílů mezi skupinami mužů a žen v textových parametrech jednotlivých textů (kategoriích a intexech). Statisticky významné rozdíly byly zjištěny u *Vocative indexu*, *Singulary P indexu*, Per2 a Per3 indexů, u indexů Deg1, Deg2 a Deg3 a Phr1 per W, což je počet frazémů ku počtu slov. Tyto rozdíly byly pokaždé zjištěny jen v některých textech a někdy rovněž jen pro jedno pohlaví.

*Vocative index* se významně lišil u mužů v prvním a druhém textu, nižším průměrem u mužů této výzkumné skupiny, stejně tak u žen, ale pouze ve druhém textu. Index Per2, tedy 2. osoba, se u obou pohlaví odlišoval pouze ve druhém textu, také nižším průměrem u výzkumné skupiny. Indexy Deg1 a Deg2 popisující počet přídavných jmen a 2. stupně se významně lišily u prvních dvou pozitivně zaměřených textů, ale pouze u mužů. Muži měli u přídavných jmen 1. stupně (Deg 1) nižší průměr, u 2. stupně (Deg 2) zase vyšší. Phr1, tedy frazémy, se lišily ve třetím textu (stížnost) u mužů a ve druhém (dopis z dovolené) u žen. V obou případech měla vyšší průměr výzkumná skupina. *Singulary P index*, který zjišťuje počet zájmen v první osobě čísla jednotného, se lišil pouze ve třetím textu u žen (vyšší průměr výzkumná skupina), Per3 in-

dex v druhém textu u mužů (vyšší průměr výzkumná skupina) a Deg3 index ve čtvrtém (také vyšší průměr výzkumná skupina). Hladiny významnosti jednotlivých rozdílů jsou podle zvýraznění patrné v tabulce.

*Tab. 3.4.5.4.2.B – Relativizované textové parametry – rozdíly mezi skupinami: Mann-Whitney U Test, signifikance (N=200, N=72)*

<b>Parametr</b>	<b>Muži</b>				<b>Ženy</b>			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Singulary index	0,93	0,72	0,50	0,53	0,63	0,45	0,49	0,81
Vocative index	0,01	0,00	0,49	0,55	0,08	0,02	0,11	0,57
Singulary P index	0,41	0,29	0,47	0,68	0,70	0,81	0,02	0,20
Per1 index	0,64	0,22	0,76	0,74	0,68	0,18	0,08	0,67
Per2 index	0,22	0,00	0,10	0,29	0,12	0,04	0,70	0,36
Per3 index	0,18	0,02	0,38	0,79	0,30	0,52	0,17	0,57
TenF index	0,29	0,90	0,81	0,15	0,14	0,31	0,41	0,46
TenP index	0,92	0,11	0,14	0,60	0,63	0,21	0,81	0,08
TenR index	0,59	0,07	0,13	0,21	0,93	0,08	0,78	0,09
Deg1 index	0,01	0,02	0,54	0,14	0,31	0,88	0,43	0,79
Deg2 index	0,01	0,02	0,88	0,72	0,60	0,68	0,61	0,73
Deg3 index	0,20	0,09	0,55	0,00	0,76	0,31	0,10	0,87
Negativity index	0,84	0,09	0,77	0,25	0,92	0,94	0,56	0,87
Pasivum index	0,37	0,31	0,41	0,90	0,51	0,70	0,61	0,26
Imperfectum index	0,39	0,45	0,66	0,57	0,22	0,20	0,95	0,47
Atypical words index	0,34	0,14	0,76	0,22	0,32	0,05	0,27	0,51
Var1 per W	0,40	0,99	0,43	0,42	0,42	0,56	0,54	0,20
Phr1 per W	0,16	0,13	0,04	0,25	0,80	0,02	0,23	0,59

## Lingvistické indexy

Z následující skupiny lingvistických indexů byl statisticky významný rozdíl mezi skupinami potvrzen u šesti indexů: *Coherence index*, *Formality index*, *Trager index*, *Agressiveness index*, *Readines to action index* a *Activity index*. Pouze *Cohorence index* se lišil u všech textů u mužů a u dvou u žen (druhý a čtvrtý text – obojí neformální texty) a ve všech případech šlo o statisticky významný rozdíl na hladině významnosti 0,01. Výzkumná skupina měla průměrnou hodnotu tohoto indexu vždy vyšší. *Trager index*, *Readines to action index*, *Aggressiveness index* a *Activity index* se významně liší pouze v posledním z textů – dopise s omluvou, a pouze u žen, kde jsou u výzkumné skupiny všechny tyto indexy nižší. *Formality index* je významně odlišný u čtvrtého textu u mužů a jeho průměrná hodnota je také u výzkumné skupiny nižší než u kontrolní.

Tab. 3.4.5.4.2.C – Lingvistické indexy – rozdíly mezi skupinami: Mann-Whitney U Test (uvedena signifikance) (N=200, N=72)

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Coherence index	0,00	0,00	0,00	0,00	0,93	0,00	0,06	0,00
Pronom index	0,39	0,18	0,68	0,98	0,06	0,13	0,21	0,21
Formality index	0,83	0,44	0,96	0,01	0,46	0,50	0,46	0,53
Trager index	0,40	0,52	0,19	0,99	0,64	0,13	0,84	0,02
Readines to action index	0,67	0,68	0,08	0,73	0,45	0,73	0,58	0,05
Aggressiveness index	0,49	0,94	0,13	0,73	0,93	0,09	0,86	0,01
Activity index	0,93	0,48	0,22	0,43	0,51	0,27	0,97	0,02
Autosemantic index	0,74	0,16	0,17	0,78	0,95	0,51	0,32	0,42

## Výrazy s emočním významem – sémantické proměnné

Všechny čtyři sémantické proměnné, textové parametry, týkající se emočního významu výrazů se statisticky významně lišily v prvních dvou (pozitivních) textech jak u mužů, tak u žen, a to na hladině významnosti 0,01. V případě stížnosti a dopisu s omluvou už je to složitější. U dopisu s omluvou byly zjištěny statisticky významné rozdíly ve všech čtyřech proměnných u mužů, u žen pouze ve třech proměnných (kromě Emo B – negativní význam). Ve třetím textu u mužů byl zjištěn rozdíl pouze v kategorii Emo B, zatímco u žen byly zjištěny statisticky významné rozdíly i v jedné kategorii slov s pozitivním emočním významem (Emo A) a v obou kategoriích s negativním (Emo B a Em2 B).

V motivačním dopise, dopise z dovolené a dopise s omluvou (texty 1, 2 a 4) měla výzkumná skupina vždy nižší průměr slov s pozitivním emočním nábojem (kategorie Emo A a Em2 A) a vyšší průměr emočních slov s nábojem negativním. U stížnosti je tomu ale přesně naopak – v kategoriích slov s negativním emočním nábojem mají výzkumné skupiny u mužů (Emo B) i u žen (Emo B i Em2 B) nižší průměr a kategorie Emo A (významně odlišná pouze u žen) vykazuje u výzkumné skupiny průměr naopak vyšší.

Tab. 3.4.5.4.2.D – Sémantické textové parametry – rozdíly mezi skupinami: Mann-Whitney U Test, signifikance ( $N=200$ ,  $N=72$ )

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Emo A per W	0,00	0,00	0,77	0,02	0,00	0,00	0,00	0,03
Emo B per W	0,00	0,00	0,01	0,02	0,00	0,00	0,00	0,18
Em2 A per W	0,00	0,00	0,29	0,00	0,00	0,00	0,12	0,00
Em2 B per W	0,00	0,00	0,06	0,00	0,00	0,00	0,03	0,00

Pokud se podíváme na počet rozlišujících proměnných v jednotlivých tabulkách, zdá se, že největší počet rozdílů byl u mužů nalezen v prvních dvou textech – tedy v motivačním dopise, a zvláště pak v dopise z dovolené. U žen je textem s nejvíce rozdíly dopis s omluvou a až po něm dopis z dovolené.

## **Výsledky a vyjádření k hypotézám**

Hypotéza o větším zastoupení zájmen první osoby čísla jednotného u klinické skupiny byla potvrzena v textu stížnost, a to u žen. Hypotézy o nižším zastoupení slov s pozitivním emočním nábojem a vyšším zastoupením slov s negativním emočním nábojem u klinické skupiny byly potvrzeny.

### **3.4.5.4.3 Vztahy mezi textovými parametry a DASS-D (deprese)**

---

Těsnost vztahu mezi textovými parametry, které vykazovaly vždy alespoň v jednom z textů u mužů anebo u žen statisticky průkazný rozdíl mezi expozicemi v textech kvótního a klinického souboru, byly testovány výpočtem Spearmanova korelačního koeficientu. Výsledky jsou prezentovány v tabulkách zvlášť pro každé pohlaví a stručně komentovány.

## **Slovní druhy a detailní morfosyntaktické proměnné**

Korelační koeficienty mezi skórem ve škále deprese (DASS\_D) a jednotlivými lingvistickými charakteristikami textů ukazují na značné mezipohlavní diference.

Kategorie slovních druhů častěji korelovaly v textech mužů než žen. Přídavná jména (POSA) negativně korelovala ve třech textech mužů, jedná se o nízké korelace na hladině významnosti 0,05. Předložky (POSR) negativně korelovaly v druhém textu u mužů a ve třetím u žen, síla vztahu je také nízká.

Pozitivní korelace na hladině významnosti 0,01 byla zjištěna ve třech textech mužů u sloves (POSV) a zde korelace dosahují i střední síly. U žen byla pozitivní korelace DASS s touto kategorií zjištěna pouze v jednom textu, na volnější hladině významnosti ( $p < 0,05$ ) a menší síly. Dále byla u mužů v prvních dvou textech zjištěna nízká pozitivní korelace s příslovci (POSD). Naopak pouze u žen, zato ve všech jejich textech, pozitivně korelovala interpunkce (POSZ).

Mezi kategoriemi určující slovní druhy podrobněji byly u osobních zájmen (SbPP) a neurčitých číslovek (SbPa A) zjištěny pozitivní korelace dosahující střední síly v textech mužů i žen. Osobní zájmena pozitivně korelovala ve všech textech u mužů a ve třech u žen. Neurčité číslovky ve všech textech obou pohlaví, pokaždé středně silně a na hladině významnosti 0,01. Negativní korelace střední síly byly zjištěny u obyčejných předložek (SbPR) ve všech textech mužů i žen na hladině významnosti 0,01. Dále byly zjištěny negativní korelace u sloves ve tvaru přítomného nebo budoucího času (SbPB) i u slovesných tvarů minulého aktivního příčestí (SbPp) ve všech textech žen, ale pouze v jednom (text 4) u mužů, tyto vztahy jsou nízké až střední síly.

Tab. 3.4.5.4.3.A – Vztah mezi mírou deprese a expozicí morfosyntaktických proměnných: Spearman. korelace ( $N_{muži}=127, N_{ženy}=145$ )

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
POSA per W	-0,1	-0,2*	-0,2*	-0,2*	0,0	0,0	-0,1	0,1
POSP per W	-0,1	0,0	0,1	0,0	0,0	0,1	0,1	0,0
POSV per W	0,2*	0,2*	0,3*	0,1	0,1	0,0	0,2*	-0,1
POSD per W	0,2*	0,2*	0,2	0,0	0,0	0,0	0,0	-0,1
POSR per W	-0,1	-0,2*	-0,2	-0,1	-0,1	0,0	-0,3*	0,1
POSZ per Tok	0,1	0,0	0,2	0,1	0,2*	0,2*	0,2*	0,2*
SbPB per W	-0,1	0,0	0,0	-0,2*	-0,2*	-0,2*	-0,2*	-0,3*
SbPP per W	0,2*	0,2*	0,2*	0,3*	0,3*	0,1	0,3*	0,4*
SbPR per W	-0,4*	-0,4*	-0,4*	-0,3*	-0,4*	-0,3*	-0,4*	-0,3*
SbPa A per W	0,3*	0,4*	0,3*	0,3*	0,4*	0,4*	0,3*	0,3*
SbPp A per W	-0,1	-0,1	-0,1	-0,2*	-0,2*	-0,3*	-0,2*	-0,3*

## Morfosyntaktické relativizované proměnné

Ve skupině lingvistických proměnných zahrnující relativizované proměnné je výskyt signifikantních korelačních vztahů spíše sporadický. *Vocative index* a *Deg2 index* pro druhý stupeň přídavných jmen korelují negativně a s malou silou vztahu u jednoho textu – *Vocative*



*index* v motivačním dopise u mužů, Deg 2 u žen.

U Deg 1, tedy 1. stupně přídavných jmen, byla v motivačním dopise u žen zjištěna pozitivní korelace malé síly. Phr1, kategorie pro frazémy, koreluje pozitivně ve formálních textech (motivační dopis a stížnost) u mužů. Jedná se nízkou korelací v prvním textu, ve třetím se síla korelace blíží ke střední velikosti.

Tab. 3.4.5.4.3.B – Vztah mezi mírou deprese a expoziční relativizovaných proměnných: Spearmanova korelace ( $N_{\text{muži}}=127, N_{\text{ženy}}=145$ )

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Vocative index	-0,2*	-0,1	-0,1	0,0	-0,1	-0,1	0,1	-0,1
Singulary P index	-0,1	0,1	0,0	0,0	0,0	0,2	0,1	0,1
Per2 index	0,0	-0,1	-0,1	0,0	0,1	0,0	0,0	0,1
Per3 index	0,0	0,1	-0,1	-0,1	-0,1	-0,1	-0,1	0,0
Deg1 index	-0,1	-0,1	0,1	0,0	0,2*	0,0	0,0	0,1
Deg2 index	0,1	0,0	0,0	0,0	-0,2*	-0,1	0,0	0,0
Deg3 index	0,0	0,1	-0,1	0,0	0,0	0,0	0,0	-0,1
Phr1 per W	0,2*	0,1	0,3*	0,0	0,0	0,0	0,1	0,0

## Lingvistické indexy

Korelace lingvistických indexů se skóre DASS-D byla častěji zjištěna v textech mužů než žen, kde byly korelace zjištěny pouze u stížnosti.

*Formality index* negativně koreloval v dopise z dovolené, stížnosti a dopise s omluvou u mužů, u žen pouze ve stížnosti. Zjištěné vztahy byly nízké až střední intenzity.

Ostatní indexy korelovaly pozitivně. *Trager index* pozitivně koreloval se škálou ve všech textech mužů, ale v žádném textu u žen. *Readines to action index* pozitivně koreloval ve všech textech u mužů, kde dosahoval nízké až střední síly a ve stížnosti u žen (nízká korelace).

*Aggressiveness index* pozitivně koreloval ve třech textech u mužů, kde byla zjištěna nízká až střední pozitivní korelace a také ve stížnosti u žen (nízká korelace). *Activity index* vykazuje korelaci pouze s jedním textem – stížností, u mužů. Je to nízká síla vztahu a pozitivní korelace.

Zatímco v oblasti rozdílů mezi výzkumnou a kontrolní skupinou se zdál třetí text, stížnost, pro hledání psycholingvistických souvislostí nejméně nosný (indexy se statisticky významně lišily pouze u žen ve čtvrtém textu), při hledání spojitých korelačních vztahů se ukazuje stížnost jako nevhodnější text pro detekci těchto vztahů. U žen byla shledána signifikantní korelace DASS-D s indexy zjištěna pouze v tomto textu a u mužů dosahovaly korelace právě ve stížnosti největší síly.

Tab. 3.4.5.4.3.C – Vztah mezi mírou deprese a lingvistickými indexy: Spearmanova korelace ( $N_{muži}=127, N_{ženy}=145$ )

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Coherence index	0,1	0,1	0,0	0,0	0,0	0,0	0,1	0,1
Formality index	-0,1	-0,3*	-0,3*	-0,2*	-0,1	-0,1	-0,2*	0,0
Trager index	0,2*	0,2*	0,3*	0,2*	0,0	0,0	0,2	0,0
Readines to action index	0,2*	0,3*	0,4*	0,2*	0,1	0,0	0,2*	0,0
Aggressiveness index	0,2*	0,2*	0,3*	0,1	0,1	0,0	0,2*	-0,1
Activity index	0,2	0,2	0,3*	0,2	0,0	0,1	0,2	0,0

## Výrazy s emočním významem – sémantické proměnné

U všech čtyř sledovaných sémantických proměnných byla v prvních dvou (pozitivních) textech jak u mužů, tak u žen zjištěna korelace na hladině významnosti 0,01. Síla zjištěných vztahů je u mužů nízká a středně silná, u žen mají všechny vztahy zjištěné v prvních dvou textech střední sílu.

Kategorie výrazů s pozitivním významem korelují se skóre na škále DASS-D negativně, zatímco kategorie výrazů s negativním emočním významem korelují pozitivně. Korelace popsanych směrů a nízké síly byly zjištěny u proměnných Emo B, Em2 A a Em2 B u mužů také v textu dopis s omluvou. Poněkud překvapivý je výsledek korelace u žen v textu stížnost, kde byla zjištěna pozitivní korelace u Emo A, což je opačný směr korelace než korelace slov s pozitivním emočním nábojem zjištěné v prvních dvou textech u obou pohlaví.

Tab. 3.4.5.4.3.D – DASS-D a textové parametry:  
Spearmanova korelace ( $N_{muži}=127$ ,  $N_{ženy}=145$ )

Parametr	Muži				Ženy			
	TXT1	TXT2	TXT3	TXT4	TXT1	TXT2	TXT3	TXT4
Emo A per W	-0,3*	-0,3*	0,1	-0,2	-0,3*	-0,4*	0,2*	-0,1
Emo B per W	0,3*	0,4*	-0,1	0,2*	0,4*	0,4*	0,0	0,0
Em2 A per W	-0,3*	-0,3*	0,1	-0,2*	-0,3*	-0,4*	0,1	-0,1
Em2 B per W	0,3*	0,4*	-0,1	0,2*	0,3*	0,3*	0,1	0,1

## Výsledky a vyjádření k hypotézám

V žádném z textů nebyla zjištěna očekávaná korelace zájmen v první osobě se skóre na škále DASS-D. Hypotézy o negativní korelaci skóre DASS-D a slov s pozitivním emočním významem a pozitivní korelaci skóre DASS-D a slov s negativním emočním významem byly potvrzeny v pozitivně zaměřených textech – motivačním dopise a dopise z dovolené u obou pohlaví. Hypotéza o pozitivní korelaci slov s negativním emočním nábojem byla navíc potvrzena i u textu dopis s omluvou, ale pouze u mužů.

### 3.4.5.4.4 Prediktivní modely – logistická regrese

Logistická regrese byla provedena zvlášť pro každé pohlaví a každý ze čtyř textů, získali jsme celkem 8 prediktivních modelů.

#### Motivační dopis – muži

Prediktivní model deprese u mužů z motivačního dopisu je výsledkem 23 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,37$  a zahrnuje 5 proměnných – POSP per W (zájmena), POSR per W (předložky), SbPR per W (obyčejné předložky), Deg2 index (2. stupeň přídavných jmen) a *Readines to action index*. Příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tabulce 3.4.5.4.4.A.

Tab. 3.4.5.4.4.A – Textové parametry predikující depresi mužům z textu 1: logistická regrese ( $n=126$ )

Text. parametry	B (S.E.)	Sig.
POSP per W	-11,8 (7,0)	0,091
POSR per W	22,0 (12,1)	0,068
SbPR per W	-20,1 (5,4)	0,000
Deg2 index	10,4 (5,2)	0,043
Readines to action index	4,4 (1,2)	0,000
Constant	-3,3 (1,8)	0,064

Na základě identifikovaných proměnných bylo 67 osob správně vyhodnoceno jako nedeprativních (*true negatives*), 29 správně jako depresivních (*true positives*). 10 osob bylo nesprávně zařazeno do kategorie depresivní (*false positive*), 20 osob bylo nesprávně zařazeno do kategorie nedeprativních (*false negative*). Pro tento model je hodnota *precision* = 0,74, *recall* = 0,59 a *F* = 0,66.

Tab. 3.4.5.4.4.B – Klasifikační tabulka predikce deprese mužům z textu 1: logistická regrese (n=126)

Pozorováno	Predikováno	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	67	10
DASS-D > 6	20	29

### Motivační dopis – ženy

Prediktivní model deprese u žen z motivačního dopisu je výsledkem 23 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,30$  a zahrnuje 5 proměnných – POSV per W (slovesa), POSZ per Tok (interpunkce), Deg2 index (2. stupeň přídavných jmen), *Activity index* a Emo A per W (slova s pozitivním emočním významem). Příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tabulce 3.4.5.4.4.C.

Tab. 3.4.5.4.4.C – Textové parametry predikující depresi ženám z textu 1: logistická regrese (n=141)

Text. parametry	B (S.E.)	Sig.
POSV per W	29,6 (10,2)	0,004
POSZ per Tok	15,9 (7,8)	0,042
Deg2 index	-10,5 (5,0)	0,034
Activity index	-9,5 (4,5)	0,035
Emo A per W	-39,8 (9,7)	0,000
Constant	-2,2 (1,5)	0,115

Na základě identifikovaných proměnných bylo 79 osob správně vyhodnoceno jako nedeprativních (*true negatives*), 25 správně jako depresivních (*true positives*). 12 osob bylo nesprávně zařazeno do kategorie depresivní (*false positive*), 25 osob bylo nesprávně zařazeno do kategorie nedeprativních (*false negative*). Pro tento model je hodnota *precision* = 0,68, *recall* = 0,50 a *F* = 0,57.

Tab. 3.4.5.4.4.D – Klasifikační tabulka predikce deprese ženám z textu 1: logistická regrese (n=141)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	79	12
DASS-D > 6	25	25

### Dopis z dovolené – muži

Prediktivní model deprese u mužů z dopisu z dovolené je výsledkem 25 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,35$  a zahrnuje 3 proměnné – SbPR per W (obyčejné předložky), Deg2 (2. stupeň přídavných jmen) a Em2 B per W (slova s širším negativním emočním významem). Příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tabulce tab. 3.4.5.4.4.E.

Tab. 3.4.5.4.4.E – Textové parametry predikující depresi mužům z textu 2: logistická regrese (n=77)

<b>Text. parametry</b>	<b>B (S.E.)</b>	<b>Sig.</b>
SbPR per W	-17,1 (8,2)	0,036
Deg2 index	-10,7 (5,6)	0,071
Em2 B per W	43,8 (25,6)	0,086
Constant	0,3 (0,9)	0,694

Pro tento model je hodnota *precision* = 0,70, *recall* = 0,52 a *F* = 0,60.

Tab. 3.4.5.4.4.F – Klasifikační tabulka predikce deprese mužům z textu 2: logistická regrese (n=77)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	44	6
DASS-D > 6	13	14

### Dopis z dovolené – ženy

Prediktivní model deprese u žen z dopisu z dovolené je výsledkem 20 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,40$ . Zařazené proměnné, příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tab. 3.4.5.4.4.G.

Tab. 3.4.5.4.4.G – Textové parametry predikující depresi ženám z textu 2: logistická regrese (n=92)

<b>Text. parametry</b>	<b>B (S.E.)</b>	<b>Sig.</b>
POSA per W	170,1 (88,4)	0,054
POSV per W	-135,5 (68,1)	0,047
POSD per W	154,6 (82,9)	0,062
POSZ per Tok	20,1 (10,0)	0,044
Degi index	8,1 (4,7)	0,087
Trager index	0,7 (0,4)	0,081
Activity index	124,0 (64,7)	0,055
Emo B per W	49,4 (14,1)	0,000
Constant	-82,1 (37,6)	0,029

Pro tento model je hodnota *precision* = 0,72, *recall* = 0,64 a *F* = 0,68.

Tab. 3.4.5.4.4.H – Klasifikační tabulka predikce deprese ženám z textu 2: logistická regrese (n=92)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	47	9
DASS-D > 6	13	23

### Stížnost – muži

Prediktivní model deprese u mužů ze stížnosti je výsledkem 22 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,35$ . Zařazené proměnné, příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tab. 3.4.5.4.4.I.

Tab. 3.4.5.4.4.I – Textové parametry predikující depresi mužům z textu 3: logistická regrese (n=114)

<b>Text. parametry</b>	<b>B (S.E.)</b>	<b>Sig.</b>
POSA_per_W	-18,4 (8,2)	0,026
POSP_per_W	-13,6 (7,6)	0,071
SbPR_per_W	-14,6 (4,6)	0,001
Deg1_index	8,3 (3,9)	0,035
Readines_to_action_index	1,8 (1,0)	0,069
Em2_A_per_W	27,7 (14,2)	0,051
Constant	-5,3 (4,0)	0,182

Pro tento model je hodnota *precision* = 0,70, *recall* = 0,52 a *F* = 0,60.



Tab. 3.4.5.4.4.J – Klasifikační tabulka predikce deprese mužům z textu 3: logistická regrese (n=114)

Pozorováno	Predikováno	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	55	11
DASS-D > 6	14	34

### Stížnost – ženy

Prediktivní model deprese u žen ze stížnosti je výsledkem 22 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,34$ . Zařazené proměnné, příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tab. 3.4.5.4.4.K.

Tab. 3.4.5.4.4.K – Textové parametry predikující depresi ženám z textu 3: logistická regrese (n=125)

Text. parametry	B (S.E.)	Sig.
POSR per W	-38,9 (10,7)	0,000
SbPB per W	-15,0 (7,1)	0,035
SbPP per W	34,1 (14,0)	0,015
Per2 index	-5,9 (3,6)	0,098
Em2 A per W	-33,1 (18,8)	0,079
Em2 B per W	25,4 (15,4)	0,100
Constant	4,8 (1,9)	0,012

Pro tento model je hodnota *precision* = 0,71, *recall* = 0,52 a *F* = 0,6.

Tab. 3.4.5.4.4.L – Klasifikační tabulka predikce deprese ženám z textu 3: logistická regrese (n=125)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	74	9
DASS-D > 6	20	22

### Dopis s omluvou – muži

Prediktivní model deprese u mužů z textu omluvy je výsledkem 22 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,38$ . Zařazené proměnné, příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tab. 3.4.5.4.4.M.

Tab. 3.4.5.4.4.M – Textové parametry predikující depresi mužům z textu 4: logistická regrese (n=112)

<b>Text. parametry</b>	<b>B (S.E.)</b>	<b>Sig.</b>
POSR per W	34,3 (17,3)	0,048
SbPR per W	-30,9 (11,0)	0,005
Vocative index	-44,7 (28,3)	0,114
Coherence index	-0,04 (0,002)	0,063
Formality index	-0,062 (0,027)	0,024
Emo B per W	44,6 (16,6)	0,007
Constant	-0,3 (1,0)	0,755

Pro tento model je hodnota *precision* = 0,75, *recall* = 0,51 a *F* = 0,61.

Tab. 3.4.5.4.4.N – Klasifikační tabulka predikce deprese mužům z textu 4: logistická regrese (n=112)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	64	7
DASS-D > 6	20	21

### Dopis s omluvou – ženy

Prediktivní model deprese u žen z textu omluvy je výsledkem 25 kroků logistické regrese provedené metodou *backward stepwise likelihood ratio*. Výsledný model má hodnotu Nagelkerke  $R^2=0,16$ . Zařazené proměnné, příslušné koeficienty B včetně intervalu spolehlivosti a hodnoty signifikace jsou uvedeny v tab. 3.4.5.4.4.O.

Tab. 3.4.5.4.4.O – Textové parametry predikující depresi ženám z textu 4: logistická regrese (n=127)

<b>Text. parametry</b>	<b>B (S.E.)</b>	<b>Sig.</b>
POSV per W	-12,3 (7,4)	0,096
SbPP per W	25,7 (7,6)	0,001
Trager index	0,049 (0,023)	0,038
Constant	1,3 (1,9)	0,487

Pro tento model je hodnota *precision* = 0,56, *recall* = 0,32 a *F* = 0,41.

Tab. 3.4.5.4.4.P – Klasifikační tabulka predikce deprese ženám z textu 4: logistická regrese (n=127)

<b>Pozorováno</b>	<b>Predikováno</b>	
	DASS-D ≤ 6	DASS-D > 6
DASS-D ≤ 6	72	11
DASS-D > 6	30	14

### 3.4.5.4.5 Výsledky a souhrnné vyjádření k získaným modelům

Pro přehlednost uvádíme tabulku srovnání hodnot Nagelkerke  $R^2$ , *recall*, *precision* a jejich harmonického průměru pro každý z osmi získaných prediktivních modelů.

Podle hodnoty Nagelkerke  $R^2$ , které zjednodušeně řečeno odpovídá poměru rozptylu popsaného modelem, pak vidíme, že s výjimkou omluvy u žen jsou všechny modely přibližně srovnatelné (Nagelkerke  $R^2$  nabývá hodnot od 0,30 do 0,40). Nejpříznivější hodnoty Nagelkerke  $R^2$  má u mužů dopis s omluvou (0,38) a u žen dopis z dovolené (0,40), který má také nejvyšší hodnoty *precision*, *recall* a F.

Kdybychom sílu modelu posuzovali podle harmonického průměru (F) hodnot *precision* a *recall*, potom se opět jako nejslabší prediktivní model ukazuje model generovaný z textu omluvy psaného ženami (F kleslo pod 0,5), síla predikce deprese z textu se v tomto případě v jistém slova smyslu blíží pravděpodobnosti hodu mincí. Všechny ostatní modely vykazují hodnoty F nad pomyslnou hranici 0,5, a to v rozmezí od 0,57 do 0,73. Na základě této hodnoty lze konstatovat, že **nejsilnější prediktivní model deprese pro muže je (poněkud překvapivě) z formálního řečového registru textu stížnosti, pro ženy z neformálního řečového registru textu dopisu z dovolené.**

Tab. 3.4.5.4.4.Q – Shrnutí modelů predikujících deprese

Text	Model	Nag. $R^2$	Precision	Recall	F
TXT1	muži	0,37	0,74	0,59	0,66
	ženy	0,30	0,68	0,50	0,57
TXT2	muži	0,35	0,70	0,52	0,60
	ženy	0,40	0,72	0,64	0,68
TXT3	muži	0,35	0,76	0,71	0,73
	ženy	0,34	0,71	0,52	0,60
TXT4	muži	0,38	0,75	0,51	0,61
	ženy	0,16	0,56	0,32	0,41

### 3.4.5.5 Výsledky

---

Cílem této studie bylo ověřit souvislosti mezi lingvistickými parametry a depresí, dříve zjištěných v jiných jazycích, také v českých textech, případně popsat vztahy nově objevené. Ověřováno bylo zejména častější užívání zájmena v první osobě a slov s negativním emočním významem, a naopak menší užívání slov s pozitivním emočním významem u osob s depresí.

Očekávaný **vztah mezi užíváním emočně nabitých slov a depresí** byl na našem výzkumném souboru a českém jazykovém podkladě potvrzen: míra deprese (počet příznaků deprese) korelovala podle očekávání s výskytem slov s negativním emočním významem v pozitivních textech (motivační dopis a dopis z dovolené) v celém vzorku, u mužů také v neformálním negativním textu – dopise s omluvou. Pozitivních emočních slov užívaly osoby s depresí v pozitivních textech a Omluvě naopak méně, s rostoucím počtem příznaků deprese klesal výskyt emočně pozitivních slov. Tyto lingvistické jevy pravděpodobně odráží prožívání osob popsané psychologickými modely deprese. Podle Lewinsohna (MacPhillamy & Lewinsohn, 1974) získávají depresivní osoby z událostí méně potěšení. To by vysvětlovalo zejména méně pozitivně emočních slov a více negativně emočních slov v dopise z dovolené, kde má pisatel popisovat pozitivní zážitek, na který chce nalákat blízkou osobu. Beckova kognitivní triáda, podle které jedinec negativně nazírá na sebe sama a okolní svět, může dobře vysvětlit menší počet pozitivně emočních slov jak v dopise z dovolené, tak v motivačním dopise, který je zaměřen na osobu pisatele.

Zajímavé trendy týkající se emočně nabitých slov se ukázaly v textu stížnost. U žen kladně koreloval počet pozitivních emočně nabitých výrazů s příznaky deprese, zatímco v ostatních textech měl tento korelační vztah opačné znaménko. Rovněž rozdíly mezi klinickým a kvótním souborem v užívání emočních slov jsou ve stížnosti opačného charakteru než ve všech ostatních textech. Depresivní ženy v tomto jednom druhu textu užívaly více pozitivních slov a méně negativních slov než ženy nedeprativní, depresivní muži pak užívali při psaní stížnosti méně negativních slov. Odlišnost detekovaného vztahu mezi emočně

nabítkými výrazy a depresí v textu stížnost by mohla odrážet specifikum tohoto textu: scénář stížnosti předpokládá napsání textu formálního charakteru zaměřeného negativně směrem k jiné osobě nebo okolí pisatele. Výsledek je pak možné interpretovat jako paradoxní důsledek negativního sebezaměření osob s depresí (viz např. Beckův kognitivní model deprese): depresivní osoby jsou pak umírněnější v projevování negativních emocí a hodnocení zaměřeném na druhé lidi z toho důvodu, že negativní zaměření směřují pouze či především na sebe samotné. Zadání evokující odklon negativního usuzování od sebe směrem k druhým možná vede k narušení depresivní sebezaměřené pozornosti a nevědomému sklonu vzniklou disonanci neutralizovat pozitivním zaměřením na druhé. Je-li tomu tak, potom by u osob s depresí mělo být pozorováno základní kognitivní nastavení známé z transakční analýzy jako schéma: „já nejsem ok, ty jsi ok“ (Harris, 2002). Opačný trend v užití pozitivních a negativních výrazů ve stížnosti možná lze interpretovat také na podkladě atribuční teorie deprese. Gladstone a Kaslow (1995) zjistili, že k rozvoji deprese významně přispívá atribuční styl, v rámci kterého mají depresivní osoby sklon u negativních událostí hledat chybu u sebe („za neúspěch si mohu sám“), zobecňovat („cokoliv dělám, špatně skončí“) a za projev stálých charakteristik („jsem špatný“), naopak u pozitivních událostí hledají příčinu mimo sebe a považují je za ojedinělý a náhodný jev („měl jsem štěstí“). Pokud tedy v dopise stížnost byla negativní událost a priori spojena s druhými lidmi, nikoli s pisatelem, zadání bylo v rozporu s atribučním myšlenkovým schématem (příčina negativní události je dle scénáře evidentně mimo působnost pisatele), což teoreticky mohlo u depresivních osob vést k nevědomému sklonu vyhnout se negativním výrazům a posílit podíl pozitivních výrazů v textu.

Očekávaný **vztah mezi užíváním zájmen v první osobě a depresí** nebyl na našem výzkumném souboru a českém jazykovém podkladě potvrzen. Vztah byl očekáván na základě výsledků zahraničních studií a podpořen teorií deprese Pyszczynského a Greenberga (1987), podle které je deprese spojena s na sebe zaměřenou pozorností. Očekávaný vztah byl nalezen pouze v textu stížnost u žen, ve kterém ženy z klinického souboru užívaly ve statisticky významně větším poměru

přivlastňovací osobní zájmena v singuláru nežli ženy z kvótního výběru, avšak v žádném jiném textu nebyl rozdíl potvrzen a v žádném textu mužů ani žen nebyl potvrzen korelační vztah s mírou aktuálně prožívané (počtem příznaků) deprese. Nabízí se vysvětlení, že čeština neužívá osobní zájmena takovým způsobem jako angličtina nebo jiné germánské jazyky, protože v češtině je podmět často vyjadřován osobou slovesa. Toto vysvětlení by bylo dostačující, kdyby osoby s depresemi nebo úzkostmi častěji užívaly první osobu. Mezi výzkumnou a kontrolní skupinou však nebyl zjištěn rozdíl v užívání slov v první osobě v žádném z textů.

Dalším z možných důvodů, proč se v našich textech neprojevovalo sebezaměření v podobě zvýšeného užívání osobních zájmen singuláru, by mohla být specifická zadání textů, která se od textů používaných v zahraničních výzkumech liší tím, že jde o dopisy adresované druhé osobě s cílem něco druhé osobě sdělit nebo ji ovlivnit, například ji přesvědčit, aby za pisatelem přijela v dopise z dovolené, požádat o pomoc ve stížnosti nebo požádat o odpuštění v dopise s omluvou. V jiných výzkumech byly obvykle zkoumány autobiografické vzpomínky nebo texty a rozhovory volnějšího zadání a rovněž spíše autobiografického charakteru (Brockmeyer et al., 2013; Rude, Gortner, & Pennebaker, 2004). Je možné, že v textech adresovaných konkrétní osobě (například příteli, řediteli firmy atd.) se depresivním osobám podařilo na sebe zaměřenou pozornost odklonit. Navíc, některé studie naznačují, že i na sebe zaměřený, například autobiografický text, je třeba patřičně detailně specifikovat, aby se očekávaný vztah projevil. Například Jarrold et al. (2011) zkoumající tzv. seberefrenční slova u malého vzorku starších mužů zjistili větší výskyt těchto slov u depresivní skupiny pouze v souvislosti s jednou konkrétní autobiografickou otázkou z celého rozhovoru.

Roli při detekování vztahu mezi užíváním osobních zájmen singuláru a depresí může hrát i rozlišování a specifikace různých poddruhů deprese. Zimmermann et al. (2017), kteří analyzovali přepis polostrukturovaného rozhovoru s hospitalizovanými pacienty, nezjistili žádnou souvislost mezi současně zjištěnými depresivními příznaky a zájmeny, ale až mezi výskytem příznaků po propuštění a objektivním a přivlast-

ňovacím zájmenem. Brockmeyera et al. (2015) předpokládali větší výskyt zájmena v první osobě jednotného čísla u osob s chronickou depresí, ale zjistili jej pouze u osob s nechronickou depresí. V naší studii byla u klinické populace sledována deprese jako součást klinického obrazu duševních poruch (depresivní poruchy, úzkostné poruchy, závislost na alkoholu) a charakter deprese nebyl blíže specifikován (nebylo zaznamenáváno, zda se jedná o chronickou depresí, či nikoli), exaktně měřena byla „pouze“ míra aktuálně prožívané deprese. Zdá se tedy, že pro detekování vztahu mezi užíváním osobních zájmen v singuláru a depresí je nezbytné depresi zevrubněji kategorizovat.

Explorace provedená na textech v českém jazyce ukazuje na existenci vztahu mezi depresí a dalšími textovými proměnnými. Například českými texty se ukázal jako průkazný **vztah mezi osobními zájmeny (obecně, nikoli pouze v singuláru), neurčitými číslovkami a depresí**. Průkazně odlišně často je užívaly osoby z klinického souboru oproti kvótnímu souboru, statisticky průkazná byla korelace mezi užíváním těchto proměnných a aktuálně prožívanou mírou deprese. Tyto vztahy se podle výsledků naší studie jeví jako nezávislé na registru (byly detekovány ve všech textech) i na pohlaví.

Jeden z nejsilnějších vztahů vykazovalo užívání určitých druhů sloves ve spojitosti s depresí. **Méně užívaly depresivní osoby slovesných tvarů v přítomném nebo budoucím čase a tvary minulého aktivního přičestí**, tj. více využívaly jiné podoby minulého času. Užívání těchto slovesných kategorií klesalo s přibýváním depresivních příznaků zejména v textech žen (zde byly vztahy silnější). Je zajímavé, že celkový počet sloves spolu s depresivními příznaky stoupal, a to hlavně u mužů. Klíčové pro vztah k depresi je tedy časování sloves – s rostoucí depresí se snižuje užívání přítomného a budoucího času v poměru k minulému. Tento výsledek logicky koresponduje s charakteristickým pohledem na budoucnost osob s depresí. Například Beckova (1979) kognitivní triáda deprese ve třetím bodu popisuje charakteristický negativní pohled osob s depresí na budoucnost: budoucnost je buď vnímána negativně („nic nezvládnou, nečeká mne nic dobrého, všechno, co bude, bude ještě horší než dnes“), anebo není vnímána vůbec. Tomuto kognitivnímu nastavení v jazyce odpovídá nízký poměr sloves v čase přítomném a budoucím.



Negativní vztah k depresi dále vykazovaly předložky: **počet předložek s rostoucím skóre aktuálně prožívané míry deprese klesal**, klinický soubor používal předložek méně než kvótní. Nalezený vztah patří ke dvěma nejsilnějším. Předložky jsou neohebný slovní druh, který v češtině stojí vždy před podstatným jménem, zájmenem, číslovkou nebo přídavným jménem. Jedná se o synsémantikum, tj. slovní druh, který sám o sobě nenese vlastní význam, ten vzniká až spojením s dalším slovním druhem (Cvrček a kol., 2015). Předložky plní funkci vyjadřování bližších okolností, konkretizují smysl a význam zejména podstatných jmen (například je značný rozdíl být před lesem, za lesem nebo v lese). V souladu s touto základní lingvistickou funkcí předložek Tausczik a Pennebaker (2010) považují výskyt předložek za výraz konkrétnosti. Čím více předložek, tím konkrétnější a přesnější bývá sdílená informace. Rozdíl ilustrujeme na jednoduchém příkladu dvojic výroků: (1a) *preferuji neseďavé zaměstnání*, (1b) *preferuji zaměstnání s možností volného pohybu*, (2a) *je tam pláž a spousty turistů*, (2b) *je tam pláž a na ní spousty turistů*, (3a) *byl slyšet křik a cinkot lahví*, (3b) *v domě byl slyšet křik a cinkot lahví*, (4a) *chtěl bych o tom ještě jednou promluvit*, (4b) *chtěl bych s tebou o tom ještě jednou promluvit*. Konkrétnost výsledného textu může odrážet dominanci specifických druhů kognice (například zaměření pozornosti na konkrétní reálné jevy, převahu konkrétního myšlení oproti abstraktnímu) a ukotvenost v realitě (ne-bytí tady a nyní). Neurofyziologicky by se mělo odlišné zaměření pozornosti projevit v činnosti prefrontální kůry (Dušek & Večeřová-Procházková, 2017). Předložky tak ukazují, že **s rostoucí mírou aktuálně prožívané deprese klesá konkrétnost a specifčnost textu, potažmo kognitivních procesů**.

Této interpretaci nasvědčují i další jazykové jevy odhalené v textech pisatelů s depresí a uvedené výše: zvýšené užívání osobních zájmen a neurčitých číslovek. Osobní zájmena jsou slovním druhem, který ve větě plní zpravidla funkci podmětu, předmětu nebo přívlastku, přičemž obvykle zastupuje podstatné nebo přídavné jméno (odtud i označení zájmeno – zastupuje nějaké konkrétní jméno). Neurčité číslovky jsou slovní druh, který označuje slova číselného významu, avšak na rozdíl od určitých číslovek neurčité číslovky nevyjadřují přesný, ale jen přibližný, relativní počet (například několik lidí, málo zkušeností, mno-

ho starostí). Nízká míra konkrétnosti je tak v obou slovních druzích imanentně obsažena. Snížená schopnost nebo neochota se konkrétně a přesně vyjádřit by mohla také být důvodem k uchýlení se k použití frazémů, jejichž užití vykazovalo těsnou souvislost s depresí (frazémy častěji užívali muži z klinického souboru v textu stížnost, ženy z klinického souboru v dopise z dovolené, u mužů navíc průkazně koreloval výskyt frazémů s mírou aktuálně prožívané deprese v obou formálních textech, tj. ve stížnosti i v Žádosti o zaměstnání). **Zvýšené užívání osobních zájmen, neurčitých číslovek a frazémů podporuje vyslovenou teorii o vztahu mezi depresí a sníženou konkrétností jazykových prostředků (fiktivního) textu.**

Další souvislost nalézáme mezi **depresí a sníženým užíváním vokativu a druhé osoby**. Oba tyto vztahy jsou podle našeho názoru projevem převahy na sebe zaměřené pozornosti a omezení projevů vztahu k okolí, které jsou popisovány jako symptomy deprese (Beck, 1970, Pyszczynski & Greenberg, 1987, Dušek & Večeřová-Procházková, 2010). Vokativ, tedy oslovení, označuje osobu, na kterou se pisatel se svým sdělením obrací. Oslovit či titulovat danou osobu lze pomocí funkce (pane řediteli), jména nebo jiným titulem (drahá slečno). V případě neformálního oslovení může použitý výraz vyjadřovat vztah mluvčího či pisatele k dané osobě (např. zdvořilost jména, oslovení *kamaráde* nebo *můj drahý*). Vynecháním oslovení se pisatel vzdává možnosti vyjádřit vzájemný vztah. Respondenti z klinického souboru užívali průkazně méně oslovení v dopisech z dovolené, muži též v motivačních dopisech (tento vztah byl potvrzen i negativní korelací výskytu vokativu s rostoucí mírou aktuálně prožívané deprese). Analogicky se projevuje deprese nižším užíváním druhé osoby. Rozdíly v užívání mezi klinickou a kvótní populací se projevily v dopisech z dovolené, kde muži i ženy výzkumné skupiny užívali druhou osobu méně. Menší počet slov v druhé osobě koresponduje s poznatky rumunské studie Trifu et al. (2017), kteří zjistili méně zájmen v druhé osobě v textech osob s depresí. Snížené užívání vokativu a druhé osoby v závislosti na míře deprese může být **odrazem na sebe zaměřené pozornosti a sníženého vztahu k okolí, tj. popisovaných projevů deprese.**

Jako výrazná intervenující proměnná se v této práci potvrdilo pohlaví. Muži se stoupající četností depresivních příznaků užívali méně přídavných jmen a více sloves, což se promítlo také do velikostí hodnot indexů, které jsou z těchto kategorií počítány. *Trager index*, *readiness to action index* a *aggressiveness index* pozitivně korelovaly s příznaky deprese ve všech nebo (v případě *aggressiveness indexu* téměř ve všech) textech mužů a v textu stížnost dosahovaly korelace střední síly. Tyto indexy mají odrážet emocionalitu pisatele a jeho sdělení, konkrétně *trager index* souvisí s emocionální stabilitou a *aggressiveness* vyjadřuje emocionální vzrušení a připravenost k jednání (Sboev, Gudovskikh, Rybka, & Moloshnikov, 2015). U žen dva z těchto indexů korelovaly pouze ve stížnosti, a to s malou silou vztahu. Navíc ženy výzkumné skupiny osob s depresemi a úzkostmi měly v dopisech s omluvou nižší skóre těchto indexů než kontrolní skupina. Zdá se, že u žen se emocionální nestabilita projevuje jiným způsobem, který tolik nesovisí s vyjadřováním akce a připraveností jednat.

Rovněž index formality koreloval častěji a s větší silou v textech mužů než žen, jednalo se o negativní korelaci – s četností příznaků formalita klesala. Formální řeč se podle Mairesse, Walkerové, Mehla a Moora (2007) vyznačuje tím, že je méně nejednoznačná. Pokles formality s rostoucími depresivními příznaky tedy odpovídá trendu naznačenému menším počtem předložek. Vyšší index koherence u klinické skupiny byl zjištěný ve všech textech mužů, ale i neformálních textech žen. Ruská studie (Litvinova, Seredin, Litvinova, & Zagorovskaya, 2016) zjistila vyšší hodnoty tohoto koeficientu u osob s vyšším rizikem sebedestruktivního chování, které se mimo jiné vyznačovaly depresivitou a emoční labilitou.

Výrazný vliv na jednotlivé textové parametry měl kromě pohlaví také charakter textu podle jeho zadání. To můžeme vidět i u logistických modelů, které měly predikovat depresi (skóre DASS-D  $\geq 6$ ) na základě těchto proměnných. Každý z celkem osmi modelů zahrnuje odlišné textové parametry, přestože se některé lingvistické kategorie, např. slova s emočním významem nebo předložky, opakují ve více modelech. Nejsilnějším prediktivním modelem u žen je model pro druhý text – dopis z dovolené s  $R^2=0,4$  a s nejvyššími hodnotami *precision*

a *recall* a jejich harmonickým průměrem ( $F= 0,68$ ). Tento model ale zahrnoval také největší počet lingvistických proměnných. U mužů není situace tak jednoznačná. Model pro třetí text - stížnost má nejvyšší hodnoty *precision* a *recall* ( $F= 0,73$ ), jeho  $R^2=0,35$  je však o něco nižší než u modelů pro text 1 a 4. Ty mají však značně horší senzitivitu, tedy procento depresivních osob, které dokázal model predikovat.

Celkově nás získané výsledky opravňují k podpoře základního předpokladu, že osobnost a její psychické stavy se promítají do způsobu nakládání s jazykem, jejího jazykového otisku, idiolektu. Podrobné analýzy umožňují detekovat omezené množství lingvistických charakteristik idiolektu, které mají těsný vztah ke sledovaným afektivním poruchám (depresi). Vztah je však vždy ovlivněn minimálně dvěma významnými proměnnými: řečovým registrem, který se váže k dané komunikační situaci (v našem výzkumu daném zadáním psaného textu) a pohlavím pisatele.

### 3.4.5.5 Výsledky

---

Získané výsledky nás opravňují k formulaci následujících tvrzení o vztazích mezi aktuálně prožívanou mírou deprese (DASS\_D) a textovými charakteristikami:

1. **Aktuálně prožívaná míra deprese se promítá do textu.** Existenci vztahu potvrzují výsledky naší studie, stejně jako řada zahraničních studií.
2. **Transjazykově spolehlivým indikátorem deprese v textu je přítomnost emočně zabarvených slov.** Výsledky studie na českém jazykovém podkladu se shodují se zahraničními. Lépe se tento vztah projevuje u textů produkovaných na základě pozitivních scénářů (žádost o zaměstnání a dopis z dovolené).
3. Na základě neobsahových, pouze morfo-syntaktických textových proměnných lze s akceptovanou mírou chyby odlišit klinickou populaci (s diagnostikovanou poruchou nálad) od neklinické. **Trans-textově a trans-genderově jsou v českém textu spolehlivým indikátorem deprese osobní zájmena a neurčité**

**číslovky.** Míra aktuálně prožívané deprese koreluje transtextově a transgenderově s výskytem dvou jmenovaných textových proměnných a navíc s **užíváním předložek** v textech.

4. Některé textové charakteristiky pravděpodobně odrážejí **odklon zaměření pozornosti osob s depresí směrem od budoucnosti a přítomnosti** (vyšší poměr sloves v minulém čase bez aktivního přičestí).
5. **S rostoucí mírou aktuálně prožívané deprese klesá konkrétnost a specifičnost textu** (i kognice). Textové charakteristiky, které reflektují tento vztah, jsou: klesající počet předložek, osobních zájmen, neurčitých číslovek a vokativu v textu a rostoucí počet frazémů.
6. **S rostoucí mírou aktuálně prožívané deprese klesá používání výrazů vyjadřujících vztah k druhým.** Textové charakteristiky, z nichž odvozujeme tuto souvislost, jsou užívání vokativu a druhé osoby. Tyto výsledky korespondují s výsledky některých zahraničních studií.
7. Korelační koeficienty mezi skórem ve škále deprese a jednotlivými lingvistickými charakteristikami textů ukazují na **značné mezipohlavní diference.**
8. **Lze vytvořit prediktivní modely,** které na základě vybraných morfo-syntaktických charakteristik textu detekují ohroženou populaci s vysokou mírou aktuálně prožívané deprese. Celková prediktivní síla vytvořených modelů dosahuje průměrné hodnoty 70 %. Na základě této hodnoty lze konstatovat, že nejsilnější prediktivní model deprese pro muže je (poněkud překvapivě) z formálního řečového registru textu stížnosti, pro ženy z neformálního řečového registru textu dopisu z dovolené.

## 4 Závěr

V předchozím textu byla podrobně popsána východiska projektu CPACT, jeho design i výstupy několika studií, které v něm byly realizovány. Jedná se samozřejmě pouze o výseč z celé problematiky, která časově spadá především do třetího roku řešení projektu (tj. 2018). Následující období bychom proto rádi věnovali přípravě a publikaci dalších textů, které původní několikaletý horizont projektu výrazně přesáhnou a přinesou nové informace k oblastem, kterým jsme se doposud jen málo věnovali. Vzhledem k ohromnému množství dat získaných v rámci výzkumu by mohla být tato snaha velkým přínosem pro rozvoj psycholingvistického bádání nejen v českém či evropském kontextu.

Je vhodné zmínit, že na CPACT v tomto směru bezprostředně navazují i plány dalších projektů. Jedním z nich je výzkum „Analýza verbální a nonverbální komunikace v kontextu osobnostních charakteristik komunikující osoby“, který je připravován ve spolupráci s Fakultou elektrotechnickou Českého vysokého učení technického v Praze a Fakultou aplikovaných věd Západočeské univerzity v Plzni. Jedná se o základní výzkum zaměřený na studium vlivu psychologicky relevantních charakteristik komunikujících osob nejen na podobu jejich verbální, ale nově i nonverbální komunikace. Další připravované projekty se zaměřují např. na počítačnou analýzu textů ve směru k predikci úspěšnosti léčby hospitalizovaných osob, případně predikci rizikového chování u pachatelů trestné činnosti.

Projekt CPACT je rovněž spojen s řadou publikovaných i připravovaných kvalifikačních prací studentů, zejm. na pracovištích členů projektového týmu. Mezi názvy a témata prací nalezneme například: Psychologické koreláty morfologických charakteristik textu u klinické populace; Kvantitativní psycholingvistická analýza textů produkovaných v kontextu anxiety a deprese; Specifika v jazykovém projevu mužů v protialkoholní léčbě; Diskrepance mezi dotazníkovým sebeposouzením a posouzením druhou osobou a její kvalitativní reflexe rozborem textu posuzovaného; Jak dobře se vzájemně známe? Míra shody mezi

sebeuposouzením a posouzením osobnosti; Kvantitativní lingvistická analýza psaného textu ve vztahu k osobnostním rysům pisatele se zaměřením na smysl pro humor; Analýza laických odhadů rysů osobnosti pisatele na základě morfologických charakteristik neformálních textů; Analýza laických odhadů rysů osobnosti na základě analýzy přepisu fiktivního přijímacího pohovoru; Možnosti predikce motivační úrovně z textu; Smysl pro humor a depresivita či Fenomén negace a jeho psychologické konsekvence.

Výše zmíněné výstupy, projekty a plány odkazují k ohromnému výzkumnému potenciálu, který počítačnická psycholinguvistická analýza textu nabízí. Určitým limitujícím faktorem je však množství času a energie, který sběr materiálů a příprava/nastavení metod zabírá. Zvláště náročným procesem je například příprava ústních komunikátů, resp. jejich přepis, který, vzhledem ke značné chybovosti, nemohou zajistit momentálně dostupné aplikace. V tomto směru bude tedy velmi přínosný vývoj softwaru, který umožní efektivní automatizovaný přepis češtiny. Dalším limitem je pak samozřejmě sběr psychologických dat, resp. jejich formát a relevance. Některé současné psychodiagnostické metody bývají kritizovány za problematickou validitu (zejm. predikční) nebo např. umělé vymezení psychologických konstruktů, na které jsou zaměřeny. Bylo by jistě vhodné do budoucna využívat pouze ty metody, které budou ve vztahu k chování a prožívání respondentů maximálně relevantní a které zahrnou i různé perspektivy, jimiž je možné na člověka nahlížet (viz asymetrie posouzení).

Bylo by zde možné zmínit ještě mnoho věcí, které jsou ve vztahu k našemu výzkumu, a celkově psycholinguvistickému bádání, perspektivní, či naopak limitující. Je ale zřejmé, že se tento směr, přinejmenším v českém kontextu, teprve formuje a že narůstající pozornost odborné i laické veřejnosti nepochybně dodá velký impuls k jeho rozvoji a proměně. Už jen z tohoto důvodu stojí za to dál bádát a „hledat mezi lidmi a slovy“.

*Dalibor Kučera*

# 5 Literatura



Adam, R. (2009). Za empiricky adekvátní teorii písemné komunikace. *Slovo a slovesnost*, 69, 221–231.

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science*, 7, 32–39.

Allik, J., Realo, A., Mõttus, R., Esko, T., Pullat, J., & Metspalu, A. (2010). Variance determines self-observer agreement on the Big Five personality traits. *Journal of Research in Personality*, 44(4), 421–426.

Allport, G. W. (1937). Personality: A psychological interpretation.

Altmann, G. T. (2006). History of psycholinguistics. *The encyclopedia of language and linguistics*, 257–265.

Andersen, S. M., Glassman, N. S., & Gold, D. A. (1998). Mental representations of the self, significant others, and nonsignificant others: Structure and processing of private and public aspects. *Journal of Personality and Social Psychology*, 75, 845–861

Arbib M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–124.

Argamon, S., Koppel, M., Pennebaker, J., & Schler, J. (2007) Mining the blogosphere: age, gender, and the varieties of self-expression. *First Monday*, 10/2007, 12.

Balaščíková, V., & Blatný, M. (2003). Determinanty výběru strategií zvládnání zátěže: osobnost, vnímání situace, sebepojetí. *Zprávy – Psychologický ústav AV ČR*, 9(2),1–19.

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, 17, 327–335.

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, 18, 359–372.

Bartoň, T., Cvrček, V., Čermák, F., Jelínek, T., & Petkevič, V. (2009). *Statistika češtiny*. Praha: Nakladatelství Lidové noviny.

Beer, A., & Vazire, s. (2017). Evaluating the predictive validity of personality trait judgments using a naturalistic behavioral criterion: a preliminary test of the self-other knowledge asymmetry model. *Journal of Research in Personality*, 70, 107–121.

Beer, A., & Watson, D. (2008). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality*, 76(3), 535–559.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

Berelson, B. (1952). *Content analysis in communication research*. Glencoe: Free Press.

Bernard, J. D., Baddeley, J. L., Rodriguez, B. F., & Burke, P. A. (2016). Depression, Language, and Affect: An Examination of the Influence of Baseline Depression and Affect Induction on Language. *Journal of Language and Social Psychology*, 35(3), 317–326.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: University Press.

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational linguistics*, 19(2), 219–241.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92(1), 119.

Bjekić, J., Lazarević, L., Erić, M., Stojimirović, E., & Đokić, T. (2012). Development of Serbian dictionary for automatic text analysis (LIWCser). *Psihološka istraživanja* 15(1), 85–110.

Bland, J. M., & Altman, D. G. (1994). Statistics notes: Some examples of regression towards the mean. *Bmj*, 309(6957), 780.

Blatný, M. (2010). 100 let výzkumu osobnosti v psychologii. In M. Blatný, M. Hřebíčková, K. Millová, A. Plháková, P. Říčan, A. Slezáčková, & I. Stuchlíková, *Psychologie osobnosti: Hlavní témata, současné přístupy* (s. 11–20). Praha: Grada.

Blatný, M., Hřebíčková, M., Millová, K., Říčan, P., Slezáčková, A., & Stuchlíková, I. (2006). *Psychologie osobnosti. Hlavní témata, současné přístupy*. Praha: Grada.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., Jimeno Yepes, A., Koehn, P., Logacheva, V., Monz, C., Negri, M., Neveol, A., Neves, M., Popel, M., Post, M., Rubino, R., Scarton, C., Specia, L., Turchi, M., Verspoor, K., Zampieri, M. (2016). Findings of the 2016 Conference on Machine Translation (WMT16). *Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers*. Berlin, Germany: Association for Computational Linguistics, s. 131–198.

Brockmeyer, T., Zimmermann, J., Kulesa, D., Hautzinger, M., Bents, H., Friederich, H. C., ... & Backenstrass, M. (2015). Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in psychology*, 6, 1564.

Brucks, M. S., Holtzman, N. S., Back, M. D., Pennebaker, J. W., & Mehl, M. R. (2015). Narcissism and the use of personal pronouns revisited. *Journal of personality and social psychology*, 109(3), e1-e15.

Burato, K. R. S. B., Crippa, J. A. S., & Loureiro, S. R. (2009). Validity and reliability of the social phobia safety behaviour scale in social anxiety. *Rev. Psiq. Clín.*, 6(5), 175-81.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (s. 3-52). San Diego, CA: Academic Press.

Carey, A. L., Brucks, M. S., Küfner, A. C., Holtzman, N. S., Back, M. D., Donnellan, M. B., Carley, K. (1993). Coding Choices for Textual Analysis: A Comparison of content analysis and map analysis. In *Sociolog. Methodology*, 23, 75-126.

Cegala, D. J. (1989). A study of selected linguistic components of involvement in interaction. *Western Journal of Communication (includes Communication Reports)*, 53(3), 311-326.

Cimrman, J., & Smoljak, L. (1992). *Divadlo Járy Cimrmana: Vyšetřování ztráty třídní knihy (Vol. 2)*. Paseka.

Cleuziou, G., & Poudat, C. (2007, February). On the impact of lexical and linguistic features in genre-and domain-based categorization. In *International Conference on Intelligent Text Processing and Computational Linguistics* (s. 599-610). Berlin, Heidelberg: Springer.

Cohen, A. S., Minor, K. S., Baillie, L. E., & Dahir, A. M. (2008). Clarifying the linguistic signature: Measuring personality from natural speech. *Journal of Personality Assessment*, 90, 559–563.

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology*, 60, 884–894

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092–1122.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110–117.

Corral, A., Boleda, G., & Ferrer-i-Cancho, R. (2015). Zipf's law for word frequencies: Word forms versus lemmas in long texts. *PLoS one*, 10(7), e0129031.

Crocker, M. W. (2012). *Computational psycholinguistics: An interdisciplinary approach to the study of language (Vol. 20)*. Springer Science & Business Media.

Cvrček, V. & Chlumská, L. (2015). Simplification in translated Czech: a new approach to type-token ratio. *Russian linguistics* 39(3), 309–325.

Cvrček, V. et al. (2010). *Mluvnice současné češtiny 1*. Praha: Karolinum.

Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková A. & Zásina, A. J. (2018a). From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory*. [online]. DOI: <https://doi.org/10.1515/cllt-2018-0020>

Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A. & Zásina, A. J. (2018b). Variabilita češtiny: multidimenzionální analýza. *Slovo a slovesnost*, 78, 293–321.

Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A. & Zásina, A. J. (n.d.). *Register variability of elicited texts*. Manuscript in preparation.

Čechová, M., Krčmová, M. & Minářová, E. (2008). *Současná stylistika*. Praha: NLN.

Čermák, F. et al. (1983–2009). *Slovník české frazeologie a idiomatiky (SČFI), vol. 1–4*. Academia/Leda, Prague.

Čermák, F. (2009): *Slovník české frazeologie a idiomatiky*. Praha: Leda.

Čermák, F. (2014). Lexis in spoken and written language. In *Jazyk a slovník* (s. 299–303). Praha: Karolinum.

Čermák, F., Doležalová-Spoustová, D., Hlaváčová, J., Hnátková, M., Jelínek, T., Kocek, J., Kopřivová, M., Křen, M., Novotná, R., Petkevič, V., Schmiedtová, V., Skoumalová, H., Šulc, M., Velíšek, Z. (2005). *SYN2005: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupné z <http://www.korpus.cz>

Černý, J. (1996). *Dějiny lingvistiky*. Praha: Votobia.

ČSÚ – Český statistický úřad (2015). *Věk a vzdělání populace*. Dostupné z: <https://www.czso.cz>

Demjén, Z. (2014). Drowning in negativism, self-hate, doubt, madness: Linguistic insights into Sylvia Plath's experience of depression. *Communication & medicine*, 11(1), 41.

Dewaele, J. M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49(3), 509–544.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, 16, 81–106.

Dunning, D., & McElwee, R. (1995). Idiosyncratic Trait Definitions: Implications for Self-Description and Social Judgment. *Journal of Personality and Social Psychology*, 68(5), 936–946.

Durkheim, É. (1951). *Suicide: A study in sociology*. Glencoe, IL: Free Press. Dostupné z <https://archive.org/details/suicidestudyinsoodurk>

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1–23.

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A Package for Computational Text Analysis. *R journal*, 8(1).

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, 94, 334–346.

Fiedler, E. R., Oltmanns, T. F., & Turkheimer, E. (2004). Traits associated with personality disorders and adjustment to military life: Predictive validity of self and peer reports. *Military Medicine*, 169, 207–211.

Figurová, L. T. (2007). *Sociální desirabilita* (disertační práce). Brno: PsÚ FF, Masarykova univerzita.

Finkel, E. J., Simpson, J. A., & Eastwick, P. W. (2017). The psychology of close relationships: Fourteen core principles. *Annual review of psychology*, 68, 383–411.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2001). The realistic accuracy model and Brunswik's approach to social judgment. *The essential Brunswik: Beginnings, explications, applications*, 365-369

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177-182.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, 60, 773–794.

Funder, D. C., & West, Ss. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*, 61(4), 457-476.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personal. Individ. Differ.*, 7, 385–400.

Furnham, A. (1990). Language and personality. In Giles, H., Robinson, W.P. (Eds.), *Handbook of Language and Social Psychology*, New York: Wiley, s. 73-95.

Gill, A. J., Nowson, S., & Oberlander, J. (2009). What are they blogging about? Personality, topic and motivation in blogs. *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009*. San Jose, California, USA, May 17-20, 2009.

Gill, A. J., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences*, 40(3), 497–507.

Gill, A., & Oberlander, J. (2003). Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself, but Neuroticism is more of a worry. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 456-461. Boston, July 31-August 2, 2003.

Gill, J., Oberlander, J. & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences*, 40, 497-507.

Gleser, G. C., Gottschalk, L. A., & John, W. (1959). The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology*, 15, 183–191.

Golcher, F., & Reznicek, M. (2011). Stylometry and the interplay of topic and L1 in the different annotation layers in the FALKO corpus. *QITL-4 - Proceedings of Quantitative Investigations in Theoretical Linguistics 4*, 29. 03. 2011 - 31. 03. 2011, Berlin, Humboldt-Universität, 29-34.

Gottschalk, L. A., & Gleser, G. C. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Univ of California Press.

Gravetter, F. J., & Forzano, L. A. B. (Eds.) (2011). *Research Methods for the Behavioral Sciences, 4th ed.* Belmont: Cengage Learning.

Grice, J. W. (2001). Computing and Evaluating Factor Scores. *Psychological Methods*, 6(4), 430-450.

Griffin, D., & Buehler, R. (1993). Role of construal processes in conformity and dissent. *Journal of Personality and Social Psychology*, 65, 657-669

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (s. 319-359). San Diego, CA: Academic Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.

Hajič, J. (2001). *Disambiguation of Rich Inflection*, Karolinum, Praha.

Hajič, J. (2004). *Complex Corpus Annotation: The Prague Dependency Treebank*. Jazykovedný ústav L. Štúra, SAV, Bratislava, Slovensko.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42(6), 1476-1489.

Hampson, S. E. (2014). *Advances in personality psychology* (vol. 1). East Sussex, UK: Psychology Press.

Haroche, C., Henry, P., & Pêcheux, M. (1971). La sémantique et la coupure saussurienne: langue, langage, discours. *Langages*, 24, 93-106.

Hart, R. P. (1984). Systematic analysis of political discourse: The development of DICTION. *Political Communication Yearbook*, 1, 97-134.

Hartl, P., & Hartlová, H. (2000). *Psychologický slovník*. Praha: Portál.

Havigerová, J. M., Haviger, J., & Kučera, D. (2016). *Manuál ke škálování textů pro výzkum CPACT*. Unpublished manuscript.

Hayes, N. (2000). *Aplikovaná psychologie*. Praha: Portál.

Hendl, J. (2004). *Přehled statistických metod zpracování dat: Analýza a metaanalýza dat*. Praha: Portál.

Heylighen, F., & Dewaele, J. M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, 7(3), 293-340.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality*, 43, 524-527.

Hnátková, M. (2002). Značkování frazémů a idiomů v Českém národním korpusu s pomocí Slovníku české frazeologie a idiomatiky. *Slovo a slovesnost* 63(2), 117-126.

Hnátková, M. (2011). Výsledky automatického vyhledávání frazémů v autorských korpusech. In V. Petkevič & A. Rosen, *Korpusová lingvistika Praha 2011. 3 Gramatika a značkování korpusů*. Praha, Nakladatelství Lidové noviny, 2011, s. 171-185.

Hoffmannová, J., Homoláč, J., Chvalovská, E., Jílková, L., Kaderka, P., Mareš, P. & Mrázková, K. (2016). *Stylistika mluvené a psané češtiny*. Praha: Academia.

Holland, P. W., & St Wainer, H. (1993). *Differential item functioning*. Hillsdale, New Jersey: Lawrence Erlbaum.

Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of research in personality*, 45(1), 92-99.

Holtgraves, T. M. (Ed.). (2014). *The Oxford handbook of language and social psychology*. Oxford Library of Psychology.

Holtzman, N. S. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68, 63-68.

Hornová, L. (2003). *Referenční slovník gramatických termínů*. Olomouc: Univerzita Palackého v Olomouci.

Hřebíčková, M. (2003). Metodologické souvislosti výzkumu shody mezi sebezposouzením a posouzením druhými a její moderátory. *Československá psychologie*, 47(6), 533-547.

Hřebíčková, M. (2011). *Pětifaktorový model v psychologii osobnosti: přístupy, diagnostika, uplatnění*. Praha: Grada.

Hřebíčková, M., Jelínek, M., Blatný, M., Brom, C., Burešová, I., Graf, S., & Zábrodská, K. (2016). Big Five Inventory: Základní psychometrické charakteristiky české verze BFI-44 A BFI-10. *Československá Psychologie*, 60(6), 567.



Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, 36(11), 1514-1528.

Cheng, W. (2011). *Exploring corpus linguistics: Language in action*. New York: Routledge.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of research in personality*, 42(1), 96-132.

Chung, C.K., & Pennebaker, J. W. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social communication: Frontiers of social psychology*. NY: Psychology Press, 343-359.

Iacobelli, F., Gill, A., Nowson, S., & Oberlander, J. (2009). Classification of personality in large blog data. In *Proceedings of 19th Annual Meeting of the Society for Text and Discourse*. Rotterdam, The Netherlands.

Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker. *The Oxford handbook of language and social psychology*, 201-237.

Jakovljević, I., & Milin, P. (2017) The relationship between thematic, lexical, and syntactic features of written texts and personality traits. *Psihologija*, 50(1), 67-84.

Jan, P. (2011). *Dětská řeč a komunikace: poznatky vývojové psycholinguistiky*. Praha: Grada Publishing, a.s.

Jarrold, W., Javitz, H. S., Krasnow, R., Peintner, B., Yeh, E., Swan, G. E., & Mehl, M. (2011). Depression and self-focused language in structured interviews with older men. *Psychological reports*, 109(2), 686-700.

Jelínek, T. (2008). *Nové značkování v Českém národním korpusu*. Naše řeč 91(1), 13-20.

Jelínek, T., & Petkevič, V. (2011). Systém jazykového značkování současné psané češtiny. In Čermák F., *Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů*. Praha, Nakladatelství Lidové noviny, s. 154-170.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3(2), 114-158.

Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, and issues*. Scarborough: Nelson Education.

Karlík, P. et al. (1995). *Příruční mluvnice češtiny*. Praha: NLN.

Karlík, P., Nekula, M., & Pleskalová, J. (Eds.). (2002). *Encyklopedický slovník češtiny*. Praha: Lidové noviny.

Kartelj, A., Filipović, V., & Milutinović, V. (2012, May). Novel approaches to automated personality classification: Ideas and their potentials. In *MIPRO, 2012 Proceedings of the 35th International Convention* (s. 1017-1022). IEEE.

Kelley, H. H. (1982). *Personal relationships: Their structures and processes*. New Jersey: Lawrence Erlbaum.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8, 265–280.

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., Ustün, T. B., & Wang, P.S. (2009). The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiologia E Psichiatria Sociale*, 18(1), 23-33.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 567-83.

Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Res.*, 1, 15-29.

Kocovski, N. L., MacKenzie, M. B., Albiani, J. J., Battista, S. R., Noel, S., Fleming, J. E., & Antony, M. M. (2016). Safety behaviors and social anxiety: An examination of the social phobia safety behaviours scale. *Journal of Psychopathology and Behavioral Assessment*, 38(1), 87-100.

Kodýtek, V. (2008). Variace v mluvené češtině v Čechách: sonda do ORAL2006. In: M. Koprivová & M. Waclawičová (Eds.), *Čeština v mluveném korpusu* (s. 132-140). Praha: NLN.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, 64, 311–337.

Kondáš, O. et al. (1992). *Psychodiagnostika dospělých*. Martin: Osveta.

Koprivová, M., Lukeš, D., Komrsková, Z., Poukarová, P., Waclawičová, M., Benešová, L., & Křen, M. (2017). *ORAL: korpus neformální mluvené češtiny, verze 1 z 2*. 6. 2017. Ústav Českého národního korpusu FF UK, Praha 2017. Retrieved from <http://www.korpus.cz>

Krahmer, E. (2010). What computational linguists can learn from psychologists (and vice versa). *Computational linguistics*, 36(2), 285-294.

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P., & Zasina, A. (2015). *SYN2015: reprezentativní korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha. Retrieved from <http://www.korpus.cz>

Kubička, L., & Csémy, L. (1999). Psychosociální stresory, úzkost, deprese a úzus psychoaktivních látek v populaci pražských žen středního věku. *Psychiatrie*, 3(4), 219–227.

Kučera, D. (2017a). *Výzkum CPACT – Hledáme mezi lidmi a slovy*. Dostupné z <https://youtu.be/Uv--44mAC5E>

Kučera, D. (2017b). *Výzkum CPACT*. Dostupné z <http://www.facebook.com/vyzkumcpact>

Kučera, D., & Haviger, J. (2013, September). *Možnosti využití metod počítačnické lingvistiky v psychologii* [Possibilities of Application of Computational Linguistics Methods in Psychology]. Poster, SPO 2013, Třešť (CZ).

Kučera, D., & Haviger, J. (2014a, Juni). *Quantitative Psycholinguistic Analysis of Formal Parameters of Czech Text*. Poster, International Quantitative Linguistics Conference (QUALICO 2014), Olomouc (CZ).

Kučera, D., & Haviger, J. (2014b). Možnosti využití metod počítačnické lingvistiky v psychologii [Possibilities of Application of Computational Linguistics Methods in Psychology]. In M. Blatný et al. (Eds.), *Sborník příspěvků konference Sociální procesy a osobnost 2013: Včera, dnes a zítra*. Brno: Psychologický ústav AV ČR, v.v.i., s. 175-178.

Kučera, D., & Haviger, J. (2015). *Osobnost a text: Kvantitativní psycholinguistická analýza stylisticky diferencovaného českého textu* [Personality and Text: Quantitative Psycholinguistic Analysis of the Stylistically Differentiated Czech Text]. Přednáška, Social Processes and Personality Conference 2015 (SPO), Brno (CZ).

Kučera, D., & Havigerová, J. (2014a). *Sociální charakteristiky komunikátora z pohledu kvantitativní psycholinguistické analýzy* [Social Characteristics of a Communicator from the Perspective of Quantitative Psycholinguistic Analysis]. Poster, Social Processes and Personality Conference 2014 (SPO), Stará Lesná (SK).

Kučera, D., & Havigerová, J. (2014b). *Application of Computational Linguistics Methods in the Assessment of a Communicator's Psychological Characteristics*. Přednáška, IACEP Conference 2014, Budapest (HU).

Kučera, D., & Havigerová, J. M. (2015a). Interpersonální charakteristiky komunikátora z pohledu kvantitativní psycholingvistické analýzy (předběžné sdělení z výzkumné studie). In M. Bozogáňová et al. (Eds.), *Sociálne procesy a osobnosť 2014: Človek a spoločnosť: Zborník príspevkov zo 17. ročníka medzinárodnej konferencie*. Košice: Spoločenskovedný ústav SAV, s. 267-274.

Kučera, D., & Havigerová, J. M. (2015b). Zhodnocení interpersonálních charakteristik pisatele pomocí grafometrické metody [The Assessment of Interpersonal Characteristics of a Writer Using Graphometric Methods]. In E. Maierová et al. (Eds.), *PhD Existence 2015: Česko-slovenská psychologická konference (nejen) pro doktorandy a o doktorandech (Sborník odborných příspěvků)*. Olomouc: FF UPOL, s. 272-279.

Kučera, D., & Havigerová, J. M. (2015c). Computational psycholinguistic analysis and its application in psychological assessment of college students. *Journal of Pedagogy*, 6(1), 61–72.

Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, 44, 427-435.

Kuhl, J., & Kazén, M. (2002). *PSSI–Inventář stylů osobnosti a poruch osobnosti*. (přeložil J. Švancara). Praha: Testcentrum.

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology*, 25, 272-285.

Kwantes, P. J., Derbentseva, N., Lam, Q., Vartanian, O., & Marmurek, H. H. (2016). Assessing the Big Five personality traits with latent semantic analysis. *Personality and individual differences*, 102, 229-233.

Lamers, F., van Oppen P., Comijs, H. C., Smit, J. H., Spinhoven, P., van Balkom, A. J., Nolen, W. A., Zitman, F. G., Beekman, A. T., & Penninx, B. W. (2011). Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Clinical Psychiatry*, 72(3), 341-348.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259-284.

Leary, T. (1957). *Interpersonal Diagnosis of Personality*. New York: Ronald Press.

Lee, C. H., Kim, K., Seo, Y. S., & Chung, C. K. (2007). The relations between personality and language use. *Journal of General Psychology*, 134, 405–413.

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology*, 46(6), 1349-1364.

Linkov, V., & Šmerk, P. (2016). Rozdíly mezi pravdivou a lživou online textovou komunikací (Differences Between True and Deceptive Online Text Communication). *Sociální studia/Social Studies*, 6(2).

Litvinova, O., Seregin, P., Litvinova, T., & Lyell, J. (2017). Deception detection in russian texts. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, s. 43-52.

Löhken, S. (2014). *Introverti a extroverti*. Praha, CZ: Grada.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy*, 33(3), 335-343.

Luft, J., & Ingham, H. (1955). The Johari window, a graphic model of interpersonal awareness. In *Proceedings of the Western Training Laboratory in Group Development*. Los Angeles, CA: UCLA.

MacPhillamy, D. J., & Lewinsohn, P. M. (1974). Depression as a function of levels of desired and obtained pleasure. *Journal of Abnormal Psychology*, 83(6), 651.

MAFRA (2011, January 26). *Diskuse Jak říkáte svým partnerům, kromě jména?* [Online diskuse]. Získáno z: <https://www.emimino.cz/diskuse/jak-rikate-svym-partnerum-krome-jmena-60751/>

Mairesse, F., & Walker, M. A. (2011). Controlling User Perceptions of Linguistic Style: Trainable Generation of Personality Traits. *Computational Linguistics*, 37(3), 455-488.

Mairesse, F., Walker, M. A., Mehl, M., & Moore, R. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457-500.

Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2), 74-79.

Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, 41(4), 701-707.

Martindale, C. (1987). Narrative pattern analysis: A quantitative method for inferring the symbolic meaning of narratives. *Literary discourse: Aspects of cognitive and social psychological approaches*. Berlin: de Gruyter, 167-181.

McCarthy, D. (1929). A Comparison of Children's Language in Different Situations and Its Relation to Personality Traits. *The Pedagogical Seminary and Journal of Genetic Psychology*, 36(4), 583-591.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812-821.

McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78, 348-369.

McLuhan, M. (1991). *Jak rozumět médiím: extenze člověka*. Praha: Odeon.

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, 531-552.

Mehl, M. R. (2006a). The lay assessment of subclinical depression in daily life. *Psychological assessment*, 18, 340-345.

Mehl, M. R. (2006b). Quantitative text analysis. In: Eid, M., Diener, E. *Handbook of multimethod measurement in psychology*. Washington: American Psychological Association, s. 141-156.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, 90(5), 862-877.

Melartin, T. K., Rytsälä, H. J., Leskelä, U. S., Lestelä-Mielonen, P. S., Sokeero, T. P., & Isometsä, E. T. (202). Current comorbidity of psychiatric disorders among DSM-IV major depressive disorder patients in psychiatric care in the Vantaa Depression Study. *Journal of Clinical Psychiatry*, 63(2), 126-134.

Mergenthaler, E. (1996). Emotion-abstractation patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of consulting and clinical psychology*, 64(6), 1306.

Miller, J. & Weinert, R. (1998). *Spontaneous spoken language*. Oxford: Clarendon Press Oxford.

Mistrík, J. (1989). *Štylistika*. Bratislava: Slovenské pedagogické nakladateľstvo.

Molendijk, M. L., Bamelis, L., Van Emmerik, A. A., Arntz, A., Haringsma, R., & Spinhoven, P. (2010). Word use of outpatients with a personality disorder and concurrent or previous major depressive disorder. *Behaviour Research and Therapy*, 48(1), 44-51. doi: 10.1016/j.brat.2009.09.007

Mukherjee, A., & Liu, B. (2010) Improving gender classification of blog authors. In *Proceedings of the 2010 conference on Empirical Methods in natural Language Processing*. Association for Computational Linguistics, s. 207–217.

Mulac, A., Studley, L. B., & Blau, S. (1990). The gender-linked effect in primary and secondary students' impromptu essays. *Sex Roles*, 23, 439–469.

Müllerová, O. (1978). O jazykovém přizpůsobování partnerů v dialogu. *Naše řeč*, 61(2), 57-68.

Müllner, J., Ruisel, I., & Farkaš, G. (1980). *Průručka pro administráciu, interpretáciu a vyhodnocovanie dotazníka na meranie úzkosti a úzkostlivosti*. Bratislava: Psychodiagnostické a didaktické testy, n. p., Bratislava.

Nagy, W. E. and Anderson, R. C. (1984). How many words are there in printed English? *Reading Research Quarterly*, 19, 304–30.

Nebeská, I. (1992). *Úvod do psycholingvistiky*. Praha: H&H.

Nechanská, B., Jann, J., Nováková, Z., Kudrna, K., Slábová, V., & Pašingerová, R. (2017). *Psychiatrická péče 2016*. Praha: ÚZIS ČR.

Nelson Francis, W., & Kučera, H. (1964). *A standard corpus of present-day edited american english, for use with Digital Computers*. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211-236.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.

O'Connor, B., Bamman, D., & Smith, N. A. (2011, December). Computational text analysis for social science: Model assumptions and complexity. In *Second Workshop on Computational Social Science and Wisdom of the Crowds*. Dostupné z <http://www.cs.cmu.edu/~nasmith/papers/oconnor+bamman+smith.nips-ws11.pdf>

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Panicheva, P., Ledovaya, Y., & Bogolyubova, O. (2016, November). Lexical, morphological and semantic correlates of the dark triad personality traits in russian facebook texts. In *Artificial intelligence and natural language conference (AINL)*, IEEE (s. 1-8).

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6), 934–952.

Parkvall, M. (2007). Världens 100 största språk 2007. In *Nationalencyklopedin*, 32.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (s. 224–239). New York, NY: Guilford Press.

Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, 211(2828), 42–45.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296–1312.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. L., & Booth, R. J. R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin & Auckland: The University of Texas at Austin, USA & The University of Auckland, New Zealand.

Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, 3, 50–53.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality & Social Psychology*, 85, 291–301.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*. Palo Alto: Annual Reviews, 547–577.



- Pervin, L. A. (1996). *The science of personality*. New York: Wiley.
- Petkevič, V. (2006). Reliable Morphological Disambiguation of Czech: Rule-Based Approach is Necessary. In Šimková, M. (ed.), *Insight into the Slovak and Czech Corpus Linguistics*, 26–44.
- Petkevič, V. (2014). Problémy automatické morfoložické disambiguace češtiny. *Naše řeč*, 97(4), 194–207.
- Petkevič, V., & Hnátková, M. (2017). Morphological disambiguation of multiword expressions and its impact on the disambiguation of their environment in a sentence. *Jazykovědný časopis*, 2017, 68, 145–155.
- Petr, J. et al. (1986). *Mluvnice češtiny 1*. Praha: Academia.
- Pinto-Gouveia, J., Cunha, M. I., & do Céu Salvador, M. (2003). Assessment of social phobia by self-report questionnaires: the social interaction and performance anxiety and avoidance scale and the social phobia safety behaviours scale. *Behavioural and Cognitive Psychotherapy*, 31(3), 291–311.
- Piotrowski, M. (2016). *Digital Humanities, Computational Linguistics, and Natural Language Processing*. Lecture Uppsala, March 4, 2016. Dostupné z [http://stp.lingfil.uu.se/~nivre/docs/michael\\_piotrowski\\_2016.pdf](http://stp.lingfil.uu.se/~nivre/docs/michael_piotrowski_2016.pdf)
- Popescu, I. I., Best, K. H. & Altmann, G. (2007). On the dynamics of word classes in texts. *Glottometrics*, 14, 58–71.
- Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology*, 81, 639–656.
- Psychology Foundation of Australia. (2014). *Depression Anxiety Stress Scale*, 18.6. 2018. Dostupné z <http://www2.psy.unsw.edu.au/dass>
- Pulverman, C. S., Lorenz, T. A., & Meston, C. M. (2015). Linguistic changes in expressive writing predict psychological outcomes in women with history of childhood sexual abuse and adult sexual dysfunction. *Psychological Trauma: Theory, Research, Practice, and Policy*, 7(1), 50–57.
- Pustina, J., & Daňková, B. (Eds.) (2016). *Multifaktorový osobnostní profil*. 3. vyd. Praha: T&CC.
- Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of research in personality*, 46(6), 710–718.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Dostupné z <https://www.R-project.org>

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM (s. 37–44).

Raskin, R., & Shaw, R. (1988). Narcissism and the use of personal pronouns. *Journal of Personality*, 56, 393–404.

Reichel, J. (2009). *Kapitoly metodologie sociálních výzkumů*. 1. vyd. Praha: Grada Publishing.

Revelle, W. (2018). *Psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA. Dostupné z <https://CRAN.R-project.org/package=psych> Version = 1.8.4

Riggio, R. E., & Carney, D. R. (2007). *Social Skills Inventory Manual: Research Edition, Czech version* (překlad L. Krejčová). Mind Garden.

Ritchie, H., & Roser, M. (2018). *Mental Health*. Dostupné z <https://ourworldindata.org/mental-health>.

Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: the lay assessment of subclinical depression from written self-descriptions. *Journal of Personality*, 78, 575–597.

Rosenbach, C., & Renneberg, B. (2015). Remembering rejection: Specificity and linguistic styles of autobiographical memories in borderline personality disorder and depression. *Journal of behavior therapy and experimental psychiatriy*, 46, 85–92.

Říčan, P. (2010). *Psychologie osobnosti: Obor v pohybu*. Praha: Grada.

Sanford, F. H. (1942). Speech and personality: A coparative case study. *Journal of Personality*, 10(3), 169–198.

Sboev, A., Litvinova, T., Gudovskikh, D., Rybka, R., & Moloshnikov, I. (2016). Machine learning models of text categorization by author gender using topic-independent features. *Procedia Computer Science*, 101, 135–142.

Shepherd, I. F. (1919). *Handbook of mental examination methods*. New York: The Macmillan Company.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of personality and social psychology*, 67(4), 674.

Scherer, K. R., & Giles, H. (1979). *Social markers in speech*. Cambridge: Cambridge University Press.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. In *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, s. 199–205.

Schmalt, H.–D., Sokolowski, K. & Landhens, T. (2010). *Multimotivační mřížka (MMG): Příručka*. Brno: Psychodiagnostika, s. r. o.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9): e73791.

Schwarz, N Knauper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, F. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570–582.

Sikos, J., David, P., Habash, N., & Faraj, R. (2014). Authorship analysis of Inspire Magazine through stylometric and psychological features. *2014 IEEE Joint Intelligence and Security Informatics Conference*.

Skálová, K. (2017, April 27th). *Jací lidé jsou, poznají z toho, jak mluví*. *Českoobudějovický deník*. Dostupné z [https://ceskoobudejovicky.denik.cz/zpravy\\_region/jaci-lide-jsou-poznaji-z-toho-jak-mluvi-20170427.html](https://ceskoobudejovicky.denik.cz/zpravy_region/jaci-lide-jsou-poznaji-z-toho-jak-mluvi-20170427.html)

Skoumalová, H. (2011). Porovnání úspěšnosti tagování korpusu. In Petkevič V., Rosen A., Korpusová lingvistika Praha 2011, 3 *Gramatika a značkování korpusů*. Praha, Nakladatelství Lidové noviny, s. 199–207.

Slančová, D., & Slančová, T. (2012). Komunikačný register ako alternatíva funkčného štýlu (na pozadí komunikácie v športe). *Stil*, 2012, 177–187.

Smith, C. P., Atkinson, J. W., McClelland, D. C., & Veroff, J. (Eds.). (1992). Motivation and personality: *Handbook of thematic content analysis*. Cambridge University Press.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of personality and social psychology*, 30(4), 526.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety scale*. Consulting Psychologists.

Spoustová, D., Hajič, J., Votrúbec, J., Krbeč, P. & Květoň, P. (2007). The best of two worlds: Cooperation of statistical and rule-based taggers for czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing* (s. 67–74). Praha: ACL 2007.

Stepikhov, A., & Loukina, A. (2014). Annotation and personality: Individual differences in sentence boundary detection. In: Ronzhin A., Potapova R., Delic V. (Eds.) *Speech and Computer. SPECOM 2014. Lecture Notes in Com-*

puter Science, vol 8773. Cham: Springer. DOI: [https://doi.org/10.1007/978-3-319-11581-8\\_13](https://doi.org/10.1007/978-3-319-11581-8_13)

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic medicine*, 63(4), 517-522.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. Oxford, England: M.I.T. Press.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64, 479-498.

Straková, J., Straka, M. & Hajič, J. (2013). A New state of the art czech named entity recognizer. In: *Lecture Notes in Computer Science, Vol. 8082, Text, Speech and Dialogue: 16th International Conference, TSD 2013. Proceedings* (pp 68–75). New York: Springer.

Straková, J., Straka, M. & Hajič, J. (2014). Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (s. 13–18). Baltimore, MD: ACL.

Strube, M. J. (1986). An analysis of the self-handicapping scale. *Basic and Applied Social Psychology*, 7(3), 211-224.

Sumner, C., Byers, A., Boochever, R., & Park, G. (2012, December). Predicting dark triad personality traits from Twitter and a linguistic analysis of tweets. In *International Conference on Machine Learning and Applications*, Boca Raton, FL.

Svebak, S. (1996). The development of the Sense of Humor Questionnaire: From SHQ to SHQ-6. *Humor: International Journal of Humor Research*, 9, 341-361.

Svoboda, M. (1999). *Psychologická diagnostika dospělých*. Praha: Portál.

Svoboda, M. (2001). Dotazníky. In M. Svoboda, D. Krejčířová, & M. Vágnerová (Eds.), *Psychodiagnostika dětí a dospívajících* (s. 320). Praha: Portál.

Swanson, A., & Ones, D. S. (2002, April). *The effects of individual differences and situational variables on faking*. Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, Toronto.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318), 626–633. doi:10.1080/01621459.1967.10482935.

Šrámková, L., & Cakirpaloglu, P. Tvorba škály fyzického vzhledu a „body image“ v kontextu českého prostředí. *Psychologie a její kontexty*, 6(2), 95–109.

Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychological inquiry*, 4(2), 122–130.

Uličný, O. (2014). K lingvistické metodologii: termín registr a jeho typy. In *Registre jazyka a jazykovedy*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 62.

ÚZIS ČR (2008). *MKN-10 Tabelární část. Druhé aktualizované vydání, Desátá revize*. Praha: Bomton Agency.

Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90, 306–322.

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40(5), 472–481.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of personality and social psychology*, 98(2), 281.

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, 20(2), 104–108.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, 95, 1202–1216.

Veselovská, K., & Bojar, O. (2013). *Czech SubLex 1.0, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL)*. Faculty of Mathematics and Physics, Charles University. Dostupné z <http://hdl.handle.net/11858/00-097C-0000-0022-FF60-B>.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197–210.

Voňková, H. (2012). Metoda ukotvujících vinět a možnosti využití v pedagogice. *Orbis Scholae*, 6(1), 27–40.

Výrost, J. (1992). *K možnostiam vyuzitia modelu struktury crt interpersonalneho spravania J. S. Wigginsa v diagnostike*. Karpatska: SAS. Nepublikováno.

Výrost, J. (n.d.). *Social Phobia Safety Behaviors Scale (SPSBS) (Pinto-Gouveia, Cunha, Salvador, 2003): Zadání k administraci*. Nepublikováno.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self– other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78(3), 546–558.

Weintraub, W. (1981). *Verbal behavior: Adaptation and psychopathology*. Springer Publishing Company.

Weintraub, W. (1989). *Verbal Behavior in Everyday Life*. New York: Springer.

Wells, B. (2009). Advances in I/O, speedup, and universality on colossus, an unconventional computer. In *Proceedings of the 8th International Conference on Unconventional Computation, LNCS, 5715, Berlin-Heidelberg*, s. 247–261.

WFMH (2012). *Depression: A global crisis*. Occoquan, VA: WFMH.

WHO (2018a). *Depression*. World Health organisation. Dostupné z: <http://www.who.int/news-room/fact-sheets/detail/depression>

WHO (2018b). *Poruchy osobnosti a chování u dospělých*. Praha: Ústav zdravotnických informací a statistiky ČR. Retrived from <https://www.uzis.cz/cz/mkn/F60-F69.html>

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37, 395–412.

Wiggins, J. S. (1995). *IAS, Interpersonal Adjective Scales: Professional Manual*. Psychological Assessment Resources.

Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social psychology*, 99(1), 174.

Workman, L. & Reader, W. (2014). *Evolutionary Psychology: An Introduction*. New York: Cambridge University Press.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44(3), 363–373.

Zasina, A. J., Lukeš, D., Komrsková, Z., Poukarová, P. & Řehořková, A. (2018). *Koditex (A corpus of diversified texts)*. Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague.

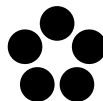
Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical psychology & psychotherapy*, 24(2), 384-391.

6

# Přílohy



# Informovaný souhlas účastníka výzkumu CPACT



**Já, níže podepsaný, svým podpisem stvrzuji svou ochotu a svobodnou vůli zúčastnit se výzkumu CPACT (Komputační psycholinguvistická analýza českého textu, podpořeného grantem GA ČR 16-19087S), realizovaného Pedagogickou fakultou Jihočeské univerzity v Českých Budějovicích v letech 2016–2018.**

Výzkum CPACT je realizován ve snaze splňovat kritéria vysoké odborné kvality a společensky prospěšných a významných cílů. Výzkum se řídí platnými etickými pravidly APA a ČMPS. Výzkum není spojen s rizikovou činností z hlediska ochrany života, zdraví a majetku. Účastníci se do výzkumu zapojují dobrovolně a za sjednanou finanční odměnu. Veškeré poskytnuté materiály jsou považovány za osobní a je s nimi takto nakládáno. Výstupy jsou publikovány pouze v anonymizované podobě, bez výslovného souhlasu účastníka nebudou původní materiály poskytnuty žádné třetí osobě. Účastníci výzkumu jsou plně informováni o požadavcích a specifikách výzkumu, souhlasí s využitím poskytnutých dat a svou účast stvrzují poučeným informovaným souhlasem.

**Prohlašuji, že jsem byl poučen o specifikách a požadavcích výzkumu CPACT, jsem s nimi seznáměn a nemám k nim výhrad. Podle zákona 101/2000 Sb. o ochraně osobních údajů ve znění pozdějších předpisů uděluji souhlas se zpracováním svých osobních údajů a s pořízením zvukových a obrazových materiálů v průběhu výzkumných sezení. V rámci výzkumu se zavazuji ke svědomitému a pravdivému vyplnění předkládaných materiálů, dodržení denního plánu výzkumu a k následné diskretnosti (tj. mlčenlivosti) o povaze a průběhu výzkumu CPACT směrem k osobám, které se výzkumu nezúčastnily). Stvrzuji rovněž to, že jsem fyzicky i psychicky připravený se výzkumu zúčastnit a nejsem si vědom toho, že by mou účastí mohly vzniknout jakékoliv škody na majetku či zdraví.**

v Českých Budějovicích dne  
(dnešní datum)

Jméno, příjmení  
a Váš podpis.

Projekt CPACT (GA ČR 16-19087S)  
Řešitel PhDr. Dalibor Kučera, Ph.D.  
Katedra pedagogiky a psychologie, Pedagogická fakulta  
Jihočeská univerzita v Českých Budějovicích  
Jeronýmova 10, 371 15 České Budějovice

E-mail: [info@cpact.cz](mailto:info@cpact.cz)  
Web: [www.cpact.cz](http://www.cpact.cz)  
Facebook: [facebook.com/vyzkumcpact](https://www.facebook.com/vyzkumcpact)  
Telefon: +420 605 983 326



## Příloha 3.1.8.A – Logo projektu



HLEDÁME MEZI LIDMI A SLOVY  
VÝZKUM GAČR 16-190875

## Příloha 3.1.8.B – Ukázka webové stránky projektu

The screenshot displays the CPACT project website. At the top, the logo is followed by navigation links: "INFORMACE O VÝZKUMU", "PODMÍNKY ÚČASTI", and "CHCI SE PŘIHLÁSIT". A language selector for "English" is also present. The main content area features a registration form titled "Přihlaste se do výzkumu CPACT!". The form text explains the study's focus on interpersonal communication and offers a 2500 Kč reward for participation. Below the form, there are sections for "INFORMACE O VÝZKUMU" and "PŘEDSTAVENÍ VÝZKUMU CPACT". The "PŘEDSTAVENÍ" section includes a video player showing the project logo and title. The footer of the page contains the text "272 | Výzkum CPACT: Komputační psycholinguvistická analýza českého textu".

# Získejte 2500 Kč za účast ve výzkumu!

## Stačí pět kroků

1



### Můžu se přihlásit?

- ✓ Jsem občanem ČR a mluvím česky.
- ✓ Můžu se zúčastnit jednodenního víkendového sezení v Českých Budějovicích.
- ✓ Vím o blízkém člověku, který se společně se mnou výzkumu zúčastní (může to být muž či žena).

### Výběr partnera

- Výzkumu se účastní dvojice osob (registrující a partner).
- Oba se dobře znají a mají spolu častý a upřímný kontakt.
- Vhodní jsou blízcí příbuzní, dlouhodobí partneři či manželé nebo dobří přátelé.



2

### Registrace přes web

- ✓ Jeden z dvojice (registrující) provede registraci obou účastníků.
- ✓ Na webu si zkontroluje, jestli je pro oba účastníky volná kapacita.
- ✓ Oba účastníci pak registraci potvrdí pomocí doručných e-mailů.



4



### Den výzkumu

- ✓ Uvítáme Vás a podrobně vysvětlíme, jak bude výzkum probíhat.
- ✓ V průběhu dne budete vyplňovat různé psychologické dotazníky, psát na počítači a zúčastníte se také dvou rozhovorů.
- ✓ Začneme v 9 hodin a skončíme přibližně v 16 hodin (včetně velké přestávky na oběd a menších přestávek na občerstvení).



3

### Termín Vaší návštěvy

- ✓ Do tří týdnů přijde registrujícímu e-mail s návrhem termínu návštěvy.
- ✓ Registrující navržené termíny probere s partnerem a vybraný termín za oba účastníky potvrdí.
- ✓ Ve stanovený den oba přijdou na jihočeskou Pedagogickou fakultu.

5



### Hotovo!

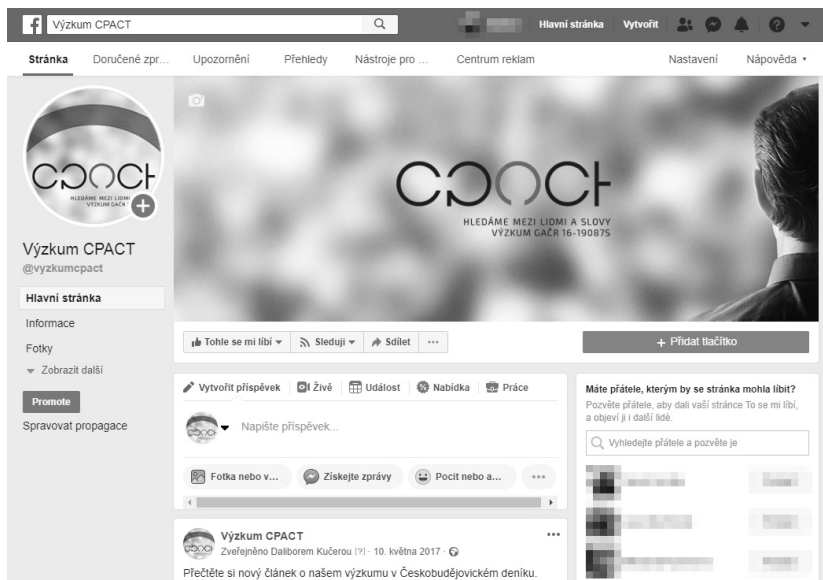
- ✓ Po Vaší návštěvě už na Vás nebudeme mít žádné další požadavky. Žádné další úkoly.
- ✓ Odměna 2500 Kč bude dvojicí vyplacena bankovním převodem na účet registrujícího.

**Výzkum CPACT** (Komputační psycholinguvistická analýza českého textu) je tříletý výzkumný projekt zaměřený na studium mezilidské komunikace. Ve výzkumu sledujeme vztahy mezi osobností člověka a slovy, která užívá. Projekt je realizován na Pedagogické fakultě Jihočeské univerzity v Českých Budějovicích za podpory Grantové agentury České republiky (16-19087S).

[www.cpact.cz](http://www.cpact.cz)



### Příloha 3.1.8.D – Ukázka prezentace na Facebooku



### Příloha 3.1.8.E – Ukázka videa na kanálu YouTube



**DENÍK V NOVÉM:** Nové rubriky, vše z vašeho okolí. Web Deníku dostal nový kabát

## Jací lidé jsou, poznají z toho, jak mluví

27.4.2017



České Budějovice – Nikoli vědci z Prahy, Olomouce nebo Brna, ale vědci z Českých Budějovic, v čele s Daliborem Kučerou z Katedry pedagogiky a psychologie Pedagogické fakulty Jihočeské univerzity (na snímku), přišli s myšlenkou nevšedního výzkumu mezilidské komunikace.



*Dalibor Kučera z Katedry pedagogiky a psychologie Pedagogické fakulty Jihočeské univerzity |*

*Foto: Deník/Klára Skálová*

### Příloha 3.3.2 – Seznam textových parametrů (proměnných)

V této příloze uvádíme seznam všech textových parametrů automaticky zjišťovaných v textech respondentů. U těch parametrů, které mohou nabývat různých hodnot, vypisujeme i všechny jejich potenciální hodnoty. Každý parametr (parametr–hodnota) je stručně vysvětlen. **Tabulka horizontálně využívá celé dvoustránky.**

Parametr	Hodnota	Statistické označení	Relativizovaná proměnná	Relativizováno k hodnotě	
1	Words	Words			
2	Sentences	Sentences			
3	Tokens	Tokens			
4	POS				
5	POS	N	POS.N	POSN_per_W	Words
6	POS	A	POS.A	POSA_per_W	Words
7	POS	P	POS.P	POSP_per_W	Words
8	POS	C	POS.C	POSC_per_W	Words
9	POS	V	POS.V	POSV_per_W	Words
10	POS	D	POS.D	POSD_per_W	Words
11	POS	R	POS.R	POSR_per_W	Words
12	POS	J	POS.J	POSJ_per_W	Words
13	POS	T	POS.T	POST_per_W	Words
14	POS	I	POS.I		
15	POS	Z	POS.Z	POSZ_per_Tok	Tokens
16	POS	X	POS.X		
17	SbP				
18	SbP	*	SbP_E		
19	SbP	,	SbP_D		
20	SbP	o	SbP.o		
21	SbP	1	SbP.1		
22	SbP	2	SbP.2		
23	SbP	3	SbP.3		
24	SbP	4	SbP.4		
25	SbP	5	SbP.5		
26	SbP	6	SbP.6	SbP6_per_W	Words
27	SbP	7	SbP.7	SbP7_per_W	Words
28	SbP	8	SbP.8	SbP8_per_W	Words
29	SbP	9	SbP.9		
30	SbP	:	SbP_C		
31	SbP	=	SbP_B		
32	SbP	?	SbP_A		
33	SbP	@	SbP.@		
34	SbP	A	SbP.A_A		
35	SbP	B	SbP.B_A	SbPB_per_W	Words
36	SbP	C	SbP.C_A		

Je-li to relevantní, uvádíme i příklady z analyzovaných textů (obvykle tři, je-li možné takové najít). Jako příklady vždy uvádíme konkrétní tvary slov (ne tedy základní tvary), malými písmeny (kromě vlastních jmen). Příklady vybíráme ze slov nejčastěji se vyskytujících ve zkoumaných textech, a to tak, aby byly pestré a dobře reprezentovaly daný parametr. Volíme tedy pokud možno více slovních druhů a více různých základních tvarů. Jsou-li tyto požadavky splněny, příklady uvádíme podle jejich frekvence.

<b>Význam atributu, popř. dvojice parametr – hodnota</b>	<b>Příklady</b>	
počet slov		1
počet vět		2
počet všech tvarů (včetně interpunkce aj.)		3
slovní druh		4
podstatné jméno	den, práce, pozdravem	5
přídavné jméno	ráda, dobrý, vážený	6
zájmeno	se, to, mi	7
číslovka	pár, dva, několik	8
sloveso	jsem, mám, vím	9
příslovce	tu, tak, už	10
předložka	na, v, s	11
spojka	a, že, i	12
částice	ahoj, opravdu, asi	13
citoslovce	pa, fajn, hele	14
interpunkce	, . !	15
neznámé či chybné slovo	cz, jobs, XY	16
podtyp slovního druhu		17
matematické operace (krát, děleno...)		18
podřadící spojka	že, když, aby	19
spojení předložky a zkráceného zájmena „-ň“ („naň“)		20
vztažné přivlastňovací zájmeno	jehož, jejíž, jejímž	21
část slova před pomlčkou		22
zkratka jako číslovka		23
vztažné nebo tázací zájmeno s adjektivním skloňováním	které, který, jaké	24
zájmeno „on“ po předložce	nimi, ní, nich	25
reflexivní zájmeno „se“ v dlouhých tvarech	sebou, sebe, sobě	26
reflexivní zájmeno „se“ (krátké tvary)	se, si, sis	27
přivlastňovací zájmeno „svůj“	svě, svůj, svou	28
vztažné zájmeno „jenž“ po předložce	němž, nichž, níž	29
interpunkce	, . !	30
číslicemi psaná číslovka	5, 3, 10	31
číslovka „kolik“ ap.	kolik	32
neznámé či chybné (neexistující) slovo	cz, jobs, XY	33
přídavné jméno obvyčné	dobrý, vážený, pracovní	34
tvar přítomného nebo budoucího času slovesa	jsem, mám, vím	35
jmenný tvar přídavného jména	ráda, spokojen, schopen	36

37	SbP	D	SbP.D_A	SbPD_per_W	Words
38	SbP	E	SbP.E_A		
39	SbP	F	SbP.F_A		
40	SbP	G	SbP.G_A		
41	SbP	H	SbP.H_A	SbPH_per_W	Words
42	SbP	I	SbP.I_A		
43	SbP	J	SbP.J_A		
44	SbP	K	SbP.K_A	SbPK_per_W	Words
45	SbP	L	SbP.L_A	SbPL_per_W	Words
46	SbP	M	SbP.M_A		
47	SbP	N	SbP.N_A		
48	SbP	O	SbP.O_A		
49	SbP	P	SbP.P_A	SbPP_per_W	Words
50	SbP	Q	SbP.Q_A	SbPQ_per_W	Words
51	SbP	R	SbP.R_A	SbPR_per_W	Words
52	SbP	S	SbP.S_A	SbPS_per_W	Words
53	SbP	T	SbP.T_A		
54	SbP	U	SbP.U_A		
55	SbP	V	SbP.V_A		
56	SbP	W	SbP.W_A	SbPW_per_W	Words
57	SbP	X	SbP.X_A		
58	SbP	Y	SbP.Y_A		
59	SbP	Z	SbP.Z_A	SbPZ_per_W	Words
60	SbP	^	SbP		
61	SbP	a	SbP.a	SbPa_A_per_W	Words
62	SbP	b	SbP.b		
63	SbP	c	SbP.c	SbPc_A_per_W	Words
64	SbP	d	SbP.d		
65	SbP	e	SbP.e		
66	SbP	f	SbP.f		
67	SbP	g	SbP.g	SbPg_A_per_W	Words
68	SbP	h	SbP.h		
69	SbP	i	SbP.i	SbPi_A_per_W	Words
70	SbP	j	SbP.j		
71	SbP	k	SbP.k		
72	SbP	l	SbP.l		
73	SbP	m	SbP.m		
74	SbP	n	SbP.n		
75	SbP	o	SbP.o		
76	SbP	p	SbP.p	SbPp_A_per_W	Words
77	SbP	q	SbP.q		
78	SbP	r	SbP.r		
79	SbP	s	SbP.s		
80	SbP	t	SbP.t		
81	SbP	u	SbP.u		
82	SbP	v	SbP.v		



zájmeno ukazovací	to, tuto, tento	37
vztažné zájmeno „což“	což, čímž, čehož	38
nesamostatná součást víceslovné předložky	vzhledem, nehledě	39
přídavné jméno slovesné odvozené od přítomného přechodníku	vedoucí, tykající	40
krátké tvary osobních zájmen („mě“, „mi“, „ti“, „mu“, ...)	mi, ti, ho	41
citoslovce	pa, faj, hele	42
vztažné zájmeno „jenž“ (tvary na j-)	jenž, jež, již	43
zájmeno tázací nebo vztažné „kdo“	kdo, koho, (o) kom	44
zájmeno neurčité „všechn“ , „sám“	vše, všechno, každý	45
přídavné jméno odvozené od slovesného tvaru minulého přechodníku		46
substantivum, obvyčejné	den, práce, pozdravem	47
samostatně stojící zájmena „svůj“, „nesvůj“, „tentam“	svoje	48
osobní zájmena (vč. tvaru „tys“)	já, vás, mne	49
zájmeno tázací/vztažné „co“, „copak“, „cožpak“	co, (o) čem, čeho	50
předložka, obvyčejná	na, v, s	51
zájmeno přivlastňovací „můj“, „tvůj“, „jeho“ (vč. plurálu)	mé, naše, vaší	52
částice (slovní druh "T")	ahoj, opravdu, asi	53
adjektivum přivlastňovací (na „-ův“ i „-in“)	Novákova, Františkových, Wortnerův	54
předložka vokalizovaná („ve“, „pode“, „ku“, ...)	ve, se, ze	55
zájmena záporná („nic“, „nikdo“, „nijaký“, „žádný“, ...)	nic, žádně, nikdo	56
neznámé nebo cizí slovo, překlep	at, Grand, la	57
zájmeno „co“ spojené s předložkou	nač	58
zájmeno neurčité	něco, nějaké, někdo	59
souřadící spojka	a, i, ale	60
číslovka neurčitá („mnoho“, „několik“ aj.)	několik, moc, mnoho	61
příslovce bez určení stupně a negace	tu, tak, už	62
kondicionální tvar	bych, by, bychom	63
číslovka druhová, adjektivní skloňování	obojí	64
slovesný tvar přechodníku přítomného	nemluvě	65
slovesný tvar: infinitiv	být, přijet, pracovat	66
příslovce s určením stupně a negace	dobře, určitě, hlavně	67
číslovky druhové „jedny“ a „nejedny“		68
slovesný tvar rozkazovacího způsobu	přijed, měj, dej	69
číslovka druhová >= 4, substantivní postavení („čtvero“)		70
číslovka druhová >= 4, adjektivní postavení, krátký tvar („čtvery“)		71
číslovky základní 1-4, „půl“, ...; sto a tisíc v nesubstantivním skloňování	dva, jeden, oba	72
slovesný tvar přechodníku minulého		73
číslovky základní >= 5	pár, pět, dvacet	74
číslovky násobné neurčité	několikrát, mnohokrát, víckrát	75
slovesné tvary minulého aktivního přičestí	byla, chtěl, zaujala	76
archaické slovesné tvary min. aktivního přičestí (zakončení „-ť“)		77
číslovky řadové	první, druhý, třetí	78
slovesné tvary pasivního přičestí	nucen, přesvědčena, řečeno	79
archaické slovesné tvary přít. a bud. času (zakončení „-ť“)		80
číslovka tázací násobná „kolikrát“		81
číslovky násobné	jednou, poprvé, dvakrát	82

83	SbP	w	SbP.w		
84	SbP	x	SbP.x		
85	SbP	y	SbP.y		
86	SbP	z	SbP.z		
87	Gen		Gen.M		
88	Gen	M	Gen.I		
89	Gen	I	Gen.F		
90	Gen	F	Gen.N		
91	Gen	N	Gen		
92	Gen	-			
93	Num				
94	Num	S	Num.S	Singulary_index	Num.S + Num.P + Num.D
95	Num	P	Num.P		
96	Num	-	Num		
97	Num	D	Num.D		
98	Cas		Cas.1		
99	Cas	1	Cas.2		
100	Cas	2	Cas.3		
101	Cas	3	Cas.4		
102	Cas	4	Cas.5		
103	Cas	5	Cas.6	Vocative_index	Cas.1 + Cas.2 + Cas.3 + Cas.4 + Cas.5 + Cas.6 + Cas.7
104	Cas	6	Cas.7		
105	Cas	7	Cas		
106	Cas	-			
107	PGe				
108	PGe	M	PGe.M		
109	PGe	I	PGe.I		
110	PGe	F	PGe.F		
111	PGe	N	PGe.N		
112	PGe	-	PGe		
113	PNu				
114	PNu	S	PNu.S	Singulary_P_index	PNu.S + PNu.P
115	PNu	P	PNu.P		
116	PNu	-	PNu		
117	Per				
118	Per	1	Per.1	Per1_index	Per.1 + Per.2 + Per.3
119	Per	2	Per.2	Per2_index	Per.1 + Per.2 + Per.3
120	Per	3	Per.3	Per3_index	Per.1 + Per.2 + Per.3
121	Per	-	Per		
122	Ten		Ten		
123	Ten	F	Ten.F	TenF_index	Ten.F + Ten.P + Ten.R
124	Ten	P	Ten.P	TenP_index	Ten.F + Ten.P + Ten.R

číslovky neurčité s adjektivním skloňováním	několikáté, několikátým, několikátá	83
zkratka, slovní druh neurčen/neznámý		84
zlomky zakončené na „-ina“		85
číslovka tázací řadová „kolikátý“		86
jmenný rod		87
rod mužský životný	pane, řediteli, vážený	88
rod mužský neživotný	den, pozdravem, dobrý	89
rod ženský	práce, ráda, byla	90
rod střední	to, vše, let	91
neurčuje se		92
číslo		93
číslo jednotné	jsem, to, mi	94
číslo množné	jsme, vás, let	95
neurčuje se		96
číslo dvojné (duál)	očima, oběma, čtyřma	97
pád		98
nominativ	to, já, den	99
genitiv	do, let, mého	100
dativ	si, mi, k	101
akuzativ	se, na, práci	102
vokativ	pane, řediteli, vážený	103
lokál	v, domě, tom	104
instrumentál	s, pozdravem, námi	105
neurčuje se		106
rod vlastníka		107
rod vlastníka mužský životný	Novákova, Františkových, Wortnerův	108
rod vlastníka mužský neživotný		109
rod vlastníka ženský	její, babiččiny, Terčíně	110
rod vlastníka střední		111
neurčuje se		112
číslo vlastníka		113
číslo vlastníka jednotné	mé, jeho, tvůj	114
číslo vlastníka množné	naše, vaší, jejich	115
neurčuje se		116
osoba		117
první osoba	jsem, mi, mám	118
druhá osoba	ti, jsi, Vás	119
třetí osoba	je, mrzí, jejich	120
neurčuje se		121
slovesný čas		122
čas budoucí	budu, pojedeme, půjdeš	123
čas přítomný	jsem, mám, vím	124

125	Ten	R	Ten.R	TenR_index	Ten.F + Ten.P + Ten.R
126	Ten	-			
127	Deg		Deg		
128	Deg	1	Deg.1	Deg1_index	Deg.1 + Deg.2 + Deg.3
129	Deg	2	Deg.2	Deg2_index	Deg.1 + Deg.2 + Deg.3
130	Deg	3	Deg.3	Deg3_index	Deg.1 + Deg.2 + Deg.3
131	Deg	-			
132	Neg		Neg		
133	Neg	A	Neg.A		
134	Neg	N	Neg.N	Negativity_index	Neg.N + Neg.A
135	Neg	-			
136	Voi		Voi		
137	Voi	A	Voi.A		
138	Voi	P	Voi.P	Pasivum_index	Voi.P + Voi.A
139	Voi	-			
140	Mod		Mod		
141	Mod	M	Mod.M_A		
142	Mod	m	Mod.m		
143	Mod	n	Mod.n		
144	Mod	-			
145	Var		Var		
146	Var	-			
147	Var	1	Var.1	Var1_per_W	Words
148	Var	2	Var.2		
149	Var	3	Var.3		
150	Var	4	Var.4		
151	Var	5	Var.5		
152	Var	6	Var.6	Var6_per_W	Words
153	Var	7	Var.7		
154	Var	8	Var.8	Var8_per_W	Words
155	Var	9	Var.9		
156	Asp		Asp		
157	Asp	B	Asp.B		
158	Asp	I	Asp.I	Imperfectum_index	Asp.I + Asp.P
159	Asp	P	Asp.P		
160	Asp	-			
161	Ran		Ran		
162	Ran	1	Ran.1		
163	Ran	2	Ran.2		
164	Ran	3	Ran.3		
165	Ran	4	Ran.4		
166	Ran	5	Ran.5		
167	Ran	6	Ran.6	Atypical_words_index	(Ran.5 + Ran.6) / Words
168	Ran	-			
169	Dem		Dem		
170	Dem	D	Dem.D	DemD_per_W	Words
171	Dem	-			

čas minulý	byla, chtěl, zaujala	125
neurčuje se		126
stupeň		127
pozitiv	dobrý, vážený, dobře	128
komparativ	dále, víc, lepší	129
superlativ	nejlepší, nejdříve, nejlépe	130
neurčuje se		131
		132
tvary není negován	jsem, den, mám	133
negace	není, nevím, nepříjemné	134
neurčuje se		135
slovesný rod		136
rod činný	jsem, mám, vím	137
rod trpný	nucen, přesvědčena, řečeno	138
neurčuje se		139
modalita		140
modální slovesa (úzké pojetí)	chtěl, mohl, musím	141
další slovesa s významem modality	žádám, nucen, donutila	142
substantiva s významem modality	žádostí, zákona, příkaz	143
ostatní případy		144
		145
neurčuje se (základní varianta)		146
rovnocenná (méně častá) varianta	vše, děkuji, mé	147
archaická nebo knižní varianta	léta, jej, odvězti	148
velmi archaický nebo hovorový tvar	vypomocť	149
archaický nebo knižní tvar	neváhej, sešli, skoč	150
hovorový tvar (akceptovaný ve standardní češtině)	dřív, nejdřív	151
obecněčeský tvar	náma, sme, budem	152
obecněřeský tvar	vem, zatím, lidma	153
zkratka	č., p., J.	154
speciální využití	dne	155
vid		156
slovesa obouvidá nebo homonymní (nerozlišeno)	stalo, jmenuji, dokážeš	157
nedokonavá slovesa	jsem, mám, vím	158
dokonavá slovesa	přijet, zaujala, napsat	159
neurčuje se		160
rank (řazení slova podle frekvence)		161
patří mezi 10 nejčastějších	a, se, na	162
patří mezi 100 nejčastějších	o, i, tak	163
patří mezi 1 000 nejčastějších	moc, práce, ráda	164
patří mezi 10 000 nejčastějších	pozdravem, ahoj, děkuji	165
patří mezi 100 000 nejčastějších	bytu, vyřízení, flexibilní	166
méně časté či nenalezeno v referenčním korpusu	koupačka, Paulát, čus	167
neurčuje se (interpunkce aj.)		168
zdrobněliny		169
zdrobnělina	sluníčko, koníčkem, vínko	170
není zdrobnělina		171

172	Vul		Vul		
173	Vul	V	Vul.V	VulV_per_W	Words
174	Vul	-			
175	Clq		Clq		
176	Clq	C	Clq.C	ClqC_per_W	Words
177	Clq	-			
178	Emo		Emo_C	Emo_C_per_W	Words
179	Emo	*	Emo_B	Emo_B_per_W	Words
180	Emo	+	Emo_A	Emo_A_per_W	Words
181	Emo	-	Emo	Emo_per_W	Words
182	Emo	_			
183	Em2		Em2_C	Em2_C_per_W	Words
184	Em2	*	Em2_B	Em2_B_per_W	Words
185	Em2	+	Em2_A	Em2_A_per_W	Words
186	Em2	-	Em2	Em2_per_W	Words
187	Em2	_			
188	Phr		Phr	Phr_per_W	Words
189	Phr	1	Phr.1	Phr1_perW	Words
190	Phr	-			
191	PhE		PhE_C	PhE_C_per_W	Words
192	PhE	*	PhE_B	PhE_B_per_W	Words
193	PhE	+	PhE_A	PhE_A_per_W	Words
194	PhE	-	PhE	PhE_per_W	Words
195	PhE	_			
196	1S		@1S	@1S_per_W	Words
197	1P		@1P	@1P_per_W	Words
198	2S		@2S	@2S_per_W	Words
199	2P		@2P	@2P_per_W	Words
200	3S		@3S	@3S_per_W	Words
201	3P		@3P	@3P_per_W	Words
202	PO1		PO1	PO1_per_W	Words
203	PO1S		PO1S	PO1S_per_W	Words
204	PO1P		PO1P	PO1P_per_W	Words
205	PP1		PP1	PP1_per_W	Words
206	PP1S		PP1S	PP1S_per_W	Words
207	PP1P		PP1P	PP1P_per_W	Words
208	PO2		PO2	PO2_per_W	Words
209	PO2S		PO2S	PO2S_per_W	Words
210	PO2P		PO2P	PO2P_per_W	Words
211	PP2		PP2	PP2_per_W	Words
212	PP2S		PP2S	PP2S_per_W	Words
213	PP2P		PP2P	PP2P_per_W	Words
214	PO3		PO3	PO3_per_W	Words
215	PP3		PP3	PP3_per_W	Words
216	V1S		V1S	V1S_per_W	Words
217	V1P		V1P	V1P_per_W	Words
218	V2		V2	V2_per_W	Words
219	Vneg		Vneg	Vneg_per_W	Words

vulgární slova		172
vulgární slovo	bordel, vůl, ožralý	173
není vulgární slovo		174
hovorová slova		175
hovorové slovo	tohle, akorát, pivko	176
není hovorové slovo		177
emocionálně nabitě slovo		178
emocionálně nabitě, neurčeno +/-	velmi, velice, vážený	179
pozitivní emocionální slovo	ráda, dobrý, děkuji	180
negativní emocionální slovo	mrzí, problém, stížnost	181
ostatní slova		182
emocionálně nabitě slovo (redukovaný seznam)		183
emocionálně nabitě, neurčeno +/-	velmi, velice, vážený	184
pozitivní emocionální slovo	ráda, dobrý, děkuji	185
negativní emocionální slovo	mrzí, problém, stížnost	186
ostatní slova		187
víceslovné obraty, frazémy		188
je součástí frazému	dobry den, děkovat předem, udělat chybu	189
není součástí frazému		190
emocionálně nabitý frazém		191
emocionálně nabitý frazém, neurčeno +/-	stát za to, zjednat nápravu, být v pořádku	192
pozitivně laděný frazém	ráj na zemi, nejlepší kamarádka, být fajn	193
negativně laděný frazém	rušit noční klid, dělat rámus, zlomit srdce	194
není součástí frazému		195
první osoba jednotného čísla	bych, mi, mám	196
první osoba množného čísla	nám, jsme, naše	197
druhá osoba jednotného čísla	ti, máš, jsi	198
druhá osoba množného čísla	vás, vaši, jste	199
třetí osoba jednotného čísla	je, mrzí, ho	200
třetí osoba množného čísla	jsou, jejich, jim	201
osobní zájmeno, první osoba	mi, mě, nám	202
osobní zájmeno, první osoba singuláru	mi, mě, já	203
osobní zájmeno, první osoba plurálu	nám, námi, nás	204
přívlast. zájmeno, první osoba	mé, naše, moje	205
přívlast. zájmeno, první osoba singuláru	mé, moje, mého	206
přívlast. zájmeno, první osoba plurálu	naše, naši, našeho	207
osobní zájmeno, druhá osoba	ti, tě, vás	208
osobní zájmeno, druhá osoba singuláru	ti, tě, tebe	209
osobní zájmeno, druhá osoba plurálu	vás, vám, vámi	210
přívlast. zájmeno, druhá osoba	vaší, vaši, tvůj	211
přívlast. zájmeno, druhá osoba singuláru	tvůj, tvé, tvoje	212
přívlast. zájmeno, druhá osoba plurálu	vaší, vaši, vaše	213
osobní zájmeno, třetí osoba	je, ho, ji	214
přívlast. zájmeno, třetí osoba	jejich, jeho, její	215
sloveso, první osoba singuláru	bych, mám, vím	216
sloveso, první osoba plurálu	jsme, máme, můžeme	217
sloveso, druhá osoba	máš, jsi, víš	218
sloveso, negace	není, nemá, nedá	219

220	Vneg1S	Vneg1S	Vneg1S_per_W	Words
221	Vneg1P	Vneg1P	Vneg1P_per_W	Words
222	Vneg2	Vneg2	Vneg2_per_W	Words
223	Aneg	Aneg	Aneg_per_W	Words
224	Dneg	Dneg	Dneg_per_W	Words
225	IE1SPV	IE1SPV	IE1SPV_per_W	Words
226	IE2PV	IE2PV	IE2PV_per_W	Words
227	W_per_S			Words
228	Lem_per_Sqrt_W			Words
229	FinV_per_S			Words
230	Intp_per_Sent			Words
231	Excl_per_Sent			Words
232	AN_per_AN+NA			Words
233	First_Conj			Words
234	Vulg_per_Sent			Words
235	Var67_per_Sent			Words
236	Coherence_index			Words
237	Pronom_index			Words
238	Formality_index			Words
239	Trager_index			Words
240	Readines_to_action_index			Words
241	Aggressiveness_index			Words
242	Activity_index			Words
243	Autosemantic_index			Words



sloveso, negace, první osoba singuláru	nevím, nechci, nemám	220
sloveso, negace, první osoba plurálu	nemáme, nejsme, nemůžeme	221
sloveso, negace, druhá osoba	nebudeš, nechceš, nemusíš	222
adjektivum, negace	nepříjemné, nerad, nedávné	223
adverbium, negace	neskutečně, nedaleko, nepřiměřeně	224
zájmeno a sloveso v 1. os. sg. v jedné větě	já jsem, já vím, já mám	225
zájmeno a sloveso ve 2. os. v jedné větě	ty víš, ty jsi, vy máte	226
počet slov na počet vět		227
počet různých základních tvarů slov na druhou odmocninu počtu slov		228
počet sloves v určitém tvaru na počet vět		229
počet interpunkčních znaků na počet vět (mimo uvozovky a závorky)		230
počet vykřičníků na počet vět		231
počet výskytů bezpříznakového pořadí adjektivum – substantivum na celkový počet adjektivních rozvití substantiv		232
počet vět začínajících spojkou na počet vět		233
počet vulgárních slov na počet vět		234
počet obecněčeských slov na počet vět		235
počet předložek, spojek a částic na 3*počet vět	Výpočet: (POST + POSJ + POSR) / (3*Sentences)	236
počet zájmen na počet sloves	Výpočet: POSP / POSN	237
míra důrazu textu na formu, opak míry důrazu na obsah	Výpočet: (POSN + POSA + POSR + POST)/2 - (POSP + POSV + POSD + POSI)/2 + 50	238
Tragerova míra aktivity	Výpočet: POSV / POSA	239
druhá míra aktivity textu	Výpočet: POSV / POSN	240
třetí míra aktivity textu	Výpočet: POSV / Words	241
čtvrtá míra aktivity textu	Výpočet: POSV / (POSV + POSA + POSD)	242
počet plnovýznamových slov na počet slov	Výpočet: (POSN + POSA + POSP + POSC + POSV + POSD) / Words	243

*Příloha 3.4.1.4 – Souhrn prokázaných vztahů mezi osobnostními charakteristikami a textovými parametry v jednotlivých textech (TXT1–TXT6) pro soubor P200 po BH (FDR) korekci (n=200)*

<b>Text</b>	<b>Blok</b>	<b>Test</b>	<b>Škála</b>	<b>Textový parametr</b>	<b><math>\rho</math></b>	<b>p_adj_BH</b>		
TXT1	PSY_S	MMG	OA	POSN_per_W	-0,231	0,035		
			OA	Formality_index	-0,230	0,035		
			SK	FinV_per_S	0,348	0,000		
			SK	GLS1	0,270	0,009		
			SK	Trager_index	0,262	0,011		
			SK	Readines_to_action_index	0,257	0,014		
			SK	Intp_per_Sent	0,233	0,034		
			SK	POSN_per_W	-0,234	0,034		
			SK	POSV_per_W	0,233	0,034		
			SK	Aggressiveness_index	0,233	0,034		
			SK	Formality_index	-0,226	0,039		
			SN	SbP8_per_W	-0,230	0,035		
			SO	Trager_index	0,310	0,003		
			SO	POSV_per_W	0,294	0,003		
			SO	Aggressiveness_index	0,294	0,003		
			SO	GLS1	0,299	0,003		
			SO	FinV_per_S	0,280	0,006		
			SO	Activity_index	0,280	0,006		
			SO	POSA_per_W	-0,266	0,010		
			SO	Readines_to_action_index	0,262	0,011		
			SO	@3S_per_W	0,249	0,020		
			SO	SbPa_A_per_W	-0,239	0,031		
			SO	Formality_index	-0,226	0,039		
				STAIX2	ANX	Intp_per_Sent	0,281	0,006
			PSY_O	SMS	CSV	Deg2_index	-0,286	0,014
					CSV	POST_per_W	0,275	0,014
			TXT2	PSY_S	BFI	S	Var67_per_Sent	-0,299
S	Var6_per_W	-0,293				0,007		
S	Negativity_index	-0,251				0,050		
S	ClqC_per_W	-0,252				0,050		
DASS1	S	POSA_per_W			-0,271	0,031		
	S	SbPa_A_per_W			-0,255	0,031		
	S	Trager_index			0,261	0,031		
PSY_O	IAS	BC			Singulary_index	0,283	0,047	
TXT3	PSY_S	DASS1			D	Intp_per_Sent	0,279	0,008
					D	FinV_per_S	0,258	0,010
			D	GLS1	0,258	0,010		
			D	Readines_to_action_index	0,253	0,010		
			D	POSN_per_W	-0,244	0,011		
			D	POSV_per_W	0,243	0,011		
			D	Formality_index	-0,246	0,011		

	D	Aggressiveness_index	0,243	0,011
	D	POSR_per_W	-0,215	0,032
	D	POSJ_per_W	0,216	0,032
	S	Intp_per_Sent	0,288	0,006
	S	POSV_per_W	0,261	0,010
	S	POSR_per_W	-0,264	0,010
	S	Aggressiveness_index	0,261	0,010
	S	FinV_per_S	0,255	0,010
	S	GLS1	0,241	0,012
	S	Readines_to_action_index	0,235	0,015
	S	Vneg_per_W	0,227	0,021
	S	POSJ_per_W	0,216	0,032
	S	Formality_index	-0,209	0,036
	S	POSZ_per_Tok	0,202	0,046
	S	Em2_B_per_W	0,201	0,046
DASS2	A	POSJ_per_W	0,298	0,006
	A	FinV_per_S	0,244	0,011
	A	Intp_per_Sent	0,250	0,011
	A	GLS1	0,243	0,011
	A	Formality_index	-0,219	0,031
	A	POSR_per_W	-0,213	0,034
	A	Trager_index	0,214	0,034
	A	POSV_per_W	0,209	0,036
	A	Readines_to_action_index	0,208	0,036
	A	Aggressiveness_index	0,209	0,036
	A	Activity_index	0,205	0,041
MMG	OA	SbPg_A_per_W	-0,294	0,018
STAIX2	ANX	Intp_per_Sent	0,361	0,000
	ANX	FinV_per_S	0,312	0,000
	ANX	W_per_S	0,272	0,004
	ANX	Em2_C_per_W	-0,233	0,026
	ANX	Emo_C_per_W	-0,229	0,026
	ANX	Aneg_per_W	-0,216	0,043
PSY_O SSI	EC	POSV_per_W	-0,250	0,041
	EC	POSJ_per_W	-0,248	0,041
	SC	POSJ_per_W	-0,249	0,041
	EC	Formality_index	0,248	0,041
	EC	Trager_index	-0,248	0,041
	EC	Aggressiveness_index	-0,250	0,041
	EC	GLS1	-0,254	0,041
TXT4 PSY_S BFI	N	PO1_per_W	0,258	0,033
	N	PO1S_per_W	0,250	0,040
	N	Trager_index	0,240	0,043
	N	SbPa_A_per_W	-0,234	0,046
	S	POSA_per_W	0,310	0,002
	S	SbPa_A_per_W	0,314	0,002
	S	Trager_index	-0,308	0,002
	S	PP1_per_W	0,248	0,040

		S	Readines_to_action_index	-0,241	0,043		
		S	GLS1	-0,239	0,043		
		S	POSN_per_W	0,234	0,046		
	DASS1	D	Intp_per_Sent	0,278	0,025		
	IAS	JK	PP1_per_W	0,301	0,015		
	MMG	SK	FinV_per_S	0,288	0,026		
	SMS	CSV	PP1_per_W	-0,251	0,047		
		CSV	V2_per_W	0,257	0,047		
		CSV	GLS7	0,248	0,047		
	STAIX2	ANX	Intp_per_Sent	0,309	0,001		
		ANX	Activity_index	0,283	0,003		
		ANX	FinV_per_S	0,240	0,021		
		ANX	POSV_per_W	0,233	0,021		
		ANX	Aggressiveness_index	0,233	0,021		
		ANX	PO1_per_W	0,227	0,021		
		ANX	PO1S_per_W	0,227	0,021		
		ANX	GLS8	-0,208	0,046		
		ANX	GLS7	0,205	0,047		
	PSY_O	BFI	O	SbPP_per_W	-0,245	0,050	
			O	Phr1_perW	-0,264	0,050	
			S	POSN_per_W	0,251	0,050	
			S	Pronom_index	-0,248	0,050	
			S	Readines_to_action_index	-0,246	0,050	
			S	PO1S_per_W	-0,244	0,050	
	TXT5	PSY_S	BFI	E	POSP_per_W	0,248	0,047
				E	Pronom_index	0,243	0,047
				N	Lem_per_Sqrt_W	-0,241	0,047
				O	Lem_per_Sqrt_W	0,342	0,000
				O	POSZ_per_Tok	-0,238	0,047
				O	Negativity_index	-0,234	0,047
				O	Em2_per_W	-0,234	0,047
				O	PhE_per_W	-0,241	0,047
				O	Vneg_per_W	-0,238	0,047
				O	Vneg1S_per_W	-0,248	0,047
		SSI	SC	Lem_per_Sqrt_W	0,344	0,000	
	PSY_O	SHQ	S	PhE_C_per_W	-0,291	0,009	
	TXT6	PSY_S	BFI	E	SbPp_A_per_W	0,275	0,006
				E	PO1_per_W	0,235	0,045
				O	POSZ_per_Tok	-0,300	0,002
				O	Emo_per_W	-0,302	0,002
				O	Em2_per_W	-0,314	0,002
				O	PhE_per_W	-0,302	0,002
				O	Lem_per_Sqrt_W	0,245	0,033
				S	@2P_per_W	0,283	0,005
				S	Per2_index	0,277	0,006
				S	V2_per_W	0,237	0,045
		IAS	FG	SbPp_A_per_W	-0,286	0,025	
			LM	Emo_per_W	-0,277	0,025	

	LM	Em2_per_W	-0,270	0,025
	LM	PhE_per_W	-0,269	0,025
	LM	POSZ_per_Tok	-0,263	0,030
SSI	ES	Emo_per_W	-0,301	0,005
	ES	Em2_per_W	-0,302	0,005
	SC	Lem_per_Sqrt_W	0,269	0,021
	SC	Emo_per_W	-0,265	0,021
	SC	Em2_per_W	-0,264	0,021
STAIX2	ANX	POST_per_W	0,268	0,014
PSY_O BFI	O	Em2_per_W	-0,278	0,036
	O	Emo_per_W	-0,261	0,044
	E	PO1_per_W	0,257	0,044

*Legenda: Blok PSY\_S – varianta dotazníkového sebesouzení (S); blok PSY\_O – varianta dotazníkového posouzení druhou osobou (O); škála – dotazníková škála (viz kap. 3.3.1).*

*S vděčností věnováno všem účastníkům výzkumu CPACT,  
našim kolegům a asistentům, našim blízkým a všem těm,  
kteří náš projekt podporovali a rozvíjeli.*



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice



GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY



HEEDÁME MEZI LIDMI A SLOVY  
VÝZKUM GAČR 16-190875