

TEMNÁ STRANA KRÁLE KAKTUSŮ: CO MAXQDA NEUMÍ A CO S TÍM DĚLAT?

Ondřej Hejnal

*Katedra antropologie, Fakulta filozofická, Západočeská univerzita v Plzni
ohejnal@gmail.com*

**The dark side of the king of the CAQDAS:
What is MAXQDA not able to do and how should
it be dealt with?**

Abstract—Without a doubt, MAXQDA, the leading CAQDAS (software), is one of the most useful and commonly used tools for qualitative data analysis. Despite this and, in a sense, outside the common praise “the King of the CAQDAS”, the aim of this methodological paper is to present a variety of troubles (and their solutions) associated with the processing of a dataset with more than 90,000 media messages in MAXDictio (MAXQDA addon). Going in steps, I will demonstrate the phases of data import, categorization of data, autocoding of sentences, word frequencies, selection of key words, dictionary-making and an analysis of co-occurrences. In addition to MAXQDA, the article will mention some minor applications/software, potentially very handy for the diverse forms of computer-assisted text analysis (CATA). Accordingly, what are the limits of MAXQDA? And if exceeded, what are the solutions? And what other software can be (or have to be) used with MAXQDA.

Keywords—CAQDAS, MAXQDA; MAXDictio; MorphoDiTa; methodology; semantic network; data processing; content analysis; co-occurrence

ÚVOD

PROGRAM MAXQDA (2016), zpravidla řazený mezi CAQDAS, již více než dvě dekády zaujímá postavení jednoho z nejlepších a nejpoužívanějších „kaktusů“, tj. pomocných nástrojů při kvalitativní analýze dat jak v podobě textu, tak i v poslední době zvuku, obrazu či videa (Koenig 2004; srov. Kuckartz a Sharp 2011; Saillard 2011; Mertl a Hejnal

2013; Hejnal a Lupták 2015; Vašát a Čermák 2015; Silver a Lewins 2014). Ve shodě se současnými trendy spojování kvalitativního a kvantitativního výzkumného designu (Lu a Shulman 2008, 107) je MAXQDA vyvíjeno podobně jako jeho přímí konkurenti se silným zřetelem na upotřebitelnost v rámci projektů různým způsobem směřující či mixující postupy obou metodologických tradic (Fielding a Lee 2002; Fielding a Cisneros-Puebla 2009; Kuckartz a Sharp 2011, 12). Třebaže se v mnoha případech jedná o míšení založené na přehnaném a příliš zjednodušujícím vymezení obou metodologií, resp. na ostrém protikladu „slova“ versus „číslo“ (Schönfelder 2011, 10), zůstává pravdou, že kvantitativní prvky a principy pronikají do MAXQDA coby nové funkce či obměny funkcí starších. Doplněk MAXDictio uplatnitelný při klasické (kvantitativní) obsahové analýze je bezesporu nejvýraznějším dokladem snahy o integraci kvantitativních prvků do MAXQDA.

Během několika měsíců roku 2012 se mi podařilo díky krátkodobému neplacenému zpřístupnění databáze Anopress vytvořit korpus všech (více než 90000) mediálních sdělení, v nichž se objevilo slovo bezdomovec, resp. základ „bezdom“ ve všech tvarech a obměnách (tj. bezdomovkyně, bezdomovectví, bezdomovství atd.) v rozmezí let 1996 až 2012. S ohledem na „zastarávání“ dat jsem se rozhodl korpus alespoň předběžně, spíše kvantitativně zpracovat. Jak již bylo řečeno, ačkoli je MAXQDA určeno převážně ke kvalitativním technikám analýzy dat, v současnosti je k máni spolu s kvantitativním modulem MAXDictio (coby součástí „plusové“, dražší

edice). V návaznosti na povahu dat coby časové řady a poměrně malého množství vodiček, co by mohlo tvořit obsah/y jednotlivých součástí datového souboru, jsem zvolil postup, který je inspirován soudobými směry analýz temporálních/dynamických sémantických sítí (Batagelj, Mrvar a Zaversnik 2002; Diesner a Carley 2004a; Diesner a Carley 2004b; Bender-deMoll a McFarland 2006; Carley et al. 2007; Leydesdorff et al. 2008; Leydesdorff a Schank 2008; Tambayong a Carley 2012; Newman, Barabasi a Watts 2006) a zároveň si vystačí s metodologickými prvky „klasické“ kvantitativní obsahové analýzy (srov. Weber 1990; Popping 2000; Riffe, Lacy a Fico 2005; Krippendorff 2012); konkrétně s těmi, u nichž MAXQDA manuál uvádí, že by si měl MAXDictio poradit.

Kromě několikaleté zkušenosti s ovládním MAXQDA bylo setrvání u tohoto softwaru motivováno předpokladem, že po kvantitativním zpracování korpusu zbydou projektové soubory, které již budou připravené pro kvalitativní analýzy, příp. budou dovolovat navrácení se k předchozí fázi a zhodnocení různých interpretací vzešlých z odlišně metodologicky a teoreticky uchopených, ale identických dat. Analytické kroky, které se více či méně překrývají s metodologickými popisy Martina Hájka (2010; 2014) a Václava Čepeláka (2013),¹ zahrnovaly import, kategorizaci dat (tj. přiřazení proměnných), automatické kódování vět s „bezdóm“ (coby jednotka KWIC), přehledy frekvencí slov (uvnitř KWIC), výběr častých a důležitých slov, vytvoření slovníku, aplikace slovníku, tj. vyhledání klíčových slov ve větě KWIC, autokódování slov ze slovníku, analýza spoluvýskytu slovníkových slov prostřednictvím „Prohlížeče vztahů kódů“ (Code Relation Browser), vizualizace spoluvýskytu kódů v MAXMaps (Import Coocuring Codes) a po vyčerpání MAXQDA bylo v plánu, po-

kud by se daný směr ukazoval jako potenciálně zajímavý a přínosný, pokračování v softwaru na analýzy sítí UCINET (2002) (popř. Gephi 2009; Bastian et al. 2009; Bastian, Heymann a Jacomy 2009).

Následujících několik kapitol vybočuje ze všech metodologických statí, které jsme spolu s kolegy inklinujícími k používání MAXQDA doposud sepsali. Na rozdíl od běžné deklarace pozitiv se popis praktických úkonů zaměřil na nejrůznější peripetie, které na uživatele nejen MAXQDA (srov. Nohlová a Nikolin 2015) mohou čekat, pokud se velikost korpusu dat vymyká běžně používaným datovým souborům. Kromě MAXQDA budou zmíněny i některé drobnější přehlížené aplikace, jejichž potenciál může být pro nejrůznější formy CATA, např. CAQDAS poměrně výrazný. Obecně je text spjat s několika otázkami: Jaké jsou tedy limity MAXQDA? Jaká mohou být zvolena řešení v případě jejich překročení? Které další softwary lze (anebo je nutné) využít spolu s MAXQDA?

Rok	Velikost projektu [MB]	Počet mediálních sdělení		
		ČR [tis.]	Četnost „bezdóm“ ^{***}	
			absolutní	relativní
1996	18,55	799	1609	0,20 %
1997	20,34	1239	2501	0,20 %
1998	19,12	1093	2362	0,22 %
1999	27,81	1410	3414	0,24 %
2000	25,94	1499	3623	0,24 %
2001	26,34	1469	3652	0,25 %
2002	31,53	1448	4105	0,28 %
2003	32,28	1501	4457	0,30 %
2004	32,56	1584	5013	0,32 %
2005	48,48	1659	5609	0,34 %
2006	44,16	1688	6265	0,37 %
2007	46,86	1905	7143	0,37 %
2008	57,50	1928	8143	0,42 %
2009	49,67	1794	6669	0,37 %
2010	53,76	1745	7291	0,42 %
2011	47,76	1740	6187	0,36 %
2012	79,59	3146	13599	0,43 %
Celkem	662,26	27646	91642	0,33 %

Pozn.: Vlastní zpracování dat z databáze Anopress.

Tabulka 1: Mediální sdělení v ČR, četnost sdělení s „bezdóm“^{***} a velikost projektů

IMPORT

Datový korpus zahrnoval před úpravami a čištěním 91642 sdělení v podobě RTF souborů (1 RTF soubor činí přibližně 40 kB). Kritériem jejich výběru byl ne/výskyt v libovolném čes-

¹ Spolu se Zdeňkem Hájkem vyvinul Martin Hájek vlastní software COOA (Co-occurrence Analysis Software), který je volně ke stažení na http://publication.fsv.cuni.cz/attachments/471_setup_COOA.exe (viděno 16. ledna 2012). Mimo frekvencí slov, KWIC atd. vypočítává COOA vybrané podobnostní míry vhodné pro aplikaci Hájkem i Čepelákem preferovaného MDS (multidimenzionálního škálování/scaling). Podobnostní analýza (Similarity Analysis) dokumentů na základě spoluvýskytu kódů se coby zabudovaná možnost objevila až v poslední verzi MAXQDA12.

kém, zmíněnou databází monitorovaném (tj. spadající do rozmezí let 1996–2012), médiu (alespoň jednoho) řetězce „bezdom“ bez ohledu na povahu jeho (textového) okolí. Jinými slovy, mediální sdělení bylo považováno za relevantní, splňovalo-li podmínku přítomnosti zmíněného řetězce; obsahově ani tematicky vymezeno nebylo. Vzhledem k tomu, že jedinou relevantní informací před zpracováním korpusu byly počty sdělení ve 204 po sobě jdoucích měsících, bylo nutné využít nějaký software, s jehož pomocí lze automatizovaně kategorizovat textové soubory a zároveň usnadňuje následnou kvalitativní, a nejlépe i kvantitativní analýzu. Nikoli překvapivě byl vybrán software MAXQDA, s nímž mám řadu (veskrze pozitivních) zkušeností a s jehož ovládním jsou na různé úrovni seznámeni i moji kolegové, potenciální spolupracovníci při analytickém zhodnocení jeho obsahu, kteří ovládají jiné součásti MAXQDA (srov. např. Krčál 2012; 2013; Toušek 2013; Hejnal a Lupták 2015; Vašát a Čermák 2015).

Jakkoli MAXQDA splňuje naznačené podmínky, datový soubor se přeci jenom na poměry softwaru určeného na kvalitativní analýzu (zpravidla nepřesahující stovky textů) zdá příliš velký. Nicméně například můj projekt s kódovanými materiály (nejvíce knihy a články v PDF) přesahující 4 Gb, resp. 1500 dokumentů a 35000 kódovaných segmentů, funguje bez větších problémů. Soudě dle manuálu k poslední (12) verzi² se velikost mediálního korpusu sice pohybuje pod hranicí teoretického maxima jednoho MAXQDA projektu, avšak s ohledem na zbývající uváděná omezení a možnost sloučit dílčí součásti mediálního korpusu později³ byla data kompletně nahrána do 17 dílčích projektových souborů (odpovídajících rokům), resp. 12 příslušných dokumentových skupin (měsíců). Spolu s tímto rozdělením se pochopitelně úměrně navýšil i počet

následných kroků při plánované („jednorázové“) aplikaci funkcí MAXDictio.

Import RTF souborů do MAXQDA projektu s dokumentovými skupinami (např. 2004-01 až 2004-12, tj. 12) se s každým dalším zpomaloval, ale nijak výrazně. Určitou „novinkou“ byl stav, kdy MAXQDA zdánlivě během importování „zamrzl“, a nebylo tedy bez vizuální informace možné odhadnout počet doposud nahraných článků/relací (zřejmě lze říci, že 500 RTF souborů trvalo dvě až tři hodiny). Situace se začala radikálněji zhoršovat od roku/projektového souboru 2005. Import byl v případě měsíců květen až prosinec stále pomalejší. Manuál řešení nenabízel, doporučení online podpory – „Co takhle importovat větší množství dokumentů přes noc?“ (e-mailová korespondence, 30. 6. 2011) – bylo v lepším případě dočasné. Po zařazení druhého počítače do procesu nahrávání dat bylo alespoň možné pracovat na již hotových.

KATEGORIZACE

Tato fáze navazuje na již několikrát popisovanou kategorizaci částečně strukturovaných dat v prostředí MAXQDA (Hejnal 2012; Hejnal a Lupták 2013a; 2013b; 2015; Mertl a Hejnal 2013): unikátní (tj. s vysokou pravděpodobností jinde v textu se neobjevující) části horní identifikační poloviny Anopress výstupu (tzv. hlavička, viz Obrázek 1) lze za běžných okolností, resp. s menšími projekty, během pár sekund vyhledat (např. „Zdroj: [mezeraj“ anebo „Mutace:“), automaticky kódovat hledaný segment s možností „celá věta“ (vzniká např. „Zdroj: Mladá Fronta Dnes“) a po odmazání hledaných řetězců (tj. „Zdroj:[mezeraj“) přiřadit zbytek („Mladá Fronta Dnes“) coby atribut / proměnnou příslušnému souboru / mediálnímu sdělení. Proběhne-li vše v pořádku, datový soubor lze filtrovat (ať už přímo v MAXQDA, či po exportu dat v jiném softwaru) na základě víceméně všech základních charakteristik zpravidla se omezujících na časové zařazení, geografickou lokalizaci, typ média (např. celostátní deníky, regionální rozhlas, televize atd.) a konkrétní deníky, televize či rozhlasové stanice. Jakkoli je tento postup s mezikrokem v podobě úpravy excelové nebo jiné tabulky stále platný, přímá konverze

² MAXQDA projekt může být naplněn 99999 dokumenty. Bez limitů jsou uváděny celkový počet kódů, kódovaných segmentů i proměnných, ale je doporučena kontrola výkonu před zpracováním 1000 dokumentů, 1000 kódů a 10000 segmentů. Není doporučeno používat více než 250 proměnných (Kuckartz 2016).

³ Původní záměr, vytvoření projektu jediného, byl tudíž nejprve odložen a po několika neúspěšných experimentech s více roky zároveň v jediném projektu i zcela zavržen.

v rámci MAXQDA (Hejnal a Lupták 2013b, 237–238) se ukazuje méně chybovou.

MONITORING MÉDIÍ 12.12.2011 18:03

AI
VÁŠ DO

Desetníky a dvacetníky skončily

Datum: 01.02.2004

Relace: Pražský deník

Zdroj: TV PRAHA

Moderátor: Martin Pouva

Oblast: Televize

Zkratka oblasti: TV

Zkratka zdroje: TVT7

Identifikace: TVT720040201010004

Název: Desetníky a dvacetníky skončily

Zpracováno: 02.02.2004 07:17:42

Klíčová slova: bezdomovců

Moderátor (Martin Pouva):

Naposledy v pátek stál před azylovým domem A

Obrázek 1: Příklad částečně strukturovaného textu (mediálního sdělení)

Kategorizace dat měla podobný průběh jako import. Zpočátku probíhalo hledání a autokódování velmi rychle, u „objemnějších roků“ docházelo ke zpomalování. Třebaže velikost dílčích projektů není nijak závratná, průměrně 39 MB, kvantita pomocných kódů (tj. ZDROJ, DATUM, OBLAST atd.) některé MAXQDA projekty doslova znehybnily a jiné připravily o základní matematické dovednosti (viz Obrázek 2). Snaha průběžně tyto kódy likvidovat⁴

⁴ Po importu proměnných, tj. přiřazení například kategorie „Blesk“ k příslušným dokumentům, se staly pomocné kódy, resp. kódované segmenty (např. „Zdroj: Blesk“) nadbytečnými, protože veškeré filtrovací funkce zastanou (přínejmenším v případě větších korpusů) mnohem rychleji právě vytvořené dokumentové proměnné. Nicméně po určitých úpravách lze použít funkce *Transform code into a document variable*, popř. *Transform code into a categorical document variable*, které vytvoří automaticky proměnnou indikující počet výskytů daného kódu (první možnost), resp. subkódů daného kódu (druhá možnost). Jinými slovy, ve stávající podobě kódy nelze použít. Pokud bychom převedli např. kód DATUM, v tabulce s proměnnými a jejich hodnotami získáme stejnojmennou proměnnou a sloupec s jedničkami (na druhou stranu toho lze využít při kontrole). Druhá možná transformace nebude nabídnuta, jelikož její podmínkou je existence subkódů, tj. kategorií proměnné. Jestliže kupříkladu vyhledáme „Oblast: Regionální deníky“, poté „Oblast: Celostátní deníky“ a následně automaticky okódujeme příslušné části textu dvěma novými kódy (např. „Regionální deníky“ a „Celostátní deníky“), které podřadíme kódu OBLAST, lze prostřednictvím transformace OBLASTi

nevedla k lepšímu výsledku, neboť rychlost mazání se v zásadě rovnala autokódování. Díky průběžnému zálohování různých verzí to nepředstavovalo příliš velký problém: proměnné ze „znehyněného“ projektu byla exportována a importována do verze bez kódů. Nicméně s ohledem na plánovanou analýzu, která existenci přínejmenším jednoho kódu (věta s „bezdom“) předpokládá, nemluvě o dalším kvalitativním kódování, se jeví moje představa vynechávající čištění dat, které by bylo v ideálním případě zbytečné, jako mylná.

Code System	Hodnota
12-2004	1072
1996-12	540
1996-11	316
1996-10	306
1996-09	270
1996-08	336
1996-07	270
1996-06	198
Code System	40024
0_IDENTIFRACE	0
Text	22600
METAFORY	0
Sets	0

Pozn.: Hodnota Code System by se měla rovnat součtu Navez a Text.

Obrázek 2: Výřez kódovníku z projektu kombinující roky 1996, 2004 a 2012

„PŘEDZPRACOVANÝ“ TEXT

Znatelné zpomalení softwaru bylo možné částečně zmírnit hromadnou úpravou, resp. pročištěním souborů před importem do MAXQDA: (1) převést RTF soubory na univerzálnější TXT s kódováním UTF-8, (2) pojmenovat je dle logického klíče (např. `rok_mesic_den_poradi`) a (3) nahradit (tj. většinou vymazat) nadbytečný text ve více souborech najednou. Ke každému kroku lze nalézt několik (placených i nezaplatněných) programů, které nebývají součástí „standardní“ výbavy počítače. Hromadné přejme-

do kategorické proměnné získat proměnnou členící korpus na dva typy deníků (toho lze dále využít ke komparativní analýze v řadě funkcí MAXQDA). Oba typy transformace, resp. *hodnoty* proměnné, jsou automaticky průběžně aktualizovány (zdůrazňuji hodnoty, tj. je-li podřazen transformovanému kategorickému kódu další sub/kód, v tomto kontextu kategorie proměnné, v tabulce proměnných nenastane žádná změna). Příslušné (transformované) kódy by pochopitelně neměly být smazány, byť v případě kategorií se vyjma znemožněné automatické aktualizace v zásadě nic nezmění a kategorie zůstávají v tabulce dokumentových proměnných.

nování nabízí zkratka CTRL+M v relativně známém Total Commander, konverzi např. RTFtoUTXT⁵ a nahrazování / promazávání textu (zejména hyperlinků) univerzální Notepad++⁶ či SarAnt (viz dále). Avšak tato fáze nabízí mnohem více: vytvoření kategorií (resp. kategorizujících kódů) dokumentů (mediálních sdělení) již během importu.

Najít (co)	Nahradit (čím)
Datum:	#TEXT
Autor:	#ENDCODE\r#CODEAutor\r
Str.:	#ENDCODE\r#CODEStrana\r
Zdroj:	#ENDCODE\r#CODEZdroj\\
Oblast:	#ENDCODE\r#CODEOblast\\
Zkratka oblasti:	#ENDCODE\r#CODEZkratka\r
Název:	#ENDCODE\r#CODENazev\r
Klíčová slova:	#ENDCODE\r#CODEClanek\r
IČ: 26694484	#ENDCODE \r

Pozn.: Testováno v Notepad++ v6.8.3

Tabulka 2: Přehled hledaných řetězců a jejich náhrad

Využijí se k tomu „předzpracované (strukturované) dokumenty“, resp. schopnost MAXQDA vytvářet dokumenty a kódy na základě rozpoznání specifických řetězců znaků – #TEXT[název], #CODE[název] (vytvoření nového kódu, příp. #CODE[Název] \Subcode[Název2] pro vytvoření hierarchie) a #ENDCODE (ukončení kódovaného segmentu) – zanesených v korpusu dat (viz Tabulku 2, první sloupec). Před nahrazováním je možné sloučit textové soubory do jediného (např. ve freeware programu TXTcollector),⁷ jelikož každý výskyt „#TEXT“ vytváří automaticky nový dokument. Namísto hromadného mazání je možné hromadně nahrazovat pomocné části dokumentu („Datum:“, „Zdroj:“, „Oblast:“ atd.) některými ekvivalentními a rozpoznatelnými řetězci (tj. „#TEXT[datum]“, „#CODEZdroj“, „#CODEOblast“ atd.). Prostřednictvím běžné funkce „najdi-nahrad“ (ať už postupně, anebo jednorázově),⁸ například ve zmíněném Notepad++ (viz Tabulku 2), se

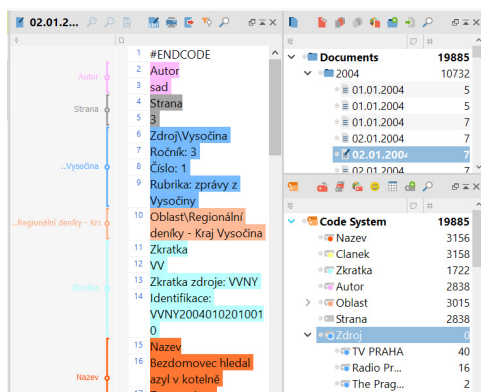
⁵ Ke stažení na <http://www.emreakkas.com/wp-content/uploads/2010/12/RTFtoUTXT.exe> (viděno 26. ledna 2016).

⁶ Konkrétní verze ke stažení na <https://notepad-plus-plus.org/> (viděno 26. ledna 2016).

⁷ Ke stažení na <http://bluefive.pair.com/TXTcollector.zip> (viděno 26. ledna 2016).

⁸ Funkční hromadné nastavení testované v Notepad++ je: (Datum:)|(Autor:)|(Str.:)|(Zdroj:)|(Oblast:)|(Zkratka oblasti:)|(Název:)|(Klíčová slova:)|(Zpracovatel: Anopress) v ko-

zkracuje čas obvykle věnovaný „doladování“ datového souboru. Projekt se stává plně funkční, resp. k analýze připravený, současně s importováním (v tomto případě spíše vytvořením) posledního dokumentu. Jakkoli chod MAXQDA setrvává v pomalejším módu, nejedná se o stav, který práci zcela znemožňoval. Krom postupu s textovými soubory nabízí MAXQDA i alternativní řešení s tabulkovým formátem (XLS, XLSX, CSV), které je ještě o něco rychlejší a umožňuje automaticky vytvořit dokumentové proměnné (viz dále). V případě některých typů vstupních dat (např. dotazníky s otevřenými odpověďmi, strukturovaných interview atd.) se jedná o bezesporu preferovaný způsob importu. Zároveň je třeba dodat, že v některých typech střídmejších analýz vyžadujících kupříkladu pouze seznam klíčových slov v kontextu (věty, odstavce atd.) s několika identifikátory (rok, měsíc, název média atd.) není nutné (díky množství pluginů) prostředí Notepad++ opouštět.



Obrázek 3: Dokumenty, kódy a ukázka textu po importování předzpracovaného TXT souboru

FREKVENCE SLOV

Vzhledem k vymezení kontextové jednotky, která odpovídala jedné větě obsahující řetězec „bezdom“, se frekvence slov omezily pouze

lonce „najít“ a (?1#TEXT)(?2#ENDCODE\r#CODEAutor\r)(?3#ENDCODE\r#CODEStrana\r)(?4#ENDCODE\r#CODEZdroj\\)(?5#ENDCODE\r#CODEOblast\\)(?6#ENDCODE\r#CODEZkratka\r)(?7#ENDCODE\r#CODENazev\r)(?8#ENDCODE\r#CODEClanek\r)(?9#ENDCODE\r) pro „nahradit“.

na takto definované části korpusu. V každém projektovém souboru byl automaticky vyhledán zmíněný řetězec a posléze automaticky okódována celá věta (kontext). První tři až čtyři projekty byly hotové, připravené k frekvenční analýze, v řádech desítek minut (průměrně 6000 vyhledaných slov a zhruba o třetinu menší počet kódovaných vět), kdežto závěrečné „dojezdovou dobu“ kódů znatelně prodlužovaly. Zpracování posledních tří trvalo již několik hodin a bylo opět doprovázeno zdánlivým, a o to nepříjemnějším, „zamrzáním“ (podobně jako v případě importu dat). Přitom pokud porovnáme MAXQDA například s bezplatným softwarem AntConc (Anthony 2016),⁹ rozdíly v čase dokončení jsou poměrně výrazné.

V obecné rovině byly seznamy frekventovaných slov získávány aplikací stejnojmenné funkce (tj. Word Frequency) kvantitativního balíku MAXDictio a postupnou tříúrovňovou aktivací: (1) všech dokumentů v projektovém souboru (tj. odpovídající jednomu roku) a (2) kódu vytvořeného v předchozím kroku, který shromažďoval věty s řetězcem „bezdom“. Tím bylo zaručeno, že se seznamy vztahují pouze a jen k nejbližšímu okolí zvoleného řetězce. K primárním „ročním“ seznamům byly díky úvodní kategorizaci projektových souborů zhotoveny dílčí přehledy, které sledovaly frekvence (v jednotlivých letech) (3) dle kategorií médií. Původních devět typů bylo s ohledem na poměr časové náročnosti / analytického přínosu zkráceno na šestici celostátní deníky, regionální deníky, (regionální i celostátní) časopisy, (regionální i celostátní) rozhlasové stanice, televizní stanice a internetová média. Během celého procesu byl kontinuálně vytvářen a aplikován tzv. stoplist, tj. seznam (z pohledu analýzy) irelevantních řetězců, které MAXDictio při vyhledávání / shromažďování slov ignoruje. Jakkoli se lze setkat s celou řadou podobných předpřipravených seznamů, data byla

filtrována výhradně řetězci vzešlými z korpusu. Výstupy frekvenčních analýz, které jsou v prostředí MAXQDA poměrně rychle hotové (řádově minuty), dovolují jednoduchým dvojklikem, popř. prostřednictvím kontextového menu, vyřadit nechtěné položky z další fáze analýzy.

Avšak je nutné se stoplistem absolvovat několik kol, tj. aplikovat frekvenční analýzu minimálně dvakrát s každou součástí korpusu (srov. Léger 2009, 15). Nakonec jsou relativně „čistě“ frekvenční seznamy sloučeny do jediné obří tabulky a zbaveny opakujících se řetězců. Sestává tedy ze čtyř sloupců udávajících absolutní a relativní četnosti, délku a podobu nalezených řetězců. Jak napovídá Tabulka 3, sloupec vyhledaných slov bude pravděpodobně mimo jiné tvořen morfologickými variantami (např. lidé, lidi nebo lidí a centra nebo centrum) určitého (menšího) počtu odpovídajících slovních základů (např. člověk a centrum).

Slovo	Délka slova	Absolutní četnost	Relativní četnost
policie	7	188	0,42
nádraží	7	157	0,35
města	5	156	0,35
lidé	4	136	0,30
lidí	4	127	0,28
problém	7	122	0,27
sociální	8	118	0,26
ulici	5	115	0,25
strážníci	9	101	0,22
soud	4	97	0,21
zemřel	6	87	0,19
centra	6	86	0,19
policisté	9	83	0,18
lidi	4	81	0,18
pomoc	5	81	0,18
jídlo	5	77	0,17
městské	7	74	0,16
vězení	6	74	0,16
naděje	6	73	0,16
centrum	7	70	0,15

Tabulka 3: Příklad (průběžného) výstupu funkce frekvence slov (TOP 20 slov)

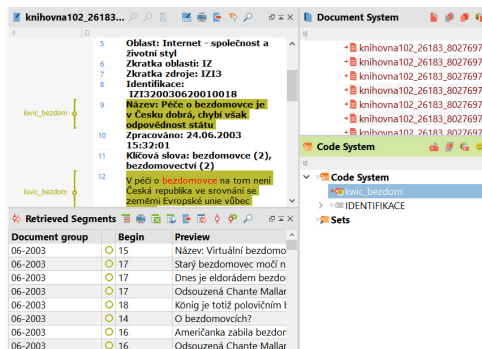
LEMMAIZACE A SLOVNÍK

Ohebnost českých slov může mít mimo jiné za následek nadreprezentaci či podreprezentaci některých slov / řetězců – méně ohebná slova (používaná v menším množství tvarů) mohou být umístěna v TOP50, kdežto ohebnější slovo

⁹ AntConc je (stejně jako zmíněný SarAnt) součástí sady multiplatformních nástrojů vyvíjených Laurencem Anthonym, která funkčně pokrývá oblasti čištění / zpracovávání dat a provedení několika typů lingvistických / obsahových analýz (k celé sadě viz <http://www.laurenceanthony.net/software.html>, viděno 26. ledna 2016). V českém sociálně-vědním prostředí je při analýzách metafor, resp. zpracování většího množství dat, nejvíce užíván Jiřím Mertlem (2013a; 2013b; 2014).

ším korpusům adekvátní, a jednak za níž přestávají být vizualizace přehledné a užitečné.¹⁴

Vzhledem k tomu, že byla předpokládána následná analýza sítí v UCINET a příprava díky aplikaci MorphoDiTa znamenala minimální úsilí, byl počet lemmat ve slovníku s průměrně třemi konkrétními slovy ponechán v poloze otázky: počet slov (první verze obsahovala 2335 slov) byl postupně snižován s cílem zjistit, do jaké míry je MAXQDA schopno se vypořádat s dalšími zátěžemi. Oproti předchozím nastavením není třeba aktivovat ani kódy, ani dokumenty, jelikož MAXDictio Kodér (Coder) umožňuje situovat prostým zatrhnutím hledání slov do oblasti konkrétního kódu, případně více kódů naráz. Nikoli překvapivě se zanedlouho ukázalo, že řády tisíců slov jsou alespoň v případě projektů obsahujících velké množství menších souborů nad síly MAXQDA. Nicméně ani snižování laťky prostřednictvím promazávání slovníku nepřinášelo zdaleka takové zrychlení vyhledávání – např. 300 lemmat odpovídalo prodlevě velmi výrazně přesahující pět hodin.



Obrázek 5: Pomocný kód s kontextovými větami

¹⁴ V rámci kvalitativních analýz může být stanovení počtu slov zcela na rozhodnutí daného autora. Kupříkladu v zásadě zcela arbitrárně jsme postupovali s Jiřím Mertlem při kvalitativní analýze metafor (Mertl a Hejnal 2013), která neměla prakticky žádné kvantitativní (statistické) ambice a frekvenční seznamy, resp. vytvořené slovníky a jejich využití měly pouze naznačit, která slova by mohla být analyticky užitečná (tj. potenciální metafory). Po prvotním seznámení s četnostmi jsme se rozhodli pro hranici 300 (nelemmatizovaných) slov. Vzhledem k tomu, že jsme neměly v úmyslu vytvářet (na základě kvantifikovaných dat) vizualizace, hranice v zásadě vyjadřuje čas, který jsme mohli analýze věnovat.

K tomu je nutné přičíst, chceme-li dále s daty pracovat v MAXQDA, dobu nutnou na převod výsledků kodéru buď do dokumentových proměnných, anebo nechat autokódovat jednotlivé věty novými kódy, jejichž názvy by odpovídaly hledaným slovům a obsahy (segmenty) větám, v nich se příslušná slova vyskytují. V situaci, kdy se hledání kodéru omezovalo na věty s „bezdom“, lze využít první, mnohem kratší dobu trvající, možnost. Elegantnější a ve vztahu k některým dalším funkcím MAXQDA příhodnější druhá alternativa je časově náročnější – téměř se vyrovná čekání na výsledek kodéru. Vyjádřeno v časových jednotkách otázka zní, zdali strávit s jedním projektem dalších pět hodin (druhá možnost), anebo u každého přechkat přibližně 90 minut (první možnost). Souhrnně to znamená, že zmíněný korpus 17 projektů by byl připraven k analýze v případě přívětivější varianty za 110,5 hodin ($[17 \cdot 5 \text{ kodér}] + [17 \cdot 1,5 \text{ dokumentové proměnné}]$), v případě elegantnější za 170 hodin ($[17 \cdot 5 \text{ kodér}] + [17 \cdot 5 \text{ autokódování výsledků kodéru}]$). Bez ohledu na některé těžko predikovatelné faktory (např. vypnutí proudu, „runtime error“¹⁵ atd.) je před uživatelem ještě téměř týden nebo více než týden čistého času práce. Zároveň se navrácí téma zmíněné „těžkopádnosti“, jelikož již první dva počtem souborů i bitovou velikostí suverénně nejmenší projekty (tj. roky 1996 a 1997) začínaly takový rys vykazovat. Ani jedna z cest k cíli patrně nevede.

„PŘEDZPRACOVANÁ“ TABULKA

Po vyhodnocení těchto dvou možností, resp. jejich *časové náročnosti*, bylo přistoupeno k jinému postupu, který sice v mnoha ohledech není optimální, zato bylo jeho zakončení opět smysluplně vyjádřitelné v řádu hodin, nikoli dnů. Cílem následujícího postupu je vytvořit nové, menší a „rychlejší“ MAXQDA projekty, které budou obsahovat pouze určité relevantní části dat (v tomto případě KWIC vět) z projektů předchozích. Z projektů byly v tabulkové podobě (Tabulka 3) prostřednictvím aktivace

¹⁵ Problém objevující se v MAXQDA verzi 10, 11 a 12 souvisí s nekompatibilitou s některými balíky „Microsoft Visual C++ [rok] [x64 nebo x86] Redistributable“. Po (čisté) instalaci balíku problém zpravidla mizí.

Document group	Document name	Segment	ŠOBLAST	ŠZDROJ
200603	200603_K001	Název: Virtuální bezdomovci	Časopisy	Týden
200603	200603_K002	Starý bezdomovec močí na veřejnosti.	Časopisy	Strategie
200603	200603_K003	Dnes je eldorádem bezdomovců a narkomanů.	Celostátní deníky	Blesk

Tabulka 4: Příklad tabulky pro import „předzpracovaných excelových souborů“

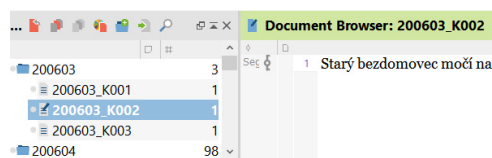
všech dokumentů a stávajícího kódu, který zachycoval věty s „bezdom“ (Obrázek 5), exportovány získané segmenty (tj. KWIC věty) s aktivovanou možností zahrnout do řádku příslušného segmentu / věty všechny proměnné / kategorie (tj. typ média a konkrétní médium). Vzniklo 17 tabulek, jejichž řádky zahrnovaly krom názvu dokumentu, dokumentové skupiny a daného segmentu (tj. KWIC věty) i datum, název média, typ nebo krajskou příslušnost. V této podobě jsou v zásadě připraveny pro import „předzpracovaných tabulek“, což je možnost ekvivalentní zpočátku zmiňované funkci pro částečně strukturované texty.

Základními parametry jsou pojmenování prvních dvou sloupců „Document group“, resp. „Document name“ a použit znak „\$“ před názvem budoucí dokumentové proměnné (viz Tabulka 4). Po kliknutí na možnost *Import documents from Excel spreadsheet* a zadání, co bude proměnná anebo kód, se začne MAXQDA projekt zaplňovat dokumenty (tj. kontextovými větami), k nimž budou automaticky přiřazeny proměnné udávající zdroj a typ média. Proto jsem v databázovém programu Access přes tabulkový Excel¹⁶ vytvořil na základě exportovaných dat z MAXQDA (tj. zhruba Tabulka 4 před zmíněnými úpravami) několik dílčích (a zpětně importovatelných) tabulek dvou základních úrovní dále členěných na základě shluku několika roků, typu média a kraje:

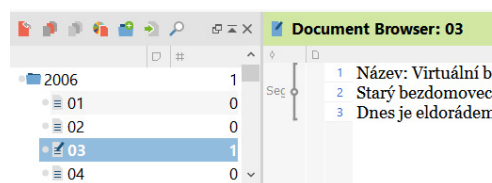
- (a) *detailní*, tj. (jedna KWIC) věta = (jeden) dokument (v projektu) a
- (b) *souhrnnou*, tj. (všechny KWIC) věty z jednoho měsíce = (jeden) dokument (v projektu).

¹⁶ Oba programy sady Microsoft Office je pochopitelně možné nahradit odpovídajícími součástmi nezaplatněných kancelářských sad – například Libre Office či Open Office.

Pro obě úrovně vzniklo devět „krajských“, šest celostátních a po jednom televizním a rozhlasovém. Příklad importu dat z Tabulky 4 („detailní“ úroveň) nabízí Obrázek 6. Modelový projekt, vytvořený na základě importu mírně pozmeněných dat z předchozí tabulky, zachycuje Obrázek 7 („souhrnná“ úroveň).



Obrázek 6: Výsledek importu „předzpracovaného excelového souboru“ (příklad detailní úrovně)



Obrázek 7: Výsledek importu „předzpracovaného excelového souboru“ (příklad souhrnné úrovně)

Pominu-li možné potíže při práci se součástmi balíků Microsoft Office (Access a Excel) či Libre Office (Base a Calc), jistá nekomfortnost spočívá v případě detailní úrovně v tom, že jedno mediální sdělení se velmi pravděpodobně rozložilo do více dokumentů a ve ztrátě širšího kontextu věty (v odstavci či celém mediálním sdělení), zatímco v případě shluknuté úrovně se ztratila část konkrétnosti vět.

Se skromnějším slovníkem (73 vybraných lemmat) a naznačeným způsobem upravenými MAXQDA projekty již bylo možné získat kvantitativní vyjádření překryvu, spolupůsoby klíčových slov (resp. lemmat). Projekty z úrovně detailní („větné“) bylo (z hlediska časové investice) efektivnější nechat

„narůst“ o proměnné, kdežto souhrnné („měsíční“) nakódovat – v obou případech jsou výstupem dvou-modální tabulky. Jediným rozdílem je možnost aplikovat na projekty z druhé jmenované úrovně funkci *Code Relation Browser*, a získat tím přímo jedno-modální matici se vzájemnými překryvy, spolupůskyty lemmat. Exportována do Excelu (nebo jeho freeware obdoby) se tento výstup stává snadno stravitelným vstupem pro UCINET, tj. program pro analýzu (nejen sociálních) sítí, s jehož pomocí se lze po několika málo kliknutích dopracovat k síťovým (či jiným) vizualizacím v zabudovaném NetDRAW, které měly sloužit coby prostředek ke *snadnému a rychlému* seznámení s obsahem datového korpusu. Avšak vzhledem k několikátýdenním potížím patrně není použití ani jednoho z těchto přívlastků empiricky obhajitelné.

ZÁVĚR

Jak jsem již podotkl, MAXQDA je řazen k nejpoužívanějším zástupcům „kaktusů“, tzv. CAQDAS, tj. softwaru určeného pro kvalitativní analýzu dat. Mezi jeho přednosti patří (přinejmenším do verze 11) příjemné pracovní prostředí a snadno naučitelné, v mnoha ohledech variabilní ovládání základních funkcí, které jsou s tímto typem softwaru spojovány – tj. zejména kódování částí textu (popř. audiovizuálních dat) a možnosti znovu / vyvolání kýžených segmentů textu (popř. zvuku / obrazu) dle vytvořených kódů. V tomto ohledu je MAXQDA spolu s Atlas.ti a NVivo právem součástí pomyslného „trojlístku“ absolutní špičky mezi CAQDAS. Všechny tři zmíněné programy nabízejí různé rozsáhlé a integrované doplňky pro kvantitativní analýzy.

Cílem předchozích kapitol bylo poukázat na některé problematické aspekty práce s kvantitativním modulem MAXDictio, který je součástí rozšířené, „plusové“ verze MAXQDA. Na základě rozboru dílčích etap zpracování mediálního datového souboru text naznačil několik problematických míst či limitací, která s sebou přináší práce s větším množstvím dat ve zmíněném doplňku, a několik návrhů či možností, co v podobných situacích dělat. Krom toho stať představila doposud nepojednané mož-

nosti při práci s MAXQDA (zejména dva typy importu „předzpracovaných“ souborů) a v příslušných částech zmiňovala i některé další, více či méně známé programy, jejichž zapojení do procesu analýzy coby určitých doplňkových „berliček“ dokumentuje Obrázek 8.

Oproti bezproblémovému chodu prvků MAXQDA určených pro kvalitativní analýzu, resp. jeho popisu v odborné literatuře, se prezentovaný postup v kvantitativním doplňku MAXDictio stává přehlídkou toho, co MAXQDA (možná prozatím, ale přeci jen spíše) *neumí*. Byť se obtíže s chodem programu přinejmenším v případě původní představy „vše v jednom“ projektu nedaly vyloučit, následné pokusy s dílčími, (mnohem) menšími (pod)soubory dat nasvědčují, že v současné podobě jsou kvantitativní prvky MAXQDA, mezi něž řadím výhradně funkce rozšíření MAXDictio (tj. výčtem frekvenční seznamy, stoplisty, slovníky, vyhledávání slovníkových slov a tzv. MAXDictio Kodér) vhodné při zpracování souborů spíše menšího rozsahu, tj. naneštěstí ve chvílích velmi malého (příp. nulového) užítku z jejich použití. Jinými slovy, MAXDictio, zdá se, funguje perfektně v případech, kdy jeho služby nejsou příliš očekávány ani vyžadovány.



Obrázek 8: Zobrazení trajektorie analýzy (od dat k vizualizaci)

PODĚKOVÁNÍ

Touto cestou bych chtěl poděkovat oběma recenzentům za kromobyčejně důkladné zhodnocení rukopisu původní verze studie.

ZDROJE PODPORY

Tato stať vznikla s podporou grantu Západočeské univerzity v Plzni „Přízpusobivost nepřizpusobivě přizpusobených: Marginalita, racionalita, chudoba“ (SGS-2015-079).

POUŽITÉ ZDROJE

- Alexa, Melina. 1997. *Computer assisted text analysis methodology in the Social Sciences*. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/97_07.pdf.
- Anthony, Laurence. 2016. *AntConc (3.4.4)* [software]. Tokyo: Waseda University. <http://www.laurenceanthony.net/software/antconc/>.
- Bastian, Mathieu, Sebastien Heymann a Mathieu Jacomy. 2009. *Gephi* [software]. <http://www.aiai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- Bastian, Mathieu, Sebastien Heymann, Mathieu Jacomy et al. 2009. „Gephi: An open source software for exploring and manipulating networks.“ *International AAAI Conference on Weblogs and Social Media* 8: 361–362.
- Batagelj, Vladimir, Andrej Mrvar a Matjaz Zaversnik. 2002. *Network analysis of texts*. Ljubljana: University of Ljubljana.
- Bender-deMoll, Skye a Daniel A. McFarland. 2006. „The art and science of dynamic network visualization.“ *Journal of Social Structure* 7 (2): 1–38.
- Borgatti, Stephen P., Martin G. Everett a L. C. Freeman. 2002. *Ucinet for Windows* [software]. Harvard: Analytic Technologies. <https://sites.google.com/site/ucinetsoftware/downloads>.
- Carley, Kathleen M., Jana Diesner, Jeffrey Reminga a Maksim Tsvetovat. 2007. „Toward an interoperable dynamic network analysis toolkit.“ *Decision Support Systems* 43 (4): 1324–47. <http://www.sciencedirect.com/science/article/pii/S0167923606000601>.
- Čepelák, Václav. 2013. „Kvalita modelů a volba parametrů počítačové textové analýzy v programu COOA.“ *Data a výzkum – SDA Info* 7 (1): 7–27. <http://dlib.lib.cas.cz/7992/>.
- Diesner, Jana a Kathleen M. Carley. 2004a. „Revealing social structure from texts.“ *Causal mapping for research in information technology* 81 (3): 65–72.
- Diesner, Jana a Kathleen M. Carley. 2004b. „Using network text analysis to detect the organizational structure of covert networks.“ In *Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference*. http://alliance.casos.cs.cmu.edu/publications/papers/NAACSOS_2004_Diesner_Carley_Detect_Covert_Networks.pdf.
- Fielding, Nigel a César A. Cisneros-Puebla. 2009. „CAQDAS-GIS Convergence Toward a New Integrated Mixed Method Research Practice?“ *Journal of Mixed Methods Research* 3 (4): 349–70. doi:10.1177/1558689809344973.
- Fielding, Nigel a Raymond M. Lee. 2002. „New Patterns in the Adoption and Use of Qualitative Software.“ *Field Methods* 14 (2): 197–216. doi:10.1177/1525822X02014002005.
- Hájek, Martin. 2014. *Čtenář a stroj: Vybrané metody sociálněvědní analýzy textů*. Praha: SLON.
- Hájek, Martin. 2010. „Počítačová textová analýza metodou sledování spoluvýskytů slov.“ *Data a výzkum – SDA Info* 4 (1): 19–37.
- Hejnal, Ondřej. 2012. „Nationalismus, multikulturalismus, sociální vyloučení a ‚sociálně nepřizpusobiví‘: Analýza dominantně politického diskursu v České republice (2006–2011).“ *AntropoWebzin* 2/2012: 47–66.
- Hejnal, Ondřej a Ľubomír Lupták. 2015. „Úvod do MAXQDA: Kvalitativní, kvantitativní a smíšený výzkumný design.“ In *Metody etnografického výzkumu*, ed. Ladislav Toušek, Gabriela Fatková, Lenka Budilová-Jakoubková, Michal Růžička a Jan Šimek. Plzeň: Západočeská univerzita v Plzni.
- Hejnal, Ondřej a Ľubomír Lupták. 2013a. „(De)sekuritizácia bezdomovectva v postsocialistickom urbánnom laboratóriu.“ In *Bezpečnostné fórum 2013: Zborník vedeckých prác*, ed. Jaroslav Ušiak, Jana Lasicová a Dávid Kollár, 478–488. Banská Bystrica: Univerzita Mateja Bela v Banskej Bystrici.
- Hejnal, Ondřej a Ľubomír Lupták. 2013b. „Nyužitie CAQDAS pri výskume sekuritizácie a desekuritizácie.“ In *Bezpečnostné fórum 2013: Zborník vedeckých prác*, ed. Jaroslav Ušiak, Jana Lasicová a Dávid Kollár, 232–239. Banská Bystrica: Univerzita Mateja Bela v Banskej Bystrici.
- Koenig, Thomas. 2004. „Routinizing frame analysis through the use of CAQDAS.“ RC33, Amsterdam. http://www.restore.ac.uk/lboro/research/frames_and_CAQDAS.pdf.
- Krčál, Petr. 2013. „Makroskopická perspektiva politické filosofie versus mikroskopická perspektiva empirie.“ *AntropoWebzin* 4/2013: 155–164.
- Krčál, Petr. 2012. „Možnosti využití dramaturgické analýzy jakožto metodologického nástroje k výzkumu bezpečnosti.“ In *Zborník príspevkov z V. medzinárodnej vedeckej konferencie*, ed. Dávid Kollár, Jana Lasicová a Jaroslav Ušiak. Banská Bystrica: Univerzita Mateja Bela v Banskej Bystrici.
- Krippendorff, Klaus. 2012. *Content analysis: An introduction to its methodology*. London: SAGE.
- Kuckartz, Anne M. a Michael J. Sharp. 2011. „Responsibility: A Key Category for Understanding the Discourse on the Financial Crisis – Analyzing the KWALON Data Set with MAXQDA 10.“ *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12: Art. 22.
- Kuckartz, Udo. 2016. *MAXQDA 12.0.3* [software]. Berlin: VERBI Software – Consult – Sozialforschung GmbH. <http://www.maxqda.com/>.
- Léger, Mathilde de Saint. 2009. *World Politics & “Parapolitics 2006”: Computer-Assisted Text Analysis of Intern*. Editions L’Harmattan.

- Leydesdorff, Loet a Thomas Schank. 2008. „Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments.“ *Journal of the American Society for Information Science and Technology* 59 (11): 1810–18.
- Leydesdorff, Loet, Thomas Schank, Andrea Scharnhorst a Wouter De Nooy. 2008. „Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling.“ arXiv: 0809.4655. <http://arxiv.org/abs/0809.4655>.
- Lu, Chi-Jung a Stuart W. Shulman. 2008. „Rigor and flexibility in computer-based qualitative research: Introducing the Coding Analysis Toolkit.“ *International Journal of Multiple Research Approaches* 2 (1): 105–117. doi:10.5172/mra.455.2.1.105.
- Mertl, Jiří. 2014. „Kvalifikační standardy lidských zdrojů aneb metafory v reformách českého vysokého školství.“ *Acta Politologica* 6 (1): 71–98.
- Mertl, Jiří. 2013a. „Analýza metafor jako potenciálně zajímavý nástroj pro analýzu bezpečnosti.“ In *Bezpečnost v době neklidu: Sborník příspěvků ze studentské konference*, ed. Oldřich Bureš, 64–72. Praha: Metropolitan University Prague Press.
- Mertl, Jiří. 2013b. „Přerodělování welfare: Nástroj pomoci nebo kontroly?“ In *Mladá politologie: Sborník z konference studentů doktorských studijních programů v oblasti politických věd*, ed. Petr Krčál a Vladimír Naxera, 139–50. Plzeň: Západočeská univerzita v Plzni.
- Mertl, Jiří a Ondřej Hejnal. 2013. „Metafory, bezdomovci a média: Metodologický nástin využití analýzy metafor a podpůrného softwaru (MAXQDA a ANTCONC).“ *AntropoWebzin* 4/2013: 145–153.
- Newman, Mark, Albert-László Barabasi a Duncan J. Watts. 2006. *The structure and dynamics of networks*. Princeton: Princeton University Press.
- Nohlová, Barbora a Pavel Nikolin. 2015. „Využití diskurzivní analýzy v aplikovaném antropologicko-sociologickém výzkumu s pomocí statistických programů HAMLET 2.0 a TextSTAT-2.9c.“ *AntropoWebzin* 3-4/2015: 85–104.
- Popping, Roel. 2000. *Computer-Assisted Text Analysis*. London: SAGE.
- Riffe, Daniel, Stephen Lacy a Frederick G. Fico. 2005. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. London: Lawrence Erlbaum Associates.
- Saillard, Elif Kuş. 2011. „Systematic versus interpretive analysis with two CAQDAS packages: NVivo and MAXQDA.“ *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12: Art. 34. <http://qualitative-research.deparraf.com/index.php/fqs/article/view/1518/3133>.
- Schönfelder, Walter. 2011. „CAQDAS and Qualitative Syllogism Logic – NVivo 8 and MAXQDA 10 Compared.“ *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12: Art. 21.
- Silver, Christina a Ann Lewins. 2014. *Using Software in Qualitative Research: A Step-by-Step Guide*. London: SAGE.
- Spoustová, Drahomíra, Jan Hajič, Jan Raab a Miroslav Spousta. 2009. „Semi-Supervised Training for the Averaged Perceptron POS Tagger.“ In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 763–71. Athens: Association for Computational Linguistics.
- Straka, Milan a Jana Straková. 2013. „Czech Models (MorFFlex CZ + PDT) for MorphoDiTa.“ <http://ufal.mff.cuni.cz/morphodita/users-manual#czech-morfflex-pdt>, <https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0023-68D8-1>.
- Straková, Jana, Milan Straka a Jan Hajič. 2014. „Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition.“ *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13–18.
- Tambayong, Laurent a Kathleen M. Carley. 2012. „Network text analysis in computer-intensive rapid ethnography retrieval: An example from political networks of Sudan.“ *Journal of Social Structure* 13 (2): 1–24.
- Toušek, Ladislav. 2013. *Prostor, transgrese a bezdomovectví*. Disertační práce. Plzeň: Západočeská univerzita v Plzni.
- Vašát, Petr a Daniel Čermák. 2015. „Referendum, nebo účet krajským vládám? Volební programy 2012.“ In *Na/O kraji: Kraje a jejich představitelé 2000–2013*, ed. Dan Ryšavý a Daniel Čermák, 126–147. Praha: SLON.
- Weber, Robert Philip. 1990. *Basic content analysis*. London: SAGE.