

Filozofická fakulta Univerzity Palackého v Olomouci
Katedra obecné lingvistiky



Analýza metafor: teorie a počítačně lingvistické přístupy

magisterská diplomová práce

Autor: Bc. Dalibor Pavlas
Vedoucí práce: Mgr. Kateřina Veselovská, Ph.D.

Olomouc
2016

Prohlášení

Prohlašuji, že jsem magisterskou diplomovou práci „Analýza metafory: teorie a počítačně lingvistické přístupy“ vypracoval samostatně a uvedl jsem veškerou použitou literaturu a veškeré použité zdroje.

V Olomouci dne 5. 5. 2016

Podpis

Tato diplomová práce vznikla v rámci SGS projektu Experimentální a kvantitativní postupy v analýze textu a jazyková diverzita, a to díky podpoře specifického výzkumu alokované Univerzitě Palackého v roce 2015.

Poděkování

Rád bych poděkoval své vedoucí práce za cenné rady a své rodině a přátelům za podporu, díky které mohla tato práce vzniknout.

Abstrakt

Název práce: Analýza metafory: teorie a komputačně lingvistické přístupy

Autor práce: Dalibor Pavlas

Vedoucí práce: Mgr. Kateřina Veselovská, Ph.D.

Počet stran a znaků: 123234 znaků, 91 stran

Počet příloh: 0

Abstrakt: Práce se zabývá možnostmi automatické analýzy metafory. Popisuje čtyři teorie metafory, které jsou nejvíce relevantní pro její počítačové zpracování (interakční teorie, porušení selekčních preferencí, teorie konceptuální metafory a teorie strukturálního mapování) a jejich specifické výhody a nedokonalosti. Ukazuje, že metafora je v jazyce všudypřítomným fenoménem a jaké benefity může přinést zvládnutí jejího počítačového zpracování. Popisuje specifické problémy, které automatickou analýzu metafory znesnadňují a ukazuje, jakými způsoby a za použití jakých zdrojů a nástrojů byla problematika automatické identifikace a automatické interpretace metafory dosud řešena, a jak by mohla být řešena v budoucnu za pomoci nových poznátků lexikálně-sémantických teorií. Popisuje situaci v českém prostředí a snaží se teoretické a praktické poznatky z předchozích částí aplikovat na češtinu, a přijít se způsoby, jakými by se šlo v českém jazyce úlohou automatické identifikace metafory co nejefektivněji zabývat.

Klíčová slova: metafora, nlp, automatická identifikace metafory, obrazné vyjadřování, komputační lingvistika, teorie metafory

Abstract

Title: Metaphor analysis: theories of metaphor and approaches in NLP

Author: Dalibor Pavlas

Supervisor: Mgr. Kateřina Veselovská, Ph.D.

Number of pages and characters: 123234 characters, 91 pages

Number of appendices: 0

Abstract: Thesis deals with possible solutions of automatic metaphor analysis. Four most prominent theories of metaphor are described (interaction theory, selectional preferences violation view, conceptual metaphor theory and structure-mapping theory) with both their advantages and pitfalls. It shows that metaphor is ubiquitous in language and what are the exact benefits of its successful automatic processing. Thesis describes specific problems which make automatic metaphor recognition and automatic metaphor interpretation such a hard task in NLP. It shows some known approaches to the task and their resources, tools and problem solutions and also some possibilities of future progression of the task using new development in lexical semantics. It then describes situation in czech linguistics and tries to apply knowledge from previous parts in order to show ways of dealing with automatic metaphor processing in czech language.

Keywords: metaphor, nlp, automatic metaphor identification, figurative language, computational linguistics, theory of metaphor

Obsah

Úvod.....	8
1 Metafora: vymezení a terminologie	10
2 Teorie metafory	13
2.1 Odklon od klasického pojetí.....	13
2.2 Interakční pojetí.....	15
2.3 Porušení selekčních preferencí.....	18
2.4 Teorie konceptuální metafory	19
2.5 Teorie strukturálního mapování	25
2.6 Možnosti aplikace teoretických přístupů.....	27
2.7 Doslovnost a obraznost	28
2.8 Stupně metaforičnosti	31
2.8.1 Přístup Patricka Hankse	32
2.8.2 Metoda Jonathana Dunna	33
2.8.3 Možnosti aplikace metod měření metaforičnosti	36
2.9 Všudypřítomná metafora.....	36
3 Analýza metafory v rámci NLP.....	38
3.1 Proč se bychom se měli pokoušet o automatické zpracování metafory.....	38
3.2 Proč je úloha složitá	40
3.3 Způsoby počítačového zpracování metafory.....	41
3.4 Zdroje informací pro počítačové zpracování metafory	42
3.4.1 Znalostní báze	42
3.4.2 Korpusy	44
3.4.3 Využití nových poznatků lexikálně-sémantických teorií.....	45
3.5 Dosavadní systémy pro automatické zpracování metafory	47
3.5.1 Základní metody evaluace výsledků	48
3.5.2 První systémy z počátku 90. let	49
3.5.3 Systémy založené na reasoningu ve zdrojové doméně.....	50
3.5.4 Automatická identifikace konceptuální metafory	52
3.5.5 Identifikace metafory založená na významech ve WordNetu.....	53
3.5.6 “Sardonicus”	54
3.5.7 Přístupy založené na clusteringu	55
3.5.8 Vyhledávání metafor na základě abstraktnosti kontextu.....	57
3.5.9 Jazykově nezávislá identifikace metafory.....	57
4 Analýza metafory v češtině.....	59
4.1 Více komplikací, méně zdrojů	59
4.2 Korpusové experimenty.....	60

4.2.1 Experiment 1: Goatlyho lingvistické ukazatele metafory	60
4.2.2 Experiment 2: Schémata bohatá na metafory v češtině	62
4.3 Využití zdrojů dostupných pro češtinu	65
4.3.1 Valenční lexikon VerbaLex	66
4.3.2 Ontologie Sholva	68
4.4 Převedení existujících systémů do češtiny	69
4.4.1 Anotace korpusu pro trénink a evaluaci	69
4.4.2 Replikace přístupu založeného na abstraktnosti kontextu	70
4.5 Shrnutí	72
Závěr	73
Použité zdroje:	74

Úvod

“[M]etaphors do not cause us to view the world in a whole new light, but rather enable us to see the world as we do.” (Martin, 1990, Preface xxi)

Počítačové zpracování přirozeného jazyka (*natural language processing*; NLP) je disciplína, která od svých počátků v 50. letech 20. století udělala nepochybně obrovské pokroky. Počítačové programy dokáží velmi spolehlivě určovat slovní druhy (PoS tagging) nebo syntaktickou strukturu vět (parsing). Na druhou stranu strojové překladače zatím nejsou schopné nahradit lidské překladače. V červnu 2014, když židovský chlapec Eugene Goostman prošel legendárním Turingovým testem, nikdo se nesnažil skutečně bojovat za jeho práva myslící bytosti. Komunikace s počítačem nám stále nepřipadá lidská, a pokroky ve výpočetní kapacitě počítačů předčí jejich pokroky ve zpracování jazyka.

Jedním z důvodů je, že se donedávna nevěnovala dostatečná pozornost tomu, že věty jazyka neznamenaají vždy to, co kombinace významů jejich slov. Lidé se často vyjadřují *obrazně*. Mezi druhy obrazného vyjadřování patří např. ironie, metonymie, hyperbola a metafora. Poslední zmíněná je tématem této práce.

Metafora je v jazyce rozšířeným fenoménem, používá se v poezii, ale také v běžné komunikaci a dokonce i ve vědě. Metafora je mluvení o něčem, jako o něčem jiném, často o něčem složitém či abstraktním, jako o něčem jednoduchém či konkrétním (příklad (1)).

(1) *Utíká nám čas.* (SYN2010)

Tato práce se zabývá možnostmi analýzy metafory v rámci NLP, je rozdělena do třech hlavních kapitol (první kapitola je pouze terminologický úvod), druhá kapitola se zabývá hlavními teoriemi metafory a problémy teoretického rázu, jako je určení hranice mezi doslovným a metaforickým. Třetí kapitola popisuje zpracování metafory v NLP, rozdělení na jednotlivé úlohy a úrovně, na kterých pracují a popisu způsobu evaluace. Čtvrtá

kapitola se snaží nastínit možnosti automatického zpracování metafory v češtině, kde dosud žádný počítačový přístup nevznikl.

K použitým příkladům: cizojazyčné (zpravidla anglické) příklady jiných autorů překládám, pokud je to možné, do češtiny. Při volnějším překladech a přizpůsobeních cizích příkladů jsem ověřoval přítomnost výsledných vět v korpusu. V některých případech jsem ponechal angličtinu, např. z důvodu ilustrace provázanosti konstituentů výroku s konkrétní znalostní bází. Pokud je to vhodné, uvádím odůvodnění svých kroků a dodatečné informace v poznámce pod čarou.

Metafora: vymezení a terminologie

Metafora je způsobem nedoslovného vyjadřování, který využívá slovník spojený s nějakým fenoménem k popisu jiného fenoménu. Většinou se setkáváme s literárně-vědnou definicí, kdy jde o přenášení významu na základě vnější či strukturální podobnosti (věta (2)). Tím se metafora liší od metonymie, která přenáší význam na základě příčinnosti nebo původu (například z autora na dílo ve větě (3)).

(2) [T]atínek je *hlava* rodiny. (SYN2010)

(3) Platón je na horní policičce. (Martin, 1991)

Tento přenos významu je v případě metafory asymetrický. Pokud obrátíme pořadí pojmů, stane se výrok buď nesmyslným (4), nebo se drasticky změní jeho význam, jako je tomu ve větě (5), kdy se obrátí sémantická orientace z negativní na pozitivní.¹

(4) Julie je slunce. / Slunce je Julie.

(5) Někteří chirurgové jsou řezníci. / Někteří řezníci jsou chirurgové.²

Podle míry ustálenosti v jazyce se používá dělení metafor na *lexikalizované* (ustálené; příklad (6)) a *aktuální* (neustálené; příklad (7)).³ Přitom se jazykovědci rozcházejí v mínění, zda jde u lexikalizovaných metafor vůbec o metafory nebo spíše o další plnohodnotné významy slov. Lexikalizace aktuálních metafor je navíc stále probíhající proces v rámci vývoje jazyka.

¹ Srovnej s bezproblémovým: "Slovníky jsou jako encyklopedie. / Encyklopedie jsou jako slovníky."

² Tento příklad se v různých obměnách často vyskytuje v psychologické literatuře, zabývající se metaforou (např. Ortony, 1979 nebo Glucksberg & Keysar, 1990).

³ Někdy se také užívá termínů *živé* a *mrtvé* metafory.

(6) *noha* stolu; *koryto* řeky

(7) Kolo dějin se nedá zastavit. (SYN2010)

Užívání metafor bývá tradičně spojováno s uměleckým jazykem a je tak považováno za jev, který se objevuje spíše neobvykle, než často, a běžné komunikační situace se bez něj obejdou (Black, 1955, 278; Lakoff & Johnson, 2014, 15).

Od 50. let 20. století bylo však filozofy analytické tradice (Black, 1955) poukazováno na to, že metaforické vyjadřování je široce používané i v běžné komunikaci a v odborném jazyce. Metafora nám totiž pomáhá vyjadřovat abstraktní koncepty za pomoci konkrétnějších pojmů. Užívání metafor tedy mnohdy není jen alternativou k doslovnému jazyku, ale jedinou možností, jak daný význam vyjádřit.

V 80. letech se pak metafora dostává do středu pozornosti lingvistiky a také kognitivních věd a psychologie. Mark Johnson a George Lakoff přicházejí ve svém slavném díle *Metafory, kterými žijeme* (1980) s teorií *konceptuální metafor*⁴. Podle Lakoffa a Johnsona existuje metafora primárně jako kognitivní a až sekundárně jako jazykový fenomén. O věcech mluvíme metaforicky, protože o nich metaforicky přemýšlíme. Toto tvrzení je v současné době lingvistickou komunitou široce přijímáno a je uznáváno i v této práci.

Projev metafor v jazykové rovině nazýváme metaforou *lingvistickou*. Můžeme rozlišit několik druhů metaforických výrazů, nejběžnějším je metafora na úrovni jednotlivých slov, tedy metafora lexikální (8). Z důvodu její jednoduchosti a zároveň vysoké relativní frekvence je lexikální metafora nejvhodnější pro strojovou detekci a interpretaci.

Dále se můžeme setkat s metaforami víceslovnými a nakonec s metaforami *rozšířenými*, kdy metaforičnost zasahuje do většího celku diskurzu. Slavným příkladem rozšířené metafor je shakespearovská analogie světa a jeviště ze hry “Jak se vám líbí” (9).

(8) To jsou jen *prázdná* slova. (Reddy, 1979)

⁴ Někdy bývá používána zkratka CMT (*conceptual metaphor theory*).

(9) Celý svět je jeviště a všichni lidé na něm jenom herci: mají své příchody a odchody; za život každý hraje mnoho rolí (...) ⁵

Ta část vyjádření, která je metaforická, tedy ta jež vykazuje známky nekongruence oproti doslovné interpretaci se nazývá *ohnisko*, zbytek je *rámcem* metafor (Black, 1955) ⁶. V příkladech užitých v této práci označuji, pokud je to možné, ohnisko metafory použitím kurzívy.

Hranici mezi ohniskem a rámcem však v některých případech není možné přesně určit, a v rámci jedné výpovědi může být použito kombinace několika metafor. Přesto se v mnoha případech jedná o užitečné dělení (Kittay, 1990, 66).

Někdy je také nesnadné, či nemožné, oddělit metaforu od jiných forem nedoslovného vyjadřování.

(10) Amerika věří v demokracii. (Hobbs, 1990)

Větu (10) je možné interpretovat jako metaforu, kdy neživou entitu "Amerika" personifikujeme a přiřkneme jí schopnost v něco věřit, avšak stejně dobře může být označena jako metonymie, kdy výraz "Amerika" nahrazuje množinu obyvatel Spojených států amerických.

Je tudíž třeba pamatovat na to, že bychom neměli chápat metaforičnost ani další jazykovědné pojmy esencialisticky. Určení výroku jako metaforický, metonymický, obrazný či doslovný je vždy záležitostí interpretace a nikoli nějaké vnitřní kvality výroku samotného. K dokonalé shodě nedochází ani při určování metaforičnosti několika lidskými anotátory (jakkoli pečlivě školenými). Od automatických systémů pro analýzu metafor tedy nemůžeme čekat zázraky ve formě stoprocentní úspěšnosti.

Následující kapitola rozebírá hlavní teoretické přístupy k fenoménu metafor.

⁵ V překladu Martina Hilského.

⁶ V originále "focus" a "frame", používám české termíny podle J. Stachové (1992); Někdy se používá také ekvivalentních termínů "tenor" a "vehicle", které pocházejí od I. A. Richardse (1936).

Teorie metafory

Metafora je předmětem zkoumání nejpozději od Aristotela. Ten ji v polovině 4. Století př. n. l. v díle *Poetika* definuje jako:

“[P]řenesení jména z jedné věci na druhou, a to buď z rodu na druh, nebo z druhu na rod, nebo z jednoho druhu na jiný, anebo podle obdoby.” (Poetika, paginace 1457b)⁷

Od té doby se pochopitelně myslitelé a vědci tématem zabývali. Oproti původnímu Aristotelovu pojetí byla z metafory vydělena metonymie jako samostatný jev, ale až do 50. let 20. století byla metafora brána jako fenomén, jehož funkce v jazyce je pouze dekorativní. Britští empirici Bacon, Hobbes a Locke kupříkladu považovali užívání metafor ve filozofických a vědeckých textech za nepřijatelné (Forrester, 2010).⁸

Odklon od klasického pojetí

Britsko-americký filozof jazyka Max Black (1955) zavádí pro dosavadní teorie termín *nahrazovací* pojetí metafory. To se na metaforické vyjádření dívá jako náhradu za doslovné vyjádření se stejným významem.

(11) Richard je *lev*. = Richard je statečný. (Black, 1955)⁹

(12) *Kořeny* rovnice

Ve větě (11) nahradil její autor slovo “statečný” slovem “lev”, úlohou čtenáře je použít metaforické vyjádření jako vodítko pro odhalení zamýšleného doslovného významu. Interpretace metafory je tedy proces podobný řešení hádanky. Proč autor nutí čtenáře

⁷ Přeložil František Groh pro vydání Společnosti přátel antické kultury z roku 1929 (Aristoteles, 1993).

⁸ Paradoxně se však sami nedokázali užívání metafor vyhnout (Forrester, 2010).

⁹ Black větu uvádí jako “poněkud nešťastný” běžně užívaný příklad z literatury s tím, že si nedovede představit, že by ji v (tehdejší) současnosti někdo opravdu použil a mínil vážně (Black, 1955, 281).

řešit hádanky? Důvody mohou být dva. Může se stát, že doslovný ekvivalent neexistuje, to je případ mnoha metafor ve vědách a matematice (příklad (12)). V případě, že je ekvivalent k dispozici, jako je tomu u Richardovy statečnosti, bude odpovědí na výše položenou otázku, že byla metafora použita ze stylistických důvodů, k potěše čtenáře. To považuje Black za nedostatečné vysvětlení existence metafory a dalších jazykových jevů. Speciálním případem nahrazovacího pojetí metafory je tzv. *srovnávací* pojetí. Dá se ilustrovat větou (13).

(13) Richard je lev. = Richard je *jako* lev (v tom, že je statečný). (Black, 1955)

V tomto pojetí, které připomíná Aristotelovo, je tedy metafora založena na podobnosti, je jakýmsi zhuštěným přirovnáním. Black však považuje srovnávací vysvětlení za příliš vágní. Podobnosti nejsou objektivně dané, mohou být různého stupně a srovnávat můžeme z různých hledisek.

Zároveň přichází s vlastním pohledem na metaforu, kterému říká *interakční*. Následná debata v analytické filozofii (Black, 1962; Hesse, 1966; Davidson, 1978; Black, 1979; Searle, 1979) posouvá metaforu více do centra zájmu a hlavně ji již vidí jako fenomén nezbytný ke komunikaci a nikoliv pouze jako ozdobný prvek jazyka.

Druhou revolucí ve vnímání metafory s daleko větším dopadem na lingvistiku a další vědní obory pak byla teorie *konceptuální metafory* (Reddy, 1979; Lakoff & Johnson, 1980), která byla základem pro nové odvětví jazykovědy, tzv. *kognitivní lingvistiku* (např. Lakoff, 1987; Langacker, 1987; Fauconnier, 1997; Fauconnier & Turner, 2003).

Teorie *strukturálního mapování*¹⁰ Dedre Gentnerové (1983), která navazuje na *srovnávací* přístupy, není teorií metafory v pravém slova smyslu a nepochází také z prostředí lingvistiky. Jde o dlouhodobě nejprominentnější psychologickou teorii popisující *analogii* (Turney, 2008). Je však obecně uznáváno, že metafora je na analogii založena (Gentner a kol., 2001; Fauconnier & Turner, 2002). Teorie strukturálního mapování je také pro tuto práci zajímavá díky tomu, že poměrně brzy po svém vydání byla formalizována a

¹⁰ Structure-Mapping Theory

implementována do programu, tzv. *Structure-Mapping Engine* (Falkenhainer, Forbus & Gentner, 1989).

Pojetí metafory jako *porušení selekčních preferencí* (Wilks, 1975; 1978) je pak teoretickým přístupem, který byl přímo navržen jako formální systém, který dokáže identifikovat nedoslovné užití jazyka včetně metafory.

Pro všechna čtyři zmíněná pojetí metafory (interakční, konceptuální, strukturálního mapování a porušení selekčních preferencí) platí, že předpokládají nějaký způsob spojení mezi dvěma oblastmi, na kterém se vznik metaforického vyjádření zakládá. Každým z těchto přístupů k metafoře se budu dále zabývat v samostatných podkapitolách.

Pochopitelně existují další teorie metafory, například teorie psychologů Glucksberga a Keysara (1990), kteří vidí metaforu jako interakci v rámci kategorií, nebo deflační přístup teorie relevance D. Sperbera a D. Wilsonové (1995; 2008), kteří tvrdí, že metafora vlastní teorii nepotřebuje. Následující čtyři probírané jsou však v literatuře často uváděny jako nejdůležitější (Tourangeau & Sternberg, 1982; Fass, 1991; Shutova, 2010b) a vychází z nich také naprostá většina komputačních přístupů.

Interakční pojetí

Jako první si představíme Blackovu *interakční* teorii. Oproti klasickému, substitučnímu pojetí, kdy metaforický význam nahrazuje doslovný, přidává další úroveň. Ohnisko metafory získává nový význam, který je rozdílný od jeho doslovného významu, ale také od významu, který by měla jakákoli doslovná náhrada. Metafora je tedy výsledkem interakce (Black, 1962, 39).

(14) John je vlk. (Ayoob, 2007)

K oné interakci nedochází mezi významy jednotlivých pojmů, ale mezi dvěma subjekty, *primárním* a *sekundárním* a jejich *soubory asociovaných implikací*.¹¹ To jsou soubory

¹¹ systems of associated commonplaces

přesvědčení o pojmech, které by očekávali členové dané komunity. U věty (14) tedy půjde o výroky spojené s vlky (ohniskem metafory) a Johnem (rámcem metafory). Black předpokládá, že se na těchto *souborech* mluvčí v naprosté většině případů shodnou, a i odborník, tedy biolog, či etolog, který má o vlčích mnoho informací rozdílných od představ většiny mluvčích, bude vědět, jaké tyto představy jsou. Nezáleží tedy na tom, zda jsou *asociované implikace* přehnané, úplně nepravdivé nebo si dokonce navzájem protiřečí, musí však jít o věci, které se mluvčím ihned vybaví, když se taková metafora objeví.¹²

Nový význam je pak tvořen průnikem obou souborů, John je viděn skrze mřížku pojmu vlk. Některé vlastnosti ohniska i rámce jsou tedy z nového významu odstraněny, například “pracuje v bance”, “žije ve smečkách” nebo “je špatný ve scrabblu” a zůstávají jen ty, které mohou být asociovány s oběma, jako je třeba “je to samotář” nebo “je zlý” (Ayoob, 2007). Black (1955; 1962) navrhuje, že jeho rozdělení teorií metafory na nahrazovací, srovnávací a interakční, může být použito jako dělení samotných nedoslovných výpovědí. Měli bychom tedy *nahrazení*, *srovnání* a pouze poslední kategorie, která nejde nahradit doslovným ekvivalentem, neboť u ní dochází ke změně významu, jak to vysvětluje interakční teorie, by byla skutečnou *metaforou*.

Blackova interakční teorie byla kritizována dalšími analytickými filozofy, zde se jen zběžně podíváme na dva způsoby kritiky, které mohou mít souvislost s automatickou analýzou metafory.

Donald Davidson (1978) se vymezuje proti postulování nového metaforického významu. Ten je podle něj jen rozšířením slovníku o další význam (viz kapitola o doslovnosti) a nikoli metaforou. Podle Davidsona nemají metafory jiný než doslovný význam, metaforičnost je až záležitostí efektu, který má daná výpověď na posluchače/čtenáře při její realizaci. Abychom pochopili metaforu, měli bychom ji tedy zkoumat z vnější, pragmatické stránky. Davidson však nabízí jen málo pozitivních návrhů, jak bychom takový výzkum měli provádět (Black, 1979; Ayoob, 2007).

¹² Tím Black odpovídá již v díle z roku 1955 na jednu z pozdějších námitek J. Searla (1977) proti interakční teorii. Ten totiž argumentuje, že ve větě “Richard je gorila”, která znamená, že je divoký, ošklivý a má sklon k násilí, by se význam nezměnil, pokud by nové poznatky v etologii takové chování u goril vylučovaly a metafora proto nemá nic společného s významem gorily, ale pouze s významem Richarda (Searle, 1977; cit. ze Stachová (1992).

John Searle argumentuje proti roli podobnosti v metaforickém myšlení. Jako příklad uvádí větu (15), původně vyřčenou Romeem v Shakespearově hře.

(15) Julie je *slunce*. (Searle, 1977)

Podle Searla patří mezi asociované implikace slunce např. “být složena z větší části z plynů” nebo “nacházet se milión mil od Země”, což rozhodně nejsou vlastnosti, které lze přiřknout Julii. Celkově nenachází mezi dvěma zmíněnými entitami žádné reálné podobnosti, které by mohly samy o sobě pomoci s interpretací výroku. Stachová (1992) však nabízí příklad takové podobnosti, společnou vlastnost “být nezbytný k životu”, kterou má pro lidi (a jiné živé tvory) slunce a pro Romea Julie, a argumentuje, že podobnost nemusí být záležitostí fyzikálních nebo chemických vlastností, ale může být abstraktnějšího rázu.

Interakční pojetí tedy trpí určitou vágností, když přijde na to, které vlastnosti “projdou” ohniskem a které budou z nového významu vyřazeny. Stejně jako kolik informací jakého druhu může být obsaženo v souborech implikací. Přesto tento přístup někteří autoři dále rozvíjejí.

Bipin Indrukha (1988) vytváří formální pojetí metafory, které nazývá teorie *omezeného sémantického přenosu*, to přímo navazuje na původní interakční pojetí.¹³ Při interpretaci metafory probíhá interakce mezi doménami reprezentovanými zdrojovým a cílovým *implikativním komplexem*. Prvky jsou přeneseny z prvního do druhého, což následně působí změny i ve zdrojové doméně. Metafora je identifikována formálně definovaným konceptem (tzv. T-MAP), sestávajícím z funkce (F), která mapuje slovník zdrojové domény do slovníku cílové, a sady vět zdrojové domény, které mají být přeneseny (S). Vhodnost metafory je hodnocena základě míry koherence, ve smyslu logické odvoditelnosti přenesené sady vět (S) z axiomů cílové domény.

Z interakčního pojetí částečně těží některé NLP systémy (explicitně to tematizuje Fass

¹³ Constrained Semantic Transference Theory; Zároveň využívá i poznatků dalších zmíněných přístupů. Metaforičnost předpokládá strukturální analogii jako u teorie *strukturálního* mapování, a terminologicky se drží modelu *zdrojová x cílová* doména, jak ho používá teorie *konceptuální metafory*.

(1991, 62) u své metody *met**, nebo Veale a Hao (2008)). Obecně je však, co se týče praktických aplikací, spíše v pozadí oproti dalším třem prominentním teoriím metafory, a její novější varianty jsou kritizovány pro zastávání zastaralého subjektivistického pojetí významu (Way, 1993).

Porušení selekčních preferencí

Pohled na metaforu jako porušení selekčních preferencí vychází z článku Yoricka Wilkse *Making preferences more active* (1978) a z jeho předchozí práce (1975) na poli tzv. preferenční sémantiky. Tento přístup je díky tomu, že nabízí způsob, jak metaforu v textu automaticky detekovat, často používanou teorií v NLP aplikacích zaměřených na analýzu metaforických výpovědí (Fass, 1991; Mason, 2004; Barnden & Lee, 2002; Krishnakumaran & Zhu, 2007).

Východiskem jsou Wilksovi tzv. *selekční omezení*, což je termín používaný generativistickou tradicí (Katz & Fodor, 1963), některá slova - týká se to hlavně sloves - mají omezený okruh sémantických typů, se kterými mohou koexistovat v rámci jedné výpovědi. Při porušení jejich selekčních omezení jsou tvořeny negramatické věty. Wilks místo toho používá pojem selekční *preference* a jejich porušování toleruje, ba dokonce s ním počítá. Takové porušení totiž indikuje použití slov v nedoslovném významu.

(16) [P]rohlásil, že je to k vzteku, kolik to auto žere. (SYN2010)¹⁴

Sloveso “žrát” běžně preferuje životný subjekt, jeho užití pro subjekt “auto” je porušením této preference, což ukazuje, že slovo “žrát” by mohlo být použito metaforicky (věta (16)). Tento přístup má některé vážné nedostatky. Jednak, jak už bylo naznačeno výše, neidentifikuje vlastně metaforu, ale obrazné použití slov obecně, výsledky analýzy budou tedy kontaminovány dalšími druhy nedoslovnosti např. metonymií, stejně jako případy

¹⁴ Větu jsem vybral z ČNK jako český ekvivalent k Wilksovu nejznámějšímu příkladu “My car drinks gasoline” (Wilks, 1978)

anomálních spojení. Tento nedostatek není fatální, pokud však chceme detekovat metaforu, je třeba doplnit rozpoznání porušení selekčních preferencí dalším druhem filtru, který oddělí ostatní druhy nedoslovnosti a anomálie (v praxi např. už (Fass, 1991)).

Palčivější problém Wilksovy teorie je naopak v tom, že lze bez problémů tvořit metaforické výpovědi bez porušení selektivních preferencí (věta (17)), některé metafory tedy zůstanou neodhaleny.

(17) Irák je dnešní Vietnam. (Barnden, 2007)¹⁵

Dále, jak poukazuje například Barnden (2007), je Wilksův přístup limitován na porušení preferencí uvnitř jedné výpovědi. V reálných komunikátech se však často stává, že se metaforičnost vyvíjí v rámci delšího úseku textu a pro interpretaci metafory je nezbytná znalost kontextu. Například věta (18) použitá v textu o manželství by měla jistě metaforický význam, ačkoli k porušení selekčních preferencí v ní nedochází.

(18) Jack a Jill se ocitli na rozcestí (Barnden, 2007)

Navzdory těmto výtkám je třeba říct, že metaforické výpovědi často selekční preference skutečně porušují, činí tak například většina konceptuálních metafor, kterým se budu věnovat v další části práce.

Teorie konceptuální metafory

Jak již bylo řečeno, metafora vždy zahrnuje mluvení o něčem jako o něčem jiném. Podle teorie konceptuální metafory (Lakoff & Johnson, 1980) ale není jazyk primární, metaforické jsou do značné míry procesy lidského myšlení (Lakoff & Johnson, 2014, 18). Například konceptuální metafora SPOR/ARGUMENTACE JE VÁLKA, která se může v jazyce

¹⁵ Tento příklad je navíc možné interpretovat jako smíšení metafory a metonymie, kdy je Irákem a Vietnamem myšlena válka v Iráku, resp. Vietnamu.

projevit třeba větami (19)-(21), vypovídá jednak o tom, že používáme pro popis našich sporů/argumentací¹⁶ stejný slovník, jako pro popis válečného konfliktu, ale ukazuje - a to v první řadě - že spor/argumentaci jako válku pojmáme, tímto způsobem o ní přemýšlíme a také v ní tímto způsobem jednáme.

(19) *Napadl každou slabinu* v mé argumentaci. (Lakoff & Johnson, 2014)

(20) Jeho argumenty jsem úplně *rozmetal*. (Lakoff & Johnson, 2014)

(21) Nikdy jsem ho v debatě *neporazil*. (Lakoff & Johnson, 2014)

Schéma konceptuální metafory tedy obsahuje dvě *domény*, *zdrojové* doméně VÁLKA koresponduje *cílová* doména SPOR/ARGUMENTACE. Abstraktnější z dvojice je ve většině případů, včetně tohoto, doménou cílovou. Ta může být někdy pojmána pouze na základě zdrojové domény, aniž by měla svůj originální systém konceptualizace (Kövecses, 2010). Tomuto vztahu zdrojové a cílové domény se říká *mapování*.¹⁷

Mezi další konceptuální metafory patří například LÁSKA JE CESTA (věta 22), VĚDOMÍ JE KŘEHKÁ VĚC (23) nebo ČAS JSOU PENÍŽE (věty 24 a 25). Poslední zmíněná dobře ilustruje, jak běžnou součástí našich životů metafora je. Ve světě ročních rozpočtů a hodinových mezd je v komunikaci neustále používána, ačkoli si většina uživatelů jazyka není vůbec vědoma, že se vyjadřuje metaforicky - protože metaforicky již přemýšlejí. Konceptualizujeme takto čas, díky našim každodenním zkušenostem s penězi, v primitivnějších, méně industrializovaných kulturách nemusí být čas pojmán tímto způsobem.

(22) Tento vztah je *slepá ulička*. (Lakoff & Johnson, 2014)

(23) *Ruplo mu v bedně*. (Lakoff & Johnson, 2014)

¹⁶ V originále "argument", slovo nemá jednoduchý český ekvivalent, který by zároveň vyjádřil oba jeho nejdůležitější významy, použitý zástupný výraz pochází z českého překladu *Metafor, kterými žijeme*. Podobného mezijazykového "šumu" je obecně v konceptuálních metaforách relativně málo oproti metaforám lingvistickým a většinou není problém nalézt český ekvivalent, to samo o sobě nepřímo podporuje adekvátnost celé teorie.

¹⁷ mapping

(24) Tento zlepšováček vám *ušetří* spoustu hodin. (Lakoff & Johnson, 2014)

(25) Ta prasklá duše mě *stála* skoro hodinu práce. (Lakoff & Johnson, 2014)

Lakoff a Johnson také musí řešit problém, kdy se objeví více kompatibilních zdrojových domén. Čas totiž nemusí být jen peníze, ale může stejně dobře platit že ČAS JE OMEZENÝ ZDROJ (a je třeba ho dobře *využít*) nebo ČAS JE CENNÉ ZBOŽÍ (a můžeme ho někomu *dát*, či od někoho *dostat*). Zavádějí proto pravidlo používání nejspécifičtějšího dostupného pojmu, tedy peněz, pro charakterizaci celého systému. Ostatní metaforické koncepty z něj vyplývají, protože peníze jsou omezeným zdrojem a omezenost přístupu k nim z nich činí cenné zboží (Lakoff & Johnson, 2014, 21).

Dalším krokem k větší uspořádanosti celé teorie je rozdělení konceptuálních metafor na *strukturální* a *orientační*. Dosavadní probrané metafory řadí autoři mezi *strukturální*, to znamená, že jeden pojem je metaforicky strukturován na základě druhého pojmu. Orientační metaforou nazývají případy, kdy jeden systém pojmů organizuje jiný. Většinou jde o organizaci různých nefyzických pojmů ve fyzickém prostoru, jako ŠTASTNÝ JE NAHOŘE a SMUTNÝ JE DOLE (věty 26 a 27), nebo PŘEDVÍDATELNÉ BUDOUCÍ UDÁLOSTI JSOU NAHOŘE (A VPŘEDU) (věta 28) a mnoho dalších.

(26) Nálada se mi dnes *zvedla*. (Lakoff & Johnson, 2014)

(27) V poslední době jsem úplně *na dně*. (Lakoff & Johnson, 2014)

(28) Co všechno je ještě letos *před* námi? (Lakoff & Johnson, 2014)

I tyto metafory založené na fyzickém ustavení lidského těla a okolního světa se mohou někdy mezikulturně odlišovat, některé kultury např. vidí budoucnost vzadu (Lakoff & Johnson, 2014, 26).¹⁸

Později Lakoff (1992) přidává druh metafor, kterým říká metafory *struktury události*¹⁹, a organizují naše myšlení o abstraktních činnostech a dějích pomocí pojmů z oblasti pohybu

¹⁸ Dodnes je vlastně doložena pouze jedna taková kultura, jde o mluvčí jazyka Aymara, kteří žijí v jihoamerických Andách, článek o jejich pojetí času vyšel v Cognitive Science (Núñez & Sweetser, 2006).

¹⁹ event structure metaphors

v prostoru. Některé metafory struktury události jsou v angličtině extrémně frekventované, například věta (29), reprezentující metaforu RŮZNÉ ZPŮSOBY DOSAŽENÍ VÝSLEDKU JSOU RŮZNÉ CESTY, ale v češtině a jiných jazycích se jich nepoužívá a český mluvčí by ve stejné situaci použil doslovné věty (30). Existují ale i takové, které v češtině ekvivalenty mají, jako věta (31), kterou Lakoff ilustruje metaforu OPATRNÉ JEDNÁNÍ JE OPATRNÝ POHYB.

(29) Do it this way. (Lakoff, 1992)

(30) Udělej to takhle.

(31) Pohybuješ se na tenkém ledě. (Lakoff, 1992)

Zvláštní pozornost si zaslouží ještě jeden druh konceptuální metafory, tzv. *potrubní* metafora. Ještě před vydáním *Metafor, kterými žijeme* ji rozpracoval Michael J. Reddy ve slavném eseji *The Conduit Metaphor* (1979).

Potrubní metafora strukturuje náš metajazyk, tedy to, jak mluvíme o jazyce. Právě už Reddy vyslovuje hypotézu, že tato metafora určuje to, jak o jazyce přemýšlíme. Funguje takto: MYŠLENKY JSOU OBJEKTY -> JAZYKOVÉ VÝRAZY JSOU NÁDOBY NA VÝZNAM -> KOMUNIKACE JE POSÍLÁNÍ MYŠLENEK V NÁDOBÁCH. Lidé vkládají své myšlenky do nádob a posílají je potrubní poštou - jazykem k dalším lidem, kteří je z nádob vytahují. Problémy v komunikaci pak musí nutně být problémy v tomto procesu.

Samo posílání je bezproblémové, nesnáze však mohou nastat u mluvčího, který nedokáže správně vložit myšlenky do slov (věty 32 a 33), nebo nádoba, kterou posílá, nejde otevřít (34). Chyby na straně posluchače jsou také možné, ale podle Reddyho pozorování je potrubní metafora daleko náchylnější k tomu, aby vinila z neúspěšné komunikace mluvčího (Reddy, 1979, 289).

(32) To jsou jenom *prázdná* slova. (Reddy, 1979)

(33) Pokus se *vtěsnat* více myšlenek *do* méně slov. (Lakoff & Johnson, 2014)

(34) Celá ta myšlenka je *zasuta* v nesmírně hutných a komplikovaných formulacích.
(Lakoff & Johnson, 2014)

Reddyho esej se nevěnuje pouze deskripci metajazyka angličtiny, zaujímá hodnotící stanovisko. To, že běžný mluvčí angličtiny přemýšlí o komunikaci jako o potrubní poště, je v podstatě druh kognitivního biasu.²⁰ Potrubní metafora zkresluje některé aspekty jazykové komunikace. Kdyby se slova chovala jako nádoby na význam, byl by význam prostě v nich a nijak by nezávisel na kontextu dané věty nebo na konkrétním posluchači, to sice může v některých případech platit, v jiných případech tomu však tak být nemusí.

“Moštové místo” z věty (35) neodkazuje k žádné konvenční entitě a věta proto bez kontextu nemá žádný smysl.

(35) Sedněte si, prosím, na to moštové místo (Lakoff & Johnson, 2014)

V konkrétní situaci, kdy byla použita, je její význam jasný. Jeden host, který v domě přenocoval, přišel dolů na snídani. Tam bylo prostřeno pro čtyři osoby, tři místa byla s pomerančovou šťávou a jedno s jablečným moštem. Místu pak označení zůstalo, i když se už dalšího dne žádný mošt nepodával (Lakoff & Johnson, 2014, 24).

Podle Reddyho dat (vzorkem je však pouze soubor výpovědí, které se autorovi podařilo vymyslet či nasbírat) je nejméně 70% metajazykových výpovědí v angličtině řízeno potrubní metaforou. Jde tedy o dominantní formu, jakési základní nastavení mluvčích. Aby mohl potrubní metaforu zkoumat a přitom nebyl ovlivněn jejím zkreslením, navrhuje Reddy alternativní, kontrastní rámec popisu komunikace, který nazývá *nástrojářovo paradigma*²¹ (Reddy, 1979, 292). Jde o přesnější ale mnohem náročnější a pomalejší systém, takže i když je uživatelům jazyka dostupný, potřeby mluvčích zvýhodňují potrubní metaforu před jinými způsoby popisu a ta nakonec vždy převládne.

Reddy, jehož práce je do velké míry ovlivněna Kuhnovou hypotézou o revolucích ve vědě (Kuhn, 1962) a Sapir-Whorfovou hypotézou o lingvistické relativitě (např. Hoier, 1954; Hill

²⁰ Termín, který se vžil i v českém prostředí, někdy se používá také pojem kognitivní zkreslení. Jde o systematické chyby v lidském myšlení, situace, kdy se lidé nevědomky nechovají jako racionální agenti a často tak jednájí v neprospěch sebe i ostatních. Známý je například confirmation bias, availability bias nebo Dunning-Kruger effect.

²¹ *Toolmaker's paradigm*

& Mannheim, 1992), se domnívá, že potrubní metafora může ovlivňovat způsoby řešení společenských problémů, vysvětluje tak některé údajně degenerativní tendence ve fungování kultury a společnosti (Reddy, 1979, 285).

Joseph Grady (1997, 1998) později rozvíjí a reviduje celou teorii tím, že zavádí pojem *primární* metafory. To jsou jednoduchá mapování, se silným zkušenostním základem (např. LOGICKÁ ORGANIZACE JE FYZICKÁ STRUKTURA). Z kombinací primárních metafor jsou pak složeny *komplexní* metafory (např. TEORIE JSOU BUDOVY).

Vidíme tedy, že nalezených konceptuálních metafor, někdy vzájemně propojených a různě se překrývajících, je obrovské množství. Pokusem o shrnutí a větší systematizaci dosavadních evidovaných mapování je *Master Metaphor List* (Lakoff, Espenson & Schwartz, 1991)²². Jde o jejich seznam s připojenými příklady, který je organizován ve formě ontologie, ovšem spíše jen vnějškově. Ve skutečnosti není strukturován podle jasného principu. Často obsahuje stejné pojmy v různých taxonomických rovinách, prvky taxonomie se vzájemně nevylučují a rozdílné názvy tříd mohou vést ke stejným třídám (Lönneker-Rodman, 2008; Lönneker-Rodman & Narayanan, 2008). Seznam ovšem ani nebyl tvořen s přihlédnutím k požadavkům počítačnické lingvistiky, nebo výpočetní techniky, a nelze jej proto za tyto nedostatky kritizovat.

Celá teorie konceptuální metafory bývá často kritizována za nedostatek empirické evidence a dokonce nedostatek zájmu o ni, a také nerespektování vývoje kognitivních věd (Shalizi, 2003), nebo přeceňování významu metafory na úkor jiných fenoménů (Pinker, 2008; Sperber & Wilson, 2008).

A je také třeba říci, že veškerou evidenci v *Metaforách, kterými žijeme* tvoří množství pečlivě vybraných příkladů, které sice dobře ilustrují popisovanou teorii a jsou touto teorií dobře vysvětleny, ale v žádném případě neposkytují reprezentativní vzorek metafor skutečně používaných v komunikaci. Není tedy zatím jasné, jak dobře celá teorie vysvětluje realitu.

Přes všechnu svou neuspořádanost ale Lakoffova teorie zároveň vnáší do myšlení o metafoře řád, a to jak teoreticky, tím že umožňuje dávat fenomény na úrovni jazyka do

²² Jde již o druhé vydání, 1. vyšlo roku 1989.

souvislosti s fenomény z psychologie a kognitivních věd, tak prakticky, tím, že se v jejím rámci dají velmi obtížně vyčíslitelné, jednotlivé metaforické entity řadit do omezeného počtu tříd. Teorie také inspirovala mnoho nových výzkumů a projektů v oblasti NLP např. (Martin, 1990; Narayanan, 1997; Shutova, 2010; Bollegala & Shutova, 2013; Shutova & Sun, 2013).

Teorie strukturálního mapování

Základem *teorie strukturálního mapování* (také *srovnávací teorie*) D. Gentnerové (1982; 1983; 1986; Gentner & Jeziorski, 1993) je tvrzení, že analogie je mapování *znalostí* z jedné domény (*bázová doména*) do druhé (*cílová doména*).²³ Než budeme pokračovat v popisu, je třeba si nastínit, jak jsou v této teorii znalosti pojímány. Gentnerová navrhuje na základě psychologických výzkumů formální systém, kde jsou znalosti reprezentovány jako síť *uzlů* a *predikátů*. Uzly reprezentují pojmy nebo soubory pojmů, chápané jako jednotky. Predikáty použité na uzly vyjadřují propozice o pojmech. Existují predikáty dvojího druhu: *atributy*, které mají jeden argument (ŽLUTÝ (x)) a *relace*, které mají dva nebo více (KOLIDUJÍ (x,y)). Pokud jsou jejich argumenty *objekty* (oba předchozí příklady), jedná se o predikáty prvního řádu, pokud je alespoň jedním argumentem predikát, jde o predikáty druhého řádu, atd. Zásadní jsou *relace* mezi objekty v rámci jedné domény. Aby se jednalo o analogii, musí tato relace korespondovat relaci v jiné doméně, vždy jeden objekt v bázové doméně jednomu v cílové. Objekty samotné si přitom nemusí být vůbec podobné (nejsou zohledňovány jejich *atributy*), důležitá je strukturální shoda. Tvorba konkrétního mapování je pak založena na principu *systematicity*, to znamená, že je preferován systém relací spojených relací vyššího řádu (například kauzální vztah), oproti stejnému počtu nezávislých shod relací. Fungování principu *systematicity*, stejně jako zohledňování relací na úkor atributů při interpretaci analogií a metafor, má opět základ v psychologickém výzkumu a bylo experimentálně testováno – lidé např. metaforu “Stonky rostlin jsou brčka” interpretují na základě systémů relací (obojí dopravuje tekutiny pro vyživování živého

²³ knowledge, base domain/target domain

organismu) a ne na základě podobnosti vlastností objektů (obojí je dlouhé a tenké) (Gentner a kol., 2001).

Analogie je tedy srovnání základě podobnosti relací, pro pochopení jejího vztahu k metafoře a dalším druhům srovnávání, použijme následujících příkladů.

(29) Planetární soustava K5 je jako naše sluneční soustava. (Gentner, 1983)

(30) Atom je jako sluneční soustava. (Gentner, 1983)

(31) Moje práce je vězení (Gentner a kol., 2001)

(32) Slunce je pomeranč. (Gentner, 1986)

Věta (29) je *doslovné srovnání*, tedy srovnání na základě shody velké části atributů i relací. Příklad (30) je *analogie*, shoda je nalezena v relacích na základě pravidel popsaných výše, atributy objektů (velikost, chemické složení) jsou ignorovány. (31) je *metafora relačního typu* - funguje na principu analogie a může být zkoumána teorií strukturálního mapování. Věta (32) je příklad spojení, které Gentnerová nazývá *pouhý výskyt*²⁴, shoda je nalezena v několika attributech a v žádných nebo v malém počtu relací, v tomto konkrétním případě se jedná metaforu, a to *atributivního typu*. Ta nemůže být popisovanou teorií zkoumána.

Metafora je tedy v tomto pojetí širší pojem, nadřazený analogii, který obsahuje některé případy analogií a některé případy pouhých výskytů. Dále pojem metafora sdružuje propozice, které mapují objekty v doménách složitějším způsobem než je korespondence jednoho s jedním, nebo výroky, ve kterých je smíšena metafora a metonymie. Tyto případy jsou také mimo dosah teorie strukturálního mapování.

Relační metafory reprezentují značnou část všech používaných metaforických výroků a kvůli jejich jednodušší struktuře se dá očekávat, že budou mít vyšší relativní frekvenci v nebásnických textech. Gentnerové teorie analogie, již od základu s vysokým stupněm formalizace, může být proto dobrým podkladem automatických systémů pro zpracování metafory.

²⁴ mere-appearance

Na jejím základě byl již v roce 1989 vytvořen simulátor analogického myšlení *Structure-Mapping Engine* (SME) (Falkenhainer, Forbus & Gentner, 1989). Ten byl, i přes některé negativní ohlasy, kritizující příliš mnoho ručně kódovaných dat na vstupu, a tedy příliš mnoho lidské namísto umělé inteligence (Chalmers, French, & Hofstadter, 1992), široce využíván v aplikacích spojených s psychologickým výzkumem. Poměrně nedávno byla vytvořena také jeho upravená verze, která za vstup používá korpusová data (Turney, 2008).

Možnosti aplikace teoretických přístupů

Nastínili jsme si zde čtyři vlivné teorie metafory, z nichž každá může sloužit za základ komputačnímu přístupu k jejímu zpracování. Každá z nich má také určitá specifika, pro která je k tomu méně vhodná.

Teorie konceptuální metafory nabízí způsob, jak jednotlivé metaforické výroky řadit do kategorií a tím, že bere metaforu jako základní pro naše myšlení nám také dává o důvod víc, proč ji zkoumat. Nenabízí však vůbec žádné vodítko k tomu jak metaforu identifikovat v textu a je velice vzdálená formálním přístupům.

Interakční teorie je jako filozofický přístup hůře aplikovatelná a její význam je spíše historický, kdy jako první počítá s metaforou jako důležitým fenoménem, který nám umožňuje vyjádřit některé myšlenky, jež by byly doslovným jazykem nedosažitelné (Black, 1955, 293). Například Veale a Hao (2008) však využívají ve svém systému pro interpretaci metafory Blackovu myšlenku o vyzdvihování a potlačování některých složek významu při metaforickém užití.

Wilksův přístup k metafoře jako porušení selekčních preferencí je zase náchylný k nalézání falešně pozitivních výsledků a zároveň není z principu schopen objevit určité druhy metafor, sám o sobě tedy není schopen dosáhnout dobrých výsledků.

Největším problémem teorie strukturálního mapování je, že je úžeji profilovaná a zabývá se metaforou jen v rámci analogie.

V praxi se proto většinou setkáme s nějakým druhem kombinace některých z těchto přístupů, nejčastěji CMT a porušení selekčních preferencí. Druhá zmíněná bývá také často modifikována a implementována na různých místech procesu zpracování metafory, např. v systému E. Shutové (2010), který parafrázuje metaforické výroky, jsou místo porušení naopak vyhledávány silné selekční preference, které značí větší míru doslovnosti a což pomáhá při rozhodovacím procesu výběru nevhodnější parafráze.

Dalším rysem, který determinuje budoucí design aplikací jsou možnosti v oblasti zdrojů znalostí jako jsou ontologie, lexikální databáze či korpusy, dostupné pro daný druh aplikace a zkoumaný jazyk.

Doslovnost a obraznost

Pokud chceme vytvářet systémy pro zpracování metafory, jednoho ze způsobů obrazného (či nedoslovného) vyjadřování, je třeba definovat si také doslovný význam. Tento problém se, stejně jako celá problematika metafory, dostal více do popředí po vydání *metafor, kterými žijeme* Lakoffa a Johnsona. Většina přístupů předtím explicitně nebo implicitně předpokládá to, co Gibbs (1984) nazývá *hypotézou doslovných významů*. Tedy předpoklad, že:

“Věty mají jasně definovaný doslovný význam a zpracování tohoto významu je nutným krokem k porozumění výpovědi.” (tamtéž, 275)²⁵

Gibbs dále shrnuje doslovnost v klasických teoriích významu, jak byla definována u G. Fregeho, či u členů *Vídeňského kruhu*. Význam vět je plně determinován významy všech slov, které věta obsahuje zkombinovanými na základě syntaktických pravidel. Některé věty mohou být víceznačné nebo naopak nemusí mít význam vůbec, pokud jsou defektně zformovány. Z použití věty mluvčím jde pak jasně zjistit doslovný význam, od kterého se

²⁵ V originálu: “[S]entences have well-defined literal meanings and the computation of the literal meaning is a necessary step on the path to understanding speakers’ utterances.”

aktuální použití může různými způsoby lišit - metaforičností, ironií, může se jednat o idiom, atd. (tamtéž, 276).

Z modifikované verze této klasické teorie u J. Searla (1979) pak derivuje interpretaci nedoslovné věty jako třífázový proces: 1) Posлуhač zpracuje doslovný výraz věty; 2) Na základě znalosti kontextu rozhodne, zda je tento význam defektní; a 3) pokud je doslovný význam nekorektní, použije posluchač své znalosti mluvních aktů k získání významu, zamýšleného mluvčím (Gibbs, 1984, 280). Gibbs následně experimentálními psychologickými metodami dokazuje, že takovým způsobem v lidské mysli chápání nedoslovných výroků neprobíhá. To samozřejmě nemusí být nutně překážkou pro přístupy, které chtějí automaticky detekovat doslovné a nedoslovné užití slov, prakticky to však překážkou bude minimálně pro určitý stupeň přesnosti těchto systémů. A naopak následování výsledků psychologických a kognitivně vědních výzkumů může být obrovskou výhodou, protože bude způsobovat progres ve všech zainteresovaných disciplínách. Takový postup nemusí být vždy možný, ale je jistě žádoucí.²⁶

Hypotéza jasně definovatelného doslovného významu je tedy těžko udržitelná a není zřejmě možné určit jasnou demarkační čáru. Podobně je tomu u dělení metafor na tzv. „mrtvé“ a „živé“, kdy živými byly v tradičním pojetí metafor myšleny imaginativní básnické obraty, zatímco často používané metafor o svůj „život“ přišly vinou sémantického posunu. Lakoff a Johnson však těmto konvenčním metaforám svou teorií život navracejí, když za jejich použitím nacházejí mapování mezi doménami. Jak upozorňuje Lakoff (1987) v kategorii mrtvých metafor se tak ocitnou velice rozdílné entity. Jednak bude obsahovat pojmy, které svou metaforičnost ztratily a dosvědčuje ji pouze jejich etymologie („pedigree“, výraz z francouzštiny, původně znamenající „noha jeřába“, metaforičnost odkazovala k tvaru podobnému grafickému vyjadřování rodokmenů) a jednak v ní budou stále používané projevy konceptuálních mapování (jako je „grasp“ ve

²⁶ V prostředí počítačové lingvistiky a umělé inteligence mluví např. již J. H. Martin (1990), v souladu s Gibbsem, o tzv. *prvotní reprezentaci*, která je v systému MIDAS fází zpracování metafor, kdy je přítomna pouze lexikální a syntaktická informace, jako o přechodné fázi, která je pro jeho systém nutným krokem, ale u které by se vůbec nemělo mluvit o významu (Fass, 1991).

smyslu pochopit). Lakoff rozlišuje celkem čtyři rozdílné kategorie výroků, které by se podle klasického chápání shodně octly pod označením mrtvá metafora.²⁷

Vhodnější tedy bude, jak navrhuje už Gibbs, dívat se na metaforičnost jako na kontinuum, jdoucí od nových, kreativních užití až k vysoce konvencionalizovaným, u kterých si ani většina rodilých mluvčích není vědoma jejich metaforičnosti. Toto kontinuum na ose metaforičnost/doslovnost se navíc samozřejmě mění v čase.

Tato neurčitost je dobře patrná na korpusových výzkumech, kde probíhá anotace metafor lidskými hodnotiteli. Příkladem může být studie Anny Barbaczy a kol. (2010)²⁸, která zahrnovala anotaci korpusu o velikosti přibližně 30 000 slov na přítomnost lingvistických a konceptuálních metafor. I přes fakt, že dva anotátoři postupovali podle jasných instrukcí (založených na metodě MIP (Praglejazz Group, 2007)), při prvním pokusu byla naměřena shoda pouze 17%. Při druhém pokusu a po upravení instrukcí bylo dosaženo 48%. To jsou překvapivě nízké hodnoty.

Na druhou stranu, v současné době nejrozsáhlejší a vysoce respektovaná korpusová studie výskytu metafor (Steen a kol., 2010), která také používá metodu MIP, dosáhla relativně velmi vysoké shody mezi anotátory 85%. To je o mnoho více než u Barbaczy a jejích kolegů, pravděpodobně vlivem kombinace pečlivějšího výcviku hodnotitelů ve zpracování úlohy a menší složitosti samotné úlohy, kdy byly anotovány pouze lingvistické metafora bez konceptuálních mapování.

Shoda 85% však není až tak vysoké číslo, pokud vezmeme v úvahu, že jde pravděpodobně o výsledek blízký tomu nejlepšímu, co dokážeme na současné úrovni metodologie výzkumu, lidských možností, a časových a finančních prostředků získat. Přesto v 15% případů selháváme. U mezních případů jednotliví hodnotitelé sice dokážou vybrat jednu z možností, ale při srovnání s dalšími často vyjde najevo, že jejich hodnocení je nekonzistentní a jde o ekvivalent hádání. Také to ukazuje, že bychom neměli mít přehnané nároky na počítačnické přístupy, zabývající se zpracováním obrazného vyjadřování.

²⁷ Andrew Goatly (1997) podobně užívá dělení na pět druhů metafor s metaforickými názvy: aktivní, unavené, spící, pohřbené a mrtvé (active, tired, sleeping, buried a dead).

²⁸ Jejím primárním cílem byla evaluace tří metod automatické detekce konceptuální metafor.

To vše neznamená, že bychom se museli pojmu doslovnosti zcela vzdát, jen není vhodné, slovy P. Hankse “na něj klást příliš velkou teoretickou váhu.” (2006, 19).

Většinovým přístupem v NLP je určení hranice mezi doslovným a metaforickým podle potřeby dané aplikace. Shutová (2015) navrhuje několik obecných zásad.

System pravděpodobně nebude potřebovat jako metaforické určovat použití “etymologicky metaforických” výrazů, jako je výše zmíněné slovo „pedigree“, které se používají pouze v jednom kontextu, také obecně vysoce konvencionalizované metafory, u kterých se nemetaforický význam používá relativně mnohem řídčeji a/nebo je jejich metaforický význam zaveden ve slovnících, takže může být zpracován existujícími systémy pro desambiguaci slovního významu (WSD).²⁹ Zbylé množství konvenčních metafor naopak bude zřejmě potřeba strojově identifikovat a interpretovat, aby s nimi mohlo být dále korektně nakládáno a totéž platí pro nové, nezavedené metafory.

Takový teoreticky neukotvený ad-hoc postup však ztěžuje evaluaci výsledků, jak vzhledem k předmětu zkoumání, protože bude snižovat už tak problémovou shodu lidských hodnotitelů, tak (a to především) oproti ostatním systémům, zpracovávajícím stejnou úlohu.

Problém je tedy v tom, že pokud si chceme ponechat binární rozdělení na metaforičnost/doslovnost, nejsme si jisti, kde udělat demarkační čáru mezi nimi. Zároveň, pokud řekneme, že jde o kontinuum, kdy doslovnost a metaforičnost jsou krajními body na škále, půjde o poměrně vágní prohlášení, které nám nijak nepomůže s řešením problému. Objevily se však přístupy, které se pokoušejí míru doslovnosti/metaforičnosti exaktně měřit. Těmi se budeme zabývat v následující kapitole.

Stupně metaforičnosti

Existují nejméně dva přístupy pro měření metaforičnosti. Prvním z nich pojetí lexikálního významu u lexikografa a korpusového lingvisty Patricka Hankse, dalším je metodika měření metaforičnosti, kterou navrhl Jonathan Dunn.

²⁹ Word sense disambiguation

Přístup Patricka Hankse

V Hanksově lexikální sémantice nemají samostatná slova význam, ale pouze neurčitý *potenciál* významu. Pomocí korpusové analýzy lze však identifikovat *schémata* jejich použití a význam přiřadit až k jednotlivým schématům. Tato metoda se nazývá *analýza korpusových schémat*³⁰ neboli CPA (Hanks, 2004a; Hanks & Pustejovsky, 2005).

Doslovný význam slova tedy může být maximálně pomocným termínem značícím tu část potenciálu významu slova, která je aktivována v jeho nejvíce *normálním* kontextu, přičemž to, do jaké míry je kontext normální, je měřitelné.

V (2006) Hanks tvrdí, že konvenční metafory, kterých je podle něj naprostá většina, se dají považovat za sekundárními významy. Neplatí však naopak, že by všechny sekundární významy byly metaforami, těch je jen malá část.

Ne všechna slova mají také schopnost být užita metaforicky, u abstraktních pojmů jako “kvantita“, je to zřejmě nemožné. Hanks rozlišuje šest charakteristik, které ovlivňují možnost metaforičnosti, patří mezi ně např. míra odbornosti (vysoce odborné výrazy nebývají příliš často používány metaforicky), frekvence (příliš vysoká absolutní frekvence výskytu vede k větší pravděpodobnosti doslovnosti) nebo sémantická třída (fyzické lokace, jako “moře“, “poušť“ nebo „džungle“, jsou častým zdrojem metaforických použití).

Míra metaforičnosti výroku pak závisí na míře shody vlastností zapojených domén, Hanks používá Blackovy termíny primární a sekundární subjekt. Čím méně společných vlastností tím větší stupeň metaforičnosti a tím větší nutnost soustředění tvůrce i interpretátora výpovědi, aby byla interpretace relevantní.

Hanks nabízí dvě případové studie na Britském národním korpusu, na kterých demonstruje různou míru metaforičnosti. Vyhledává např. výskyty “sea of [NP]“, které pak řadí podle stupně metaforičnosti na základě výše zmíněné zásady. Takže “sea of faces“ je metaforičtější než “sea of blood“, protože krev je stejně jako mořská voda (a na rozdíl od

³⁰ corpus pattern analysis; O metodě ještě bude obsírněji pojednáno v kapitole 3.4.3

tváří) tekutou a jednolitou substancí. Metodou Hanksova zkoumání je však zřejmě pouze jeho vlastní intuice aplikovaná na korpusová data.

Hlavními zjištěními jeho práce jsou: Metaforičnost roste s tím, jak se snižuje počet podobných vlastností domén. Některá schémata jsou na metafory mnohem plodnější než jiná. Čím je abstraktnější sekundární subjekt (cílová doména) metafory, tím větší je tendence k jejímu dalšímu rozvíjení v kontextu, tedy k vytvoření rozšířené metafory.

Metoda Jonathana Dunna

Dunn (2011; 2014) nabízí, aniž by odkazoval na Hanksovu práci, vlastní metodiku měření metaforičnosti, která je mnohem transparentnější. Jeho přístup předpokládá existenci metaforických mapování podle CMT a možnost jejich poznání, a využívá koncept tématických rolí (původně Fillmore, 1967), které určují vztahy mezi událostí a jednotlivými jejími participanty. Dunn navrhuje dvě hlediska, z kombinace jejichž hodnot vychází konečná míra metaforičnosti výroku, prvním je *hustota* metafory a druhým její *síla*.³¹

Hustotou metafory je míněna míra toho nakolik se ve formě lingvistické metafory projeví konceptuální metafora, na které je první založena. Například konceptuální metafora SPOR/ARGUMENTACE JE VÁLKA je ve větě (33) reprezentována jedním konstituentem zatímco ve (34) dvěma a ve (35) třemi konstituenty. Příklad (33) má tedy nejmenší a příklad (35) největší hustotu. Dunn používá výčet a definice tématických rolí konstituentů od Nirenburga a Raskina (2004), ty zahrnují: *agens, příjemce, cíl, prožívající, prostředek, místo, cesta, účel, původ a téma*, k nimž přidává ještě *událost*³².

(33) *Marry rozmetala Johnovy argumenty* – [událost] (Dunn, 2011)

(34) *Marry rozmetala Johnovy nejsilnější zbraně* – [událost, téma] (Dunn, 2011)

(35) *Nepřítel obklíčil Johnovu pevnost* – [agens, událost, téma] (Dunn, 2011)

³¹ density; strenght; Síla je v podstatě ekvivalentem vlastnosti metafory, kterou Tourangeau a Sternberg (1981) a další autoři nazývají *vhodnost* (aptness), avšak oproštěná od estetického podtextu (Dunn, 2011).

³² V totožném pořadí: agent, beneficiary, destination, experiencer, instrument, location, path, purpose, source, theme, event

Příklad (35) je navíc případem vyjádření, které Dunn nazývá *nasyčené*, to znamená, že má dostatečnou hustotu, aby jej bylo možné číst ve dvou významech, jako výrok o SPORU/ARGUMENTACI nebo jako výrok o VÁLCE, to u (33) a (34) nebylo možné, protože pouze v některých přítomných tématických rolích byl přítomen metaforický obsah.

Dunn pracovně počítá, že jednotlivé tématické role mají stejný vliv a tedy stejnou hodnotu. Potenciálním problémem je možnost různého počtu lexikálních jednotek v jednotlivých tématických rolích, který by mohl také modifikovat hodnotu vlivu daného konstituentu, Dunn však dochází ke zjištění, že tento vliv je zanedbatelný. Hustota metaforického výroku se tedy rovná počtu metaforických tématických rolí. Výsledná hodnota má pochopitelně nějaký význam až při srovnání s údajem o jiném výroku.

Druhá proměnná, tedy síla metafory, je určena jako míra metaforičnosti konceptuální metafory, na které je daný výrok založen. Ta je součtem dvou veličin: *vzdálenosti* a *směru*.³³

Vzdálenost definována jako míra ontologické odlišnosti zdrojové a cílové domény. Dunn opět používá systém Nirenburga a Raskina (2004), tentokrát pro ontologické rozdělení domén na: *duševní*, *sociální*, *nehmotné* a *fyzické*.³⁴ Přičemž s rozdíly počítá jen mezi dvěma možnými kombinacemi, pro stejný druh domény (36) a pro metaforu, která propojuje domény rozdílných druhů (37). První s menší vzdáleností má hodnotu 1, druhá má větší vzdálenost s hodnotou 2.

(36) Tenhle vztah je zlý sen [DUŠEVNÍ JE DUŠEVNÍ] (Dunn, 2011)

(37) Je na vrcholu své kariéry [SOCIÁLNÍ JE FYZICKÉ] (Lakoff & Johnson, 2014)

Druhou složkou síly metafory je směr. Ten určují pojmy použité v metaforách. Pojem z kteréhokoli výše zmíněného druhu domény může být buď *událostí* nebo *objektem*, toto rozdělení nazývá *funkce*. Pokud se metaforické mapování odehrává mezi událostmi nebo

³³ mental, social, intangible, physical

³⁴ distance, direction

objekty, má směr hodnotu 0, pokud propojuje událost a objekt, má hodnotu 1. Celý vzorec tedy vypadá takto:

(38) metaforičnost = síla x hustota

(39) metaforičnost = (vzdálenost + směr) x hustota

(40) metaforičnost = (počet druhů domén + neshodná funkce) x (počet rolí)

A aplikován na příklady:

(41) Orel je *lev* mezi ptáky (Dunn, 2011); [FYZICKÉ JE FYZICKÉ]; [objekt – objekt]; [téma]

(42) metaforičnost (41) = (1 doména + 0 neshodná funkce) x (1 role) = 1

(43) Jeho teorie je *vystavěna na pevných základech*; [DUŠEVNÍ JE FYZICKÉ]; [objekt – objekt]; [událost, téma]

(44) metaforičnost (43) = (2 domény + 0 neshodná funkce) x (2 role) = 4

Tento přístup nám tedy umožňuje metaforičnost měřit, nedává nám jasnou odpověď na to, jaké jsou nutné a dostačující podmínky pro označení výroku za metaforický (Dunn si například není jistý, zda výroky, které mají hodnotu metaforičnosti 1 označovat za metafory). Je také jasné, že kterákoli z proměnných či kterýkoli z operátorů ve vzorci, pokud by byly nastaveny chybně, by do velké míry ovlivnili výsledek, takže se snadno může stát, že celý přístup produkuje nesmyslné výsledky.

Na druhou stranu nám ukazuje některé nanejvýš zajímavé možnosti studia metafory. Protože je výsledná metaforičnost souběhem více prvků, lze některé z nich držet konstantní a sledovat, jak se mění pouze ty zkoumané, např. pokud budeme chtít zkoumat konceptuální mapování, budeme to provádět na vzorku s konstantní funkcí a počtem rolí. Toto rozdělení metaforičnosti na více proměnných, které vytvářejí hodnotu na škále, je založeno na intuicích mluvčích, kteří chápou některé výroky více a některé méně metaforicky, podle Dunna je proto žádoucí, aby takto metaforu “chápaly” i počítačové přístupy.

Možnosti aplikace metod měření metaforičnosti

I když je Hanksova CPA sama o sobě velice slibnou metodou pro budoucí přístupy k analýze metafor (viz dále v této práci), jeho přístup k měření metaforičnosti se v popsané podobě nedá aplikovat do NLP systémů.

Dunnův přístup zohledňuje, oproti Hanksovu, navíc míru proniknutí konceptuální metaforu na lingvistickou úroveň. Není však v principu kvalitativně odlišný ve smyslu objektivity. Hodnoty proměnných v jeho vzorci (např. zda budou mít různé tématické role různý koeficient metaforičnosti) totiž lze určit opět pouze na základě intuice. Lze však použít intuice více mluvčích a počítačový systém, pokud má přístup k jednoduchým ontologickým informacím, jak jsou popsány výše, je schopen s vzorcem velmi jednoduše pracovat.

Dunn svou metodu v pozměněné podobě aplikoval v (2014), kdy modifikoval svůj předchozí systém pro identifikaci metafor (2013a). Nový systém vykazuje, oproti původní verzi s binárním rozdělením na metaforické a nmetaforické výroky, mírně horší, avšak srovnatelné výsledky. Jeho výhodou však je, že nabízí možnost nastavit práh metaforičnosti na libovolnou úroveň podle potřeb aktuální úlohy a vzorku.

Otázkou je, jak hodnotit metaforu, které zasahují do větších částí diskurzu než jsou jednotlivé věty (rozšířené metaforu), a zda metoda v současné podobě nebude produkovat zkreslené výsledky, pokud na takovou metaforu narazí.

Všudypřítomná metafora

Metafora je v jazyce rozšířeným fenoménem. Dokládá to zejména množství korpusových studií na angličtině. Podle nerozsáhlejší nekomputační studie metafor (Steen a kol. 2010) je až 18,5% slov v Britském Národním Korpusu použito metaforicky. Podle dalších velkých studií (Cameron 2003; Martin 2006; Shutova & Teufel 2010) obsahuje metaforu přibližně každá třetí věta.

Co se týče ostatních jazyků, je evidence podstatně skromnější. V rámci CMT bylo provedeno několik výzkumů, zkoumající univerzalitu jednotlivých konceptuálních metafor, např. Alverson (1994) prokázal existenci konceptuálního mapování ČAS JE PROSTOR v angličtině, mandarínštině, hindštině a jazyku sesotho. To samo o sobě však neříká nic o frekvenci této metafor v jednotlivých jazycích. Navíc i když je prostorové vnímání času možná skutečnou jazykovou univerzálií, konkrétní lingvistické projevy tohoto mapování se liší.

Takové rozdíly byly zjištěny mezi angličtinou a řečtinou. Mluvčí angličtiny (podobně jako český mluvčí) běžně používá slovní spojení jako “dlouhá doba”, zatímco v řečtině jsou podstatně frekventovanější trojrozměrné metafor typu “hodně času” (Casasanto, 2013)³⁵. Dalším příkladem může být již zmíněný jazyk Aymara, jehož mluvčí vnímají budoucnost za sebou a minulost před sebou (Núñez & Sweetser, 2006).

Anatol Stefanowitsch (2004) použil metodu *analýzy metaforických schémat*³⁶, aby srovnal konceptuální metafor týkající se štěstí na korpusech angličtiny a němčiny. Zjistil mezi danými jazyky některé kulturně podmíněné rozdíly v užití metafor, a také některé metodologické komplikace podobných srovnávacích studií, plynoucí z toho, že ani zdánlivě ekvivalentní pojmy nemusí mít v různých jazycích zcela stejný význam (je například rozdíl v intenzitě emoce u anglického *happiness* a německého *Glück*).

Několik korpusových studií mezijazykového používání konceptuální metafor nabízí Silva Bratož (2004; 2012; 2013), která srovnává užívání metafor ve slovinštině a angličtině v některých oblastech diskurzu. Zjišťuje například, že v žurnalistických textech o volbách v roce 2008 se používají v obou jazycích podobné metafor, avšak jejich frekvence je v (amerických) anglických textech mnohem vyšší než ve slovinských. Podobné rozdíly naopak nepozoruje v diskurzu ochutnávání vína.

Zdá se tedy pravděpodobné, že metafora je všudypřítomným fenoménem jazyka a větší rozdíly jsou spíše otázkou stavu výzkumu jednotlivých jazyků či oblastí.

³⁵ Jde o český překlad J. Jehličky - původní článek (Casasanto, 2008)

³⁶ *metaphorical patterns analysis* (Stefanowitsch, 2006).

Analýza metafor v rámci NLP

Proč se bychom se měli pokoušet o automatické zpracování metafor

Doménově neomezená korpusová studie frekvence metaforických výpovědí v jiném než anglickém jazyce sice zatím chybí, z údajů, které máme se však zdá nepravděpodobné, že by v češtině nebo jiných jazycích byla metafora pouze okrajovým jevem. A právě její vysoká frekvence činí ze strojové detekce a interpretace metafor důležitou úlohu, která může o mnoho zpřesnit zpracování přirozeného jazyka v mnoha oblastech. Jde především o:

- **Strojový překlad:** Významná část často používaných metaforických výpovědí je specifických pro jednotlivé jazyky, či kulturní okruhy a při doslovném překladu se proto stávají těžce srozumitelnými nebo zcela nesmyslnými. Jako příklad lze uvést metaforický výrok v češtině “[P]ozorně *hltá* každé slovo.” (SYN2010), který v současné době nejužívanější automatický překladač Google Translator³⁷ přeloží do angličtiny jako nesmyslnou větu “Devouring every word carefully.”. Zároveň je ovšem třeba podotknout, že mnoho metaforických výpovědí je ekvivalentních mezi různými jazyky, a jako takové je strojový překladač, i bez nutnosti jakéhokoliv systému k práci s metaforou, přeloží korektně. Aby tedy byla NLP aplikace zpracovávající metaforu v oblasti strojového překladu v praxi použitelná, je třeba, aby byla její průměrná úspěšnost interpretace vyvážila procentuální zastoupení ekvivalentních metaforických výroků mezi danými jazyky. Tato komplikace se však stává tím méně závažnou, čím více překládaných jazyků daný program kombinuje, protože průnik používaných metaforických vyjádření se pak pochopitelně značně snižuje.
- **Opinion mining/Sentiment analysis:** Tedy extrakce postoje mluvčího k tématu, ke kterému se vyjadřuje, také patří mezi oblasti NLP, které by mohly těžit z implementace systému k interpretaci metafor. Metaforické výpovědi mají často

³⁷ Volně dostupný na <http://translate.google.com>.

silnou emocionální složku a jejich správné strojové vyhodnocení tak může přesnost daných systémů výrazně zlepšit. Například výsledky workshopu evaluací počítačnických systémů pro analýzu sémantiky SemEval 2015 ukazují, že při úloze postojové analýzy obrazného vyjadřování v příspěvcích na Twitteru (vzorek byl ručně anotován na sarkasmus, ironii, metaforu a ostatní způsoby) měly všechny zúčastněné systémy relativně nejnižší úspěšnost právě u metafory, a také, že míra úspěšnosti zpracování metafory nejvíce korelovala s celkovým skóre v testech (Ghosh, Li, Veale, Rosso, Shutova, Reyes & Barnden, 2015). Systémem, který je schopen postoj autorů metafor odhadovat je dokonce už program KARMA z roku 1997 (Narayan, 1997; 1999).

- **Information retrieval:** Protože mnoho nedoslovných výroků může vést k nesprávným tématickým okruhům. Například na dotaz “škola” se chybně vrátí texty o “gentlemanech *staré školy*” (Korkontzelos a kol., 2013, cit. ze Shutova, 2015)

Je také snadné si představit, že z počítačového “rozumění” metafoře by mohla těžit jakákoliv aplikace, zahrnující komunikaci s lidmi, přinejmenším s těmi, kteří jsou laiky v oblasti výpočetní techniky.

Kromě praktických aplikací jsou počítačové přístupy k metafoře důležité pro testování hypotéz. Introspektivně získané příklady totiž nelze brát jako dostačující evidenci pro teoretické závěry. Získávání dat pouze introspekci nebo introspektivně motivovaným ručním vyhledáváním v databázích je totiž jednak podstatně méně produktivní, jednak není relevantním způsobem určování důležitosti nalezených informací (ve smyslu centra a periferie, ani žádném jiném). Různými kognitivními zkresleními jsou navíc ovlivněni nejen autoři dané teorie, ale i lidé, kteří se s těmito teoriemi seznamují, hodnotí je - a potenciálně budou vytvářet nové hypotézy.

Kupříkladu Gentnerová a kol. (2001) upozorňují, že intuitivní chápání a ztotožnění se s konceptuálními metaforami, společně mnoha lidem, by nemělo být evidencí pro přijetí Lakoffovy teorie. Výzkumy totiž ukazují, že lidé jsou náchylní k tomu, adoptovat za vlastní interpretace nedoslovných výroků, které jsou jim podsunuty.

Keysar a Blyová (1995) to ukazují na experimentu s interpretací idiomů. Dali subjektům přečíst archaické anglické idiomy v různých kontextech. Lidé, kteří četli spojení “the goose hangs high” v příběhu se smutným koncem, chápali jeho význam opačně, než ti, kteří ho četli ve veselém příběhu, tzn. “vypadá to špatně/dobře”. Lidé z obou skupin byli zároveň přesvědčeni, že by jeho význam chápali stejně bez ohledu na kontext (Gentner a kol., 2001).

Získání “tvrdých” dat je dobrým způsobem, jak se vyvarovat co nejvíce omylů v našem uvažování o nedoslovném jazyce.

Proč je úloha složitá

Ať už jde tedy o výzkum nebo praktické NLP aplikace, je zvládnutí automatické detekce a interpretace metafory žádoucím cílem. Zároveň jde o úlohu, která vyžaduje obrovské množství kategorizovaných znalostí o světě. Její složitost je patrná i z opakovaně replikovaných výzkumů vývojových psychologů, ty již od 60. let minulého století dokazují, že děti jsou schopny korektně interpretovat metaforu až od cca 7-10 let věku, do té doby chápou výroky doslovně, např. metaforický pojem “*tvrdý* člověk” pro ně označuje člověka, který je svalnatý a tedy skutečně fyzicky tvrdý (Asch & Nerlove 1960; Gentner, 1977; Vosniadou, 1986; Geary, 2012).³⁸

Právě tyto rozsáhlé soubory vědomostí o fyzikální a kulturní realitě, které počítače postrádají a naopak lidé přirozeně a bez většího usilí získávají postupně během života (někdy se užívá termín subkognice (French, 1990)), stejně jako jsou již dopředu určeny naší evoluční historií, jsou důvodem proč je porozumění metafoře, a nedoslovnému jazyku obecně, tak náročnou úlohou v počítačnické lingvistice.

Do nedávné doby byla tak většina NLP výzkumů zaměřených na obraznost v jazyce založena z velké části na manuálně sbíraných a kódovaných datech, nebo se věnovala

³⁸ Jiná situace nastává u tvoření metaforických výpovědí, to je u dětí pozorováno téměř od chvíle, kdy začínají mluvit, podle mnoha autorů však nejde o skutečnou metaforičnost, neboť děti ještě nemají ustaveny konvenční kategorie, jejichž hranice by následně mohla metafora porušovat a jde tedy pouze o pseudometaforu, která vnějškově jako metafora působí (Vosniadou, 1986).

výzkumu daného fenoménu na ručně předpřipraveném nebo tématicky omezeném vzorku. Až v posledních letech se objevily také statistické systémy pro zpracování metafory. Ty mají oproti prvně jmenovaným několik výhod. Jsou z principu univerzálnější co se týče tématických okruhů i jednotlivých jazyků. Jejich vývoj je také z důvodu menšího zapojení úloh řešených přímo lidmi méně časově náročný. V neposlední řadě, jelikož jsou schopny okamžitě se učit z právě vzniklých textů, mohou daleko lépe držet krok s vývojem jazyka a kategorizace nových entit.

Způsoby počítačového zpracování metafory

Strojové zpracování metafory vlastně zahrnuje dvě zvláštní úlohy, *automatické rozpoznávání (identifikaci)* metafory, tedy detekci metaforických použití jazyka v textu a *automatickou interpretaci* metafory, tedy zjištění zamýšleného významu výroku.

Přitom u interpretace je navíc otázkou, jakým způsobem tuto informaci zprostředkovat dále. Jednou z možností je “překlad” rozpoznané metaforické části textu do doslovné formy, tedy interpretace metafory jako tvorba parafráze (např. Shutova, 2010). To může být vhodný způsob, pro systémy, které potřebují zpracovávat úlohy na základě “rozumění” vstupu v přirozeném jazyce.

Dalším způsobem je interpretace jako automatické vysvětlení (explanace) metafory, kdy program odhaluje vlastnosti použitých pojmů, na které má metafora poukazovat (to je např. přístup u Veale a Hao (2008) a Ovchinikové a kol. (2014)).

Systémy pro automatickou detekci se zase liší v úrovni, na které pracují. Některé systémy identifikují lingvistickou metaforu na úrovni jednotlivých nedoslovně užitých slov, nebo takové, které označí pouze větu, ve které je obsažena metafora (Krishnakumaran & Zhu, 2007; Dunn, 2013).

Přístupy se liší také v tom, se kterými slovními druhy pracují. Nejčastěji detekují metaforu reprezentovanou slovesem (např. Gedigian a kol., 2006; Birke & Sarkar (2006)), ale některé dokáží rozpoznat i metaforické adjektivum či substantivum (Krishnakumaran & Zhu, 2007)

Dále existují takové systémy, které se snaží identifikovat konceptuální metafory, na kterých jsou jejich lingvistické projevy založeny (Mason, 2004). Shutová & Sunová (2013) představily algoritmus identifikující jak metaforické výroky, tak konceptuální metafory. Některé programy provádějí jak detekci, tak interpretaci metafory (Fass, 1991; Shutova, Teufel & Korhonen, 2012), ale prozatím vždy v rámci dvou propojených systémů.

Zdroje informací pro počítačové zpracování metafory

Jak bylo řečeno výše, zpracování metafory je fenomén, který je extrémně náročný na znalosti o fungování vnějšího světa. Existuje několik způsobů, jak takové informace zprostředkovat počítačným systémům.³⁹ Můžeme je pracovním rozdělit na dvě kategorie: *Znalostní báze* obsahují potřebné znalosti přímo a byly pro jejich reprezentaci umělé inteligenci navrženy. *Korpusy* je obsahují druhotně a jejich primárním účelem je něco jiného.

Znalostní báze

Starší systémy pro zpracování metafory většinou pracují s vlastní znalostní bází ve formě ontologie. To je systém reprezentace znalostí, obsahující definice, vlastnosti a vzájemné vztahy entit (často z nějaké konkrétní oblasti). Počítačový systém tyto informace pak může využít, aby neprodukoval výsledky, které jsou v rozporu s realitou.

Problém je, že metaforické výroky vlastně v rozporu s realitou jsou (čas neběží a TEORIE NEJSOU BUDOVOY), a samotný fakt, že daný výraz porušuje pravidla dané ontologie zase není dostatečné kritérium pro jeho označení za metaforu. Některé systémy tedy používají různé způsoby reprezentace znalostí na různých úrovních procesu zpracování metafory (např. Fass, 1991). Manuální kódování těchto reprezentací je však velmi náročné. V současné době se od tvorby úzce zaměřených, ručně kódovaných zdrojů upouští.

³⁹ Reprezentace znalostí (knowledge representation; KR) je celou samostatnou disciplínou umělé inteligence.

Některé přístupy ke zpracování metafory však používají již existujících ontologií, například Dunn (2013) využívá SUMO ontologii (Niles & Pease, 2001), z níž extrahuje vlastnosti pojmů, jež pak považuje za základní významy.

Ontologickým způsobem je uspořádán i Lakoffův Master Metaphor List (Lakoff, Espenson & Schwartz, 1991), který již byl popsán v kapitole o CMT. MML je sám o sobě nevhodný pro využití v počítačové lingvistice, byl však některými autory pro tento účel modifikován (Krishnakumaran a Zhu, 2007). Často je používán k evaluaci výsledků NLP přístupů ke konceptuální metafoře (např. Mason, 2004). Inspiroval také další projekty, které se pokoušejí o tvorbu zdroje metafor ve formě seznamu mapování mezi doménami.

Jde především o MetaBank (Martin, 1994), znalostní bázi konvenčních metafor angličtiny, která obsahuje empiricky získané informace o mezidoménových mapováních, a Mental Metaphor Databank, jež je databází vytvořenou pro projekt ATT-Meta (Barnden & Lee, 2002). Ta obsahuje příklady metafor z oblasti mentálních stavů, které jsou řazeny k jednotlivým konceptuálním mapováním z MML.

Konceptuální mapování z MML využívá také Hamburg Metaphor Database (HDM; Lönneker, 2004; Reining & Lönneker-Rodman, 2007). To je databáze příkladů metaforických výroků v němčině a francouzštině (z toho francouzština převládá). Kromě MML je také jejich obsah propojen s významy v EuroWordNetu (Vossen, 1998).

WordNet

Mnoho systémů pro zpracování metafory používá lexikální databázi angličtiny WordNet (Fellbaum, 1998), její struktura byla vytvořena na základě experimentálně získaných psycholingvistických poznatků, momentálně nejaktuálnější verzí je 3.1.

WordNet seskupuje slova do řad synonym (tzv. *synsetů*). Každý význam slova reprezentuje samostatný synset⁴⁰ a každý z nich je doplněn lexikografickou definicí, či glosou. Jednotlivé synsety jsou spolu propojeny sémantickými vztahy. Databáze zahrnuje

⁴⁰ Z definice zde platí jeden synset = jeden význam ve WordNetu (*sense*), to však nemusí souhlasit s významy slov, jak jsou většinou chápány.

podstatná jména, slovesa, přídavná jména a příslovce. Typy sémantických vztahů jsou různé pro různé slovní druhy. Jde např. o hyperonymum/hyponymum, tj. vztah nadřazeného a podřazeného pojmu (hyponymum patří do kategorie, kterou reprezentuje hyperonymum), holonymum/meronymum (vztah části a celku), nebo vyplývání (sloveso předpokládá jiné sloveso). Některé ze synsetů zachycují metaforické významy např. “grasp” má ve WordNetu také význam “get the meaning of something”. Metaforické významy však nejsou zachycovány systematicky a velké množství jich schází.

Na základě původního princetonského WordNetu vzniklo množství databází pro další jazyky (např. projekty EuroWordNet (Vossen, 1998) a BalkaNet (Tufiş, Cristea & Stamou, 2004)) včetně češtiny. Jejich vývoj se však postupně zastavil a data z nich nejsou volně přístupná.

Český WordNet vznikl od roku 1998 v Centru zpracování přirozeného jazyka na Fakultě informatiky Masarykovy univerzity v Brně. Jeho upravená verze (1.9 PDT; Pala a kol., 2010) byla použita pro lexikálně-sémantické anotace Pražského závislostního korpusu (PDT). Synsety Českého WordNetu jsou propojeny s princetonským WordNetem ve verzi 2.0.

Využití WordNetu v automatickém zpracování metafor jsou různá. Pro některé systémy jde o základní zdroj dat, např. systém Krishnakumarana a Zhu (2007), který kontroluje vztah hyperonymie/hyponymie u spojení ve tvaru “substantivum1 IS A substantivum2”, a pokud ho nenalezne označí výrok za metaforu. Pro jiné funguje jako pomocná komponenta, jako u Shutové (2010), která ve svém systému pro parafrázování metaforických výroků používá WordNet pro vyfiltrování parafrází, které jsou antonymní k hledanému významu. Antonyma se totiž vyskytují ve stejných kontextech a byla by tak systémem označena jako možné parafráze.

Korpusy

Jak již bylo řečeno, korpusy explicitně neobsahují znalosti nutné k porozumění pojmům, ve smyslu jejich definic a vzájemných vztahů. Obsahují však slova a věty v mnoha reálných kontextech a jsou opatřeny anotacemi, takže jsou hodnotným zdrojem pro statistické

strojové učení. Existuje celá řada korpusů s různými způsoby anotace, v současné době jsou již dostupné dva korpusy ručně anotované na přítomnost metafory. Ty jsou, kromě využití při vývoji aplikací, vhodné k evaluaci výsledků systémů pro analýzu metafory.

Jedná se o *VU Amsterdam Metaphor Corpus* Steena a kol., (2010) a práci Shutové a Teufelové (2010). Oba přístupy vycházejí ze stejného anotačního schématu, tzv. *metaphor identification procedure* (MIP), vytvořeného skupinou Pragglejaz Group (2007). MIP identifikuje metaforu na lexikální úrovni a anotátoři při identifikaci postupují podle čtyř bodového návodu. Anotátoři určují, zda má každá lexikální jednotka základnější význam, než ten, ve kterém je právě užito, který je zároveň: konkrétnější (lépe představitelný a vázaný na smysly); spojený s tělesnými pochody; přesnější (méně vágní); historicky starší. Taková lexikální jednotka je označena za metaforickou.

VU Amsterdam Metaphor Corpus byl anotován metodou MIPVU, která oproti MIP nezohledňuje historii lexikálních jednotek. Obsahuje 200 000 slov z korpusu BNC Baby ze čtyř žánrů: akademické texty, novinové články, fikce a konverzace. Steenův výzkum označuje za metaforu i velmi konvencionalizované mezní případy, proto může být méně vhodným zdrojem pro NLP aplikace, které potřebují být zaměřeny spíše na méně konvenční metafory, které nejsou lexikalizované (Shutova, 2011). Výsledný údaj, že 18,5% lexikálních jednotek v korpusu je metaforických, tak může být nadsazený ve smyslu, jak metaforičnosti chápe většina počítačových přístupů (Dunn, 2014).

Práce Shutové a Teufelové (2010) je založena na upravené verzi MIP a je zaměřena na anotaci nejen lingvistické, ale také konceptuální metafory. Korpus je podmnožinou BNC, obsahuje 13 642 slov a je v něm označeno 241 metaforických vyjádření. Pro anotaci konceptuální metafory měli anotátoři k dispozici seznam běžných zdrojových a cílových domén, a bylo jim také povoleno vymýšlet vlastní. Korpus byl využit při testování některých počítačových přístupů k metafoře vyvíjených v rámci Computer Laboratory na univerzitě v Cambridge (např. Shutova a kol., 2010; Shutova, Teufel & Korhonen, 2012).

Využití nových poznatků lexikálně-sémantických teorií

Směrem, kam by se mohl ubírat další vývoj NLP aplikací, zabývajících se nejkompexnějšími problémy sémantiky, včetně analýzy metafory, je využití poznatků sémantických teorií Patricka Hanksa, které shrnul do *teorie norem a zneužití* (TNE; Hanks, 2013).⁴¹

Hanksův pohled na lexikální význam propojuje Pustejovského teorii *generativního lexikonu* (Pustejovsky, 1991; 1995; Pustejovsky & Boguraev, 1993) s preferenční sémantikou Yoricka Wilkse (Wilks, 1975; 1978; viz metafora jako *porušení selekčních preferencí*) a Fillmorovu sémantiku rámců (Fillmore, 1975).

Hanksovou výzkumnou metodou je tzv. *analýzu korpusových schémat*⁴² neboli CPA (Hanks, 2004; Hanks & Pustejovsky, 2005). Předpokladem jeho přístupu je, podobně jako u Fillmora i Pustejovského, že samotná slova nemají žádný význam, ten je určen až jejich použitím v kontextu. Běžné metody řešení úlohy desambiguace slovního významu (WSD) založené na enumerativních seznamech známých významů tudíž nemohou popsat produktivní jevy jako je metafora, ironie a další druhy nedoslovného vyjadřování.

Slova sama o sobě mají pouze potenciály významu, tvořené z neurčené množiny sémantických komponent. Při použití slova mohou být všechny nebo jakákoliv jejich část aktivovány. Teorie se však na této úrovni vůbec nesnaží význam definovat. Doslovný význam může být maximálně pomocným termínem značícím tu část potenciálu významu slova, která je aktivována v jeho nejvíce *normálním* kontextu (Hanks, 2006, 19). CPA je založena právě na hypotéze normálního kontextu a na možnosti měření jeho normality, k tomuto účelu používá statistické metody (např. test *bodové vzájemné informace*⁴³ (Church & Hanks, 1989)).

Jednotkami významu jsou namísto slov frazeologická *schémata* extrahovaná z korpusu, tím se rapidně snižuje míra polysémie. Schémata pak mohou být *normou*, pokud jsou užitá v běžně očekávaném kontextu, *zneužitím*, pokud se od normálního kontextu úmyslně odchyľují, nebo se jedná o chybu, pokud je odchylka neúmyslná. Jedním z typů zneužití je právě metafora.

⁴¹ Theory of Norms and Exploitations

⁴² Corpus pattern analysis

⁴³ Pointwise mutual information; PMI

Výsledkem CPA metody je také slovník slovesných schémat *Pattern Dictionary of English Verbs* (PDEV)⁴⁴, vyvíjený na Masarykově Univerzitě v Brně. Význam každého schématu je vyjádřen sadou implikatur⁴⁵ a všechny argumenty jednotlivých schémat vedou k uzlu v ontologii (Pustejovsky a kol, 2004).

Tvorba PDEV, tedy identifikace a popisy schémat, probíhá manuálně samotným Hanksem a několika dalšími lingvisty, kteří prošli speciálním výcvikem, a jde tedy o extrémně časově náročný projekt, který je prozatím daleko od dokončení. Bylo však již provedeno několik výzkumů aplikovatelnosti PDEV v NLP (např. Cinková a kol, 2010 nebo Maarouf a kol, 2014) se slibnými výsledky.

Výhledem do budoucna je nahrazení úlohy desambiguace slovního významu desambiguací schémat (*pattern disambiguation*; PD), která je díky podstatně menší míře polysémie jednodušší, a zároveň je schopna obsáhnout i obrazná a jinak nezvyklá užití jazyka.

Dosavadní systémy pro automatické zpracování metafor

V této kapitole bude pojednáno o některých dosavadních aplikacích na poli automatického zpracování metafor od počátku 90. let až do současnosti a způsobech evaluace jejich výsledků. Zabývám se těmi systémy, které byly reálně implementovány v počítačových programech. Pro získání povědomí o existujících systémech jsem použil přehledové články na toto téma (Martin, 1996; Loenneker-Rodman & Narayanan, 2008; Shutova, 2010b; Shutova 2015), které - pochopitelně - vždy popisují situaci do doby svého vydání. U většiny systémů, o kterých pojednávám, jsem se snažil popsat metodologii a architekturu jednotlivých přístupů o něco podrobněji než v právě zmíněných zdrojích. Také popisují jeden systém pro analýzu a produkci přirovnání. Jednotlivé přístupy jsou rozděleny do

⁴⁴ Dostupný na <http://pdev.org.uk/>

⁴⁵ Původní termín P. Grice (1975) je u Hankse použit v mírně odlišném smyslu, kdy není podmíněn porušením principu kooperace. Příklad: pro sloveso "file" (podat) existuje schéma [[Human = Plaintiff]] file [[Procedure = Lawsuit]], které může vyjadřovat implikaturu: Pokud podáte žalobu, jste žalobcem (plaintiff) a vyvoláváte proceduru, pomocí které vám bude, jak doufáte, nahrazena škoda, o které jste přesvědčen, že na vás byla spáchána.

podkapitol podle použitých metod, zároveň však řazení těchto kategorií téměř kopíruje vývoj způsobů zpracování dané úlohy v čase.

Základní metody evaluace výsledků

Pro určení úspěšnosti systémů pro identifikaci metafor se používá především dvou měřítek. *Precision* (P) je míra správně určených výsledků ze všech určených jako správné a *recall* (R) je míra správně určených výsledků ze všech které jsou skutečně správné (Powers, 2011). Je také možná probabilistická interpretace, kdy *precision* je pravděpodobnost, že náhodně vybraný výsledek bude správný a *recall* reprezentuje pravděpodobnost, že náhodně vybraný správný výsledek bude nalezen.

Architektura NLP aplikací je často přizpůsobena tomu, které z uvedených měřítek je v daném případě důležitější. Zda větší pokrytí vzorku, které je však vykoupeno větším počtem falešně pozitivních výsledků, nebo naopak snaha se nalézání falešně pozitivních výsledků vyhnout za cenu extrakce menšího počtu případů.

Metriky se také liší v požadavcích na testovaný vzorek. Zatímco pro měření *precision* teoreticky stačí zjišťovat správnost u výsledků určených systémem jako správné, pro *recall* je třeba mít dostatečně spolehlivě anotovaný celý vzorek. To u identifikace metafor a dalších složitých sémantických úloh v současnosti stále znamená mít k dispozici manuálně anotovaný zlatý standard (*gold standard*).

Měřítkem, které kombinuje *precision* a *recall*, je *F-measure* (někdy také *F-score*). Je definováno jako (45).

$$(45) F=2PR/P+R$$

Pochopitelně existuje mnoho dalších evaluačních metrik, tři zmíněná měřítka jsou však u systémů pro zpracování metafor klíčová a nejčastěji uváděná.

První systémy z počátku 90. let

Prvním automatickým systémem zpracovávajícím metaforu, který byl implementován ve funkčním programu, je *MIDAS* (Metaphor Interpretation, Denotation and Acquisition System) Jamese Martina (1988; 1990). *MIDAS* byl součástí programu *Unix Consultant*, automatického systému, který odpovídal na dotazy uživatelů o operačním systému Unix. Pokud *Consultant* není schopen najít odpověď na dotaz v doslovném významu, zavolá program *MIDAS*. Ten na základě porušení selekčních preferencí identifikuje metaforické výrazy a prohledá databázi pro vhodnou konceptuální metaforu, která vysvětluje danou anomálii.

Základem Martinova přístupu je tvrzení, že specifičtější metafory vychází z obecnějších. Program tedy používá hierarchicky uspořádanou databázi konvenčních metafor. Když je mu dán výrok, vyhledává k němu korespondující metaforu. Pokud takovou nenalezne, abstrahuje použité pojmy na obecnější úroveň a opakuje vyhledávání. Když je nalezena vhodná metafora, program vytvoří na bázi daného výroku nové mapování. Tímto způsobem systém získává nové metafory.

Systém posloužil pro tvorbu znalostní báze⁴⁶ konvenčních metafor angličtiny *MetaBank* (Martin, 1994).

Dan Fass (1991) vytvořil systém pro detekci nedoslovných výroků jménem *met**, který byl implementován v programu *meta5*. Systém dokáže rozlišit mezi doslovnou výpovědí, metaforou, metonymií a anomáliemi v krátkých větách v angličtině. Postup má tři fáze, nejdříve program na základě porušení selekčních preferencí rozliší doslovnost a nedoslovnost. V případě, že je fráze vyhodnocena jako obrazná, je za pomoci manuálně vytvořené databáze metonymických schémat (např. ČÁST JAKO CELEK, NÁDOBA JAKO OBSAH NÁDOBY, VÝROBCE JAKO VÝROBEK a další) testováno, zda jde o metonymii. Pokud není nalezena, systém vyhledá ve znalostní bázi *relevantní analogii*.

(46) My car drink gasoline.⁴⁷

⁴⁶ Knowledge base

Preferovaným typem slovesa *drink* je *animal*, slovo *car* není jeho hyponymem, proto jsou porušeny selekční preference, zároveň je nalezena relevantní analogie (*thing, use, energy_source*) tedy taková, která obsahuje hyperonymum jak preferovaného, tak původního agenta - jedná se o metaforu.

Výrok je tedy interpretován jako metafora, protože nesplňuje podmínku dodržení selekčních preferencí, není vyhodnocen jako metonymie a je nalezena relevantní analogie.

Pokud není splněna žádná z těchto podmínek výrok je vyhodnocen jako *anomálie*.

Podle Fasse kombinuje jeho systém rozpoznávání metafory poznatky všech čtyř hlavních teoretických přístupů k metafoře, kdy kromě již zmíněné identifikace nedoslovnosti porušením selekčních preferencí, pracuje s metaforou jako s jevem založeným na analogii stejně jako to dělají interakční teorie a srovnávací přístup D. Gentnerové. Jednotlivé relevantní analogie pak tvoří seskupení podobná konceptuálním metaforám (Fass, 1991).

Systemy založené na reasoningu ve zdrojové doméně

System *KARMA* (Knowledge-based Action Representations for Metaphor and Aspect; Narayanan, 1997) byl vytvořen v rámci *neurální teorie jazyka* (NLT), interdisciplinárního projektu Kalifornské univerzity v Berkeley, který poznatky základního výzkumu lingvistiky, kognitivní vědy, neurovědy, biologie a umělé inteligence aplikuje do NLP (např. Feldman, 2006; Bergen & Feldman, 2006).

KARMA je tématicky zaměřená na metafory struktury události v krátkých textech z mezinárodní ekonomiky a politiky, které interpretuje. Program pracuje s textovým vstupem, který prošel parsingem.

(47) Británie dělala malé krůčky (...) (Narayanan, 1997)

⁴⁷ Již jednou použitý (věta (9)) příklad (Wilks, 1978; Fass, 1991) jsem tentokrát ponechal v angličtině, aby byla patrná návaznost na prvky ve znalostní bázi Fassova systému.

System je založen na znalostní bázi, obsahující převážně údaje ze zdrojové domény, tedy pohybu v prostoru, přičemž přímo využívá výzkumů v oblasti neurálního základu provádění a vnímání pohybu v mozku primátů (pracuje s tzv. X-schématy⁴⁸). Za použití informací o zdrojové doméně program vyhodnotí výrok a výsledky na základě konceptuálních mappingů aplikuje na doménu cílovou. System převodu znalostí mezi doménami přitom obsahuje i informace o rysech či proměnných, které mají zůstat beze změny, jako je např. slovesný vid, program tedy neudělá chybu v tom, že by v interpretaci zaměnil ukončený a probíhající proces (Loenneker-Rodman & Narayanan, 2008).

Program ukazuje počátky schopnosti automatických systémů vyhodnotit autorův postoj k dějům a entitám, které se objevují v daném textu. Zatímco věta (48) pouze konstatuje kroky, které vláda učinila, věta (49) implikuje, že její autor považuje vládní regulaci trhu za nesprávný ekonomický postup.

(48) Indická vláda deregulovala trh. (Narayanan, 1999)

(49) Indická vláda uvolnila trhu smyčku. (Narayanan, 1999)

KARMA je tuto distinkci schopna rozlišovat. Cílová doména je totiž reprezentována pomocí *sítě přesvědčení*, což je metoda, která využívá teorii pravděpodobnosti (Bayesův teorém) k vyhodnocování údajů z nejistých zdrojů (Narayanan, 1999).

Podobný přístup k automatické interpretaci metafory jako Narayanan zvolili John Barnden a Mark Lee (2001), výsledkem jejich práce byl systém ATT-Meta. Ten opět používá propracovanou manuálně vytvořenou znalostní bázi pro zdrojovou doménu, o něco méně komplexní znalosti o cílové doméně a systém pro přesun informací ze zdroje k cíli, založený na teorii konceptuální metafory.

ATT-Meta analyzuje metaforické a metonymické popisy mentálních stavů. Program nedokáže přímo zpracovávat texty v přirozeném jazyce, jako vstup používá ručně kódované logické výroky reprezentující výpovědi nebo jejich fragmenty. Metoda ATT-Meta je založena na teoretickém předpokladu, že uživatel jazyka nepojímá metaforické výroky

⁴⁸ X-schemas nebo executing schemas (Loenneker-Rodman & Narayanan, 2008)

konstrukcí mnoha mapování mezi zdrojovou a cílovou doménou, ale spíše uvažováním na úrovni zdrojové domény. Systém tedy pracuje s minimem mezidoménových mapování a aplikuje komplexní rozhodovací procesy, které řeší rozpory vznikající přenesením informace mezi jednotlivými doménami a dovolují, aby měla při konstrukci interpretace kterákoli z nich prioritu.

Výhodou systému ATT-Meta je, že dokáže zpracovávat jak sériově tak paralelně smíšené metafory a také metaforické výroky v jiném než oznamovacím tvaru. Autoři přitom kritizují dosavadní nezáměr o teoretické ukotvení i automatickou analýzu obou jevů (Barnden a kol., 2002).

Automatická identifikace konceptuální metafory

Systém *CorMet* Zacharyho Masona (2004) je prvním automatickým systémem pro detekci metaforických mapování, který nepoužívá žádné speciálně pro něj ručně vybudované zdroje znalostí. Program vytěžuje z webu velké tématicky specifické korpusy a využívá lexikální databázi WordNet. Věta (50) ukazuje metaforické použití slovesa “lít” z domény LABORATOŘ v doméně EKONOMIE reprezentující programem nalezenou konceptuální metaforu PENÍZE JSOU TEKUTINA.

(50) [N]elze do něj donekonečna *lít* peníze daňových poplatníků. (SYN2010)⁴⁹

CorMet získává texty z vyhledávače *Google*, kam posílá dva druhy dotazů. První sestávají z konjunkce dvou až pěti slov náhodně vybraných z klíčových slov vztahujících se k nějakému tématu (např. EKONOMIE: *finance, investice, akcie* a další), ty slouží k vyhledání *jakýchkoli* dokumentů k danému tématickému okruhu. Dotazy druhého typu jsou použity k

⁴⁹ Větu jsem vybral z ČNK jako ekvivalent příkladu “Funds poured into his bank account.” (Mason, 2004). Metaforické použití slovesa “lít” ve smyslu “investovat peníze” se v češtině vyskytuje, je však pravděpodobné, že se používá podstatně méně než v anglicky mluvícím prostředí. V korpusu SYN2010 obsahuje v kontextu 10 tokenů napravo lemma substantiva “peníze/finance” pouze 16 výskytů lemmatu slovesa “lít” z celkových 1510. Peníze však v češtině velice často “tečou”, takže konceptuální mapování z LABORATOŘE do EKONOMIE je zde aktivní.

vyhledávání dokumentů k danému tématickému okruhu, které *obsahují určité sloveso*. Jsou složeny z konjunkce klíčových slov a disjunkce jednotlivých tvarů hledaného slovesa, které mohou nejpravděpodobněji být pouze slovesem (např. anglické *Attack* je vyhledáváno jako *Attacked OR Attacking* a není použit tvar *Attacks*, který je většinou substantivem).

Program pak ze získaných textů vybere pro jednotlivá témata několik kořenů sloves s nejvyšší relativní frekvencí. Na těchto charakteristických zástupcích se systém učí jejich selekční preference. Pokud nalezne slovesa, která preferují pojmy ze dvou tématických okruhů, určí jejich polaritu, tzn. navrhne, která doména bude zdrojová a která cílová. Např. Slovesa "lít", "téct" nebo "zmrazit" preferují v doméně LABORATOŘ pojmy ze třídy *tekutiny* a v EKONOMII pojmy ze třídy *aktiva*. Naopak slovesa "utratit" nebo "investovat", která v doméně EKONOMIE preferují *aktiva*, nemají v LABORATOŘI konkrétní selekční preference. LABORATOŘ je tedy zdroj a EKONOMIE cíl.

Systém prokázal při evaluaci proti Master Metaphor Listu 77% přesnost ve smyslu míry správně určených mapování.

Identifikace metafory založená na významech ve WordNetu

Zatímco předchozí systémy (Mason, 2004; Gedigian a kol., 2006 a Birke & Sarkar (2006)) se zaměřují pouze na metaforu vyjádřenou slovesem, Krishnakumaran a Zhu (2007) představují přístup, který dokáže detekovat i výroky, kde je metaforicky použito substantivum nebo adjektivum. Pracovně rozdělují metaforické věty na tři typy: Typ I obsahuje tvar slovesa "to be", nejčastěji ve formě "subjekt IS-A objekt" (věta (51)), u typu II je metaforicky použito sloveso, které není tvarem "to be" (52), větě typu III je použito metaforicky adjektivum (53).

(51) He *is a* brave lion. (Krishnakumaran & Zhu, 2007)

(52) He *planted* good ideas in their minds. (Krishnakumaran & Zhu, 2007)

(53) He has a *fertile* imagination. (Krishnakumaran & Zhu, 2007)

Autoři používají dvě rozdílné metody identifikace metafory zvláště pro typ I a zvláště pro typy II a III. Pro “IS-A” metafory prostě vyhledají, zda mají použítá substantiva v hierarchii WordNetu vztah hyperonyma a hyponyma, pokud není nalezen, věta je označena jako metaforická.

Pro II. a III. typ metafory navrhli Krishnakumaran a Zhu dvojrozměrnou matici, jejímiž hodnotami jsou frekvence spoluvýskytu párů *sloveso-substantivum* (tzn. typ II) a *adjektivum-substantivum* (typ III). Autoři z důvodu nedosažitelnosti korpusu zpracovaného parsingem (který by byl vhodnější), používají ke konstrukci matice bi-gramy slov z *Web IT* korpusu. Ten tvoří n-gramy (n je 1-5) anglických slov a obsahuje asi bilion tokenů extrahovaných z veřejných webových stránek. Tento přístup vede k velkým ztrátám relevantních dat, např. spojení typu “plant an idea” jím nebudou detekována. Autoři se domnívají, že tento nedostatek může být vyvážen rozsáhlostí použitého korpusu.

K evaluaci výsledků svého přístupu Krishnakumaran a Zhu použili Master Metaphor List. Ten ručně anotovali na 789 “živých” metafor a zbytek (asi 1000 vět), které považují za konvenční do té míry, že již nezapadají do jejich definice metaforického užití. Tyto “mrtvé” metafory použili při testování jako vzorek doslovných výroků. Algoritmus pro typ I dosáhl výsledků P (precision) = 70% a R (recall) = 61%. Chybou ukazující limity tohoto přístupu je např. negativně určený výrok “He is a *budding* artist”, který je metaforou typu III, ale protože obsahuje tvar slovesa “to be”, je určen jako nemetaforický. Výsledky mohou být ovlivněny tím, že algoritmy nebyly testovány na přirozených textech a kvůli použití příkladů z MML ani na skutečně doslovných větách.

“Sardonicus”

Tony Veale a Yanfen Hao (2007) vytvořili program *Sardonicus* pro vyhledávání přirovnání a generování nových přirovnání a metafor. Jedná se o tzv. *case-base reasoning* (CBR) systém, tedy takový, který provádí generalizace na základě srovnávání vlastností společných trénovacím a nově získaným datům. Jejich přístup je založený na premise, že metafora je komplexnější druh přirovnání, a že tedy lze na jednoduchých případech

(explicitní přirovnání) vytrénovat program pro zpracování těch složitých (ne-explicitní metafora).

Systém vytěžuje web pro příklady přirovnání ve tvaru “as ADJEKTIVUM as a/an SUBSTANTIVUM”. U nalezených přirovnání bylo zjištěno časté užití ironie, např. “Vlasatý jako bowlingová koule”. Příklady byly proto ručně anotovány, a přibližně 13% procent přirovnání a dokonce 20% typů přirovnání bylo označeno jako ironické. Program tak disponuje dvěma zvláštními *případovými bázemi*⁵⁰, jednou pro základní (bona-fide) užití přirovnání a jednou ironickou.

Díky tomu, že získává data z webu je systém schopen zjistit charakteristické vlastnosti entit, které lexikální databáze (jako WordNet) neobsahují a zpracovat např. metaforu “Svatby jsou pohřby” na základě frekventovaného spolu-výskytu adjektiva “nákladný”. Sardonicus také generuje metafory a přirovnání na žádost uživatele, který určí substantivum a jeho vlastnost, např. “Paris Hilton” a “štíhlý” vytvoří výsledek: “Paris Hilton je tyčka”. Pokud systém zadané slovo nezná, dokáže si k němu na webu automaticky vyhledat příklady přirovnání.

Přístupy založené na clusteringu

Birkeová a Sarkar (2006) vytvořili program *TroFi* (Trope Finder), jde o systém rozpoznávající doslovné a obrazné vyjadřování za pomoci clusteringu. Používá upravený algoritmus pro desambiguaci významu Karova a Edelmana (1998) založený na podobnosti. Ten v původní verzi vyhodnocuje podobnost mezi větami, které obsahují slovo k desambiguaci a podle toho je rozděluje do jednotlivých významů. Autoři algoritmus modifikují tím způsobem, že užití vět rozděluje pouze do dvou možných významů - doslovného a nedoslovného. Pro tvorbu sady případů nedoslovného použití používají druhou databázi vět složenou z databáze známých metafor a idiomů, a algoritmus přesunu entit z doslovné do obrazné databáze.

⁵⁰ Case-bases

System dosahuje úspěšnosti 53,8% v metodice F-score a při použití volitelné komponenty systému, tzv. aktivního učení, kdy jsou některé věty vráceny k posouzení lidským anotátorem, dosahuje program F-score 64,9%.

Loenneker-Rödman a Narayanan (2008) kritizují u tohoto přístupu nedostatečné teoretické ukotvení a z toho pramenící vágnost definic doslovného a nedoslovného vyjadřování. Taková vágnost totiž může snižovat konzistenci anotace testovacích dat a tedy i přesvědčivost výsledků.

Shutová a kol. (2010) představili systém pro identifikaci metafory, založený na clusteringu substantiv a sloves. Motivací jejich přístupu bylo pozorování výsledků automatického určování sémantických tříd slov na základě shod v kontextu za pomoci clusteringu.

Většina slov, které se ocitnou v jednom clusteru, má stejný či podobný význam, avšak abstraktní koncepty zařazené podle stejných kontextů do stejného clusteru se často významově odlišují. Autorky situaci interpretují tak, že se pojmy řadí k sobě na základě asociací se stejnou zdrojovou doménou.

Jejich systém je schopen z malého souboru ručně anotovaných metafor (*seed set*) určit další metafory ze stejné cílové domény, např. “marriage” a “democracy” se v korpusu objevují ve spojení s termíny popisující “mechanism”. Program tak dokáže z ručně anotované metafory “mend marriage” odvodit, že “the functioning of democracy” je taktéž užito metaforicky. Systém zpracovává metafory ve tvaru sloveso-subjekt a sloveso-objekt.

Autorky systém testovaly oproti *baseline* systému, který odvozuje další metafory na základě synonym slov ze *seed setu*, které získá z WordNetu. Testovaným vzorkem byly všechny texty z BNC. Metafory nalezené oběma přístupy pak byly testovány na metaforičnost lidskými anotátory. Jejich systém dosáhl přesnosti (precision) 0,79, což je o 35% více než *baseline*, při signifikantně větším počtu nalezených metafor. Co se však týče pokrytí v absolutním smyslu, je pravděpodobné, že je velmi nízké (hodnota recall nebyla měřena). K většímu pokrytí metaforických výrazů by bylo třeba použít signifikantně větší a reprezentativní *seed set*.

Přístup Shutové a kolektivu je prvním statistickým přístupem k identifikaci metafory, pracující se vstupem v čistém textu.

Vyhledávání metafor na základě abstraktnosti kontextu

Turney a kol. (2011) využívají pro identifikaci metafory fakt, že ji běžně užíváme pro vyjadřování abstraktních konceptů. Srovnávají míru konkrétnosti/abstraktnosti sloves a adjektiv oproti substantivům v jejich okolí. Výrazy, ve kterých jsou konkrétní slovesa a adjektiva užita spolu s abstraktním substantivem, jsou pak označeny jako metaforické (např. "černý humor").

Algoritmus pro výpočet abstraktnosti vychází z předchozí práce hlavního autora (Turney & Littman, 2003), zabývající se automatickým určováním sémantické orientace. Srovnává sémantickou shodu (*semantic similarity*) jednotlivých slov se souborem čtyřiceti *paradigmatických* slov (20 konkrétních a 20 abstraktních), ta jsou získána automaticky. Autoři použili psycholingvistický slovník *MRC Psycholinguistic Database Machine Usable Dictionary* (Coltheart, 1981), obsahující hodnoty abstraktnosti slov na vytrénování algoritmu, za pomoci kterého pak vygenerovali slovník o velikosti 114501 slov, každé s hodnotou abstraktnosti. Z něj automaticky extrahovali paradigmatická slova.

Pro evaluaci použili dataset z programu TroFi (Birke & Sarkar (2006)) pro slovesné a vlastní manuálně anotovaný vzorek pro adjektivní metafory, systém dosáhl F-score 0,68, resp. 0,79. Autoři přitom pro určování metaforičnosti používali pouze hodnoty abstraktnosti kontextu, u testovaných adjektiv a sloves nebyla míra jejich abstraktnosti započítávána a systém tedy implicitně předpokládal, že mají všechna stejnou pravděpodobnost být užita metaforicky.

Jazykově nezávislá identifikace metafory

V několika posledních letech se objevily snahy o metody identifikace metafory, které by byly nezávislé na konkrétním jazyce. Yulia Tsvetková a kol. (2013; 2014) toho docilují za pomoci překladu. Autoři představili přístup založený na klasifikaci podle hrubých

sémantických rysů. Systém abstrahuje data jako abstraktnost, životnost a označení pojmenovaných entit. Klasifikátor logické regrese je pak trénován na datech v angličtině a vytrénovaný model je za pomoci slovníku promítnut do ruštiny (Tsvetkov a kol., 2013), resp. ještě do španělštiny a perštiny (Tsvetkov a kol., 2014).

Systém identifikuje metaforu na úrovni celých vět. Evaluace na angličtině probíhala na datasetu TroFi (Birke & Sarkar, 2006), pro ruštinu vytvořili autoři vlastní vzorek, algoritmy dosáhly poměrně vysokých čísel v metodice F-measure: $F=0,78$, resp. $0,76$, a upravený systém pak podobných výsledků i u perštiny a španělštiny.

Analýza metafor v češtině

Automatickou identifikací ani interpretací metafor v češtině se prozatím nikdo nezabývá. V této kapitole se zaměříme na možnosti zpracování těchto úloh (s důrazem hlavně na automatickou identifikaci) pro český jazyk.

Více komplikací, méně zdrojů

Tak by se dal shrnout obecný stav NLP v češtině ve srovnání s angličtinou a také dalšími velkými jazyky, jako je němčina, španělština či ruština. Čeština má složitou morfologii a je vysoce flektivní. Na překážku je také její volný slovosled. I když typ SVO (Subjekt, Verbum, Objekt) převažuje (věta 54), je v českých větách možné používat i další druhy slovosledu (55 a 56).

(54) Syn poslal matce dárek. (Bojar a kol., 2012)

(55) Dárek poslal matce syn. (Bojar a kol., 2012)

(56) Dárek poslal syn matce. (Bojar a kol., 2012)

Přičemž přesunem jednotlivých větných členů dochází ke změně množství informace v nich obsažené. V (55) je novou informací darující a v (56) obdarovaná osoba. Čeština má také poměrně značné rozdíly mezi formami jazyka užívaného v různých komunikačních situacích (např. komunikace s úřady vs. komunikace s přáteli; Bojar a kol., 2012). To vše ztěžuje počítačové zpracování češtiny již od základních úloh morfologické a syntaktické analýzy.

V českém jazyce je také dostupná menší základna zdrojů pro zpracování přirozeného jazyka. Je to logické protože vývoj těchto zdrojů je nákladný a je tedy výhodnější realizovat ho v jazycích, v nichž bude mít větší impakt. Dobře to jde vidět např. na rozsahu sémantické sítě WordNet, kterou v některé fázi procesu prozatím využívají téměř všechny

automatické identifikátory metafor. Zatímco původní princetonský WordNet obsahuje momentálně 117659 synsetů⁵¹, jeho česká verze jich má asi čtyřikrát méně.

Co se týče výzkumu metafor, je v českém prostředí realizován hlavně v rámci kognitivní lingvistiky s důrazem na polskou (lublinskou) školou kognitivní lingvistiky a propojení s fenomenologickou filozofií. Klíčové je pro tento přístup osobní prožívání člověka a pojem *jazykového obrazu světa* (Vaňková, 2002). Metodologicky jde tedy v podstatě o přesný opak počítačových přístupů.⁵²

Korpusové experimenty

Pro získání přibližného srovnání češtiny s angličtinou v oblasti frekvence metafor a možných metod její identifikace jsem se rozhodl replikovat dva velice jednoduché korpusové experimenty.

Experiment 1: Goatlyho lingvistické ukazatele metafor

Andrew Goatly (1997) definuje soubor lingvistických ukazatelů, tedy slov a frází, jež se často objevují v okolí metafor. Identifikoval více než 100 takových ukazatelů, které rozdělil do 20 kategorií (explicitní, mimetické, percepční procesy apod.). Existuje množství studií, které tyto markery používají k výzkumu textů různých žánrů a jazyků (např. Charteris-Black & Ennis, 2001; Skorczynska & Piqué-Angordans, 2005; Skorczynska & Ahrens, 2015).

Já zde použiji velmi jednoduchý design experimentu podle Shutové, Teufelové a Korhonenové (2012), ty měřily přesnost (ve smyslu *precision*) šesti ukazatelů: *metaphorically speaking, literally, utterly, completely, so to speak*. Ty byly vyhledány v BNC a 50 náhodně vybraných kolokací na každý ukazatel bylo ručně anotováno na

⁵¹ Aktuální údaj je dostupný zde: <http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>

⁵² Např. ve (Vaňková a kol., 2005, 188-190) je sice zkratka NLP použita, je jí však míněno neurolingvistické programování (*neurolinguistic programming*), což je psychoterapeutická metoda, dlouhodobě vědeckou komunitou považovaná za pseudovědu, či šarlatánství (viz např. Thyer & Pignotti, 2015).

metaforičnost. Tento design má tu výhodu, že všechny použité markery metaforického užití jsou poměrně snadno přeložitelné do češtiny, bez pocitu nepatříčnosti některého z nich a nutnosti vynalézat jeho volnější parafrázi.

Průměrně dosáhla Shutová a její kolegyně přesnosti 0,41, jinými slovy 41% z nalezeného vzorku vět obsahujících ukazatele bylo metaforických, výsledky pro konkrétní markery jsou uvedeny v tabulce 1.

Ukazatel (Goatly, 1997)	Úspěšnost (Shutova, Teufel & Korhonen, 2012)
metaphorically speaking	0,71
Literally	0,26
Figurative	0,18
Utterly	0,32
completely	0,26
so to speak	0,71
Průměr	0,41

Tab. 1: Přesnost ukazatelů metafor v BNC

Postup jsem zopakoval na korpusu SYN2010 z ČNK. Jako ukazatele v češtině jsem použil výrazy: *metaforicky řečeno*, *doslova*, *obrazně řečeno*, *úplně*, *zcela* a *tak říkajíc*. Výsledky jsou uvedeny v tabulce 2, průměrná úspěšnost byla na češtině 0,49, tedy o něco vyšší.⁵³

Ukazatel	Relativní frekvence (SYN2010)	Absolutní frekvence (SYN2010)	Vzorek	Z toho metafor	Úspěšnost
Metaforicky řečeno	0,06	7	7	6	0,86
Doslova	50,41	6133	50	23	0,46
Obrazně	1,28	156	50	38	0,76

⁵³ Metodologickým problémem je, že jde o dva odlišné anotátory, na odlišných datech, ve dvou odlišných jazycích, u nichž nelze provést test shody anotátorů.

řečeno					
Úplně	230,14	28001	50	10	0,2
Zcela	253,09	30793	50	9	0,18
Tak říkajíc	1,38	168	50	25	0,5
				Průměr	0,49

Tab. 2: Přesnost ukazatelů metafor v SYN2010

U obou jazyků funguje s velkou mírou přesnosti explicitní ukazatel *metaforicky řečeno*, který má však velice nízkou frekvenci (pouze 7 výskytů u obou korpusů), a nepříliš vysoké procento úspěšnosti mají markery *úplně* a *zcela* (*utterly* a *completely*). Největší rozdíl mezi danými jazyky byl u nalezen explicitního ukazatele *obrazně řečeno* (*figuratively*), který, při srovnatelné absolutní frekvenci (ČNK: 156 a BNC: 125), má v češtině přesnost 76%, zatímco v angličtině pouze 18%.

Základním problémem lingvistických ukazatelů je, že z principu nacházejí pouze malé procento ze všech metafor (*recall*), protože většina z nich není v textu žádným podobně zjevným způsobem označena.

Obě verze tohoto experimentu ukazují, že v omezeném vzorku metafor, které lze touto metodou detekovat se navíc nachází poměrně velké procento šumu. Metoda tedy sama o sobě pro identifikaci metafor nepostačuje a byla by použitelná pouze jako pomocná komponenta pracující v rámci komplexnějšího systému (Shutova, Teufel & Korhonen, 2012).

Experiment 2: Schémata bohatá na metafor v češtině

Jednou z hypotéz Patricka Hankse ve studii (2006) je, že některá slova, potažmo některá korpusová *schémata*, jsou na metaforu bohatší než jiná. Demonstruje to dvěma případovými studiemi na BNC. V první z nich vyhledává metaforické výskyty “a sea of

[NP]”. Provedení podobného experimentu na ČNK a srovnání výsledků nám může pomoci odhadnout, jak moc se fungování metafor v češtině a angličtině liší, a v čem se shoduje.

Hanksova pozorování se dají shrnout následovně. V BNC je:

- a) slovo “sea” frekventované a většina jeho užití není metaforických.
- b) konstrukce “a sea of [NP]” extrémně bohatá na metafory.

Pokusil jsem se zjistit platnost ekvivalentních výroků v korpusu SYN2010 a české a anglické verzi InterCorpu v8.

Ad a) Hanks uvádí, že v BNC je 11565 výskytů slova “sea” a při manuálním prozkoumávání stovek náhodně vybraných kolokací, nenarazil na jedinou metaforu (2006, 23). Kvůli jeho časové nenáročnosti jsem zvolil podobně nesystematický postup, vyhledal jsem lemma substantiva “moře” v korpusu SYN2010⁵⁴ a ručně vyhledával metaforičnost u prvních 150 náhodně promíchaných kolokací. Výskytů celkově je 15021, jde tedy přibližně o 1%.

Metaforických užití jsem našel 9 (např. “moře písku”, “moře plyšáků”, “moře chmelu”), to znamená 6% všech zkoumaných. Pokud bychom brali vážně Hanksovu evidenci, je to o 9 více než v BNC, ani z jednoho provedení experimentu však nelze vyvozovat žádné skutečné důsledky kvůli malé prozkoumané části vzorku. Kdybychom jev přesto chtěli vysvětlit, nabízí se srovnání geografické polohy České republiky a Britských ostrovů. Angličané pravděpodobně potřebují častěji mluvit o moři v doslovném významu.

Zajímavějším zjištěním je, že všech 9 nalezených metafor je realizováno typem konstrukce, kterým se zabývám v testování hypotézy b).

Ad b) Konstrukci “a sea of [NP]” jsem do češtiny převedl jako: “moře + substantivum v genitivu”.⁵⁵ V korpusu SYN2010 bylo nalezeno 670 výskytů, přítomnost metafory jsem vyhodnotil u poloviny, tj. 335, náhodně promíchaných výsledků. 29 výsledků jsem následně vyřadil, 23 protože byly chybně vyhledány (většinou nesouhlasil pád druhého substantiva), 6 jsem vyřadil, protože šlo o vlastní jména (Moře Sirén; Moře Krizí).

⁵⁴ Dotaz v cql: [lemma="moře" & tag="N.*"]

⁵⁵ Dotaz v cql: [lemma = "moře"][tag = "N...2.*"]

Ze zbylých 306 nebylo metaforicky užitých pouze 5 (např. věty (57) a (58)).

(57) [M]ůžete se octnout v pravěkém *moři období* Jury (SYN2010)

(58) [R]odová linie keporkaků (...), kteří po *mořích světa* zpívají (...) své magické romance (SYN2010)

Jeden výsledek jsem označil za sporný (59), kvůli blízkosti zdrojové a cílové domény, spíše bych se však přikláněl k metaforické interpretaci.

(59) *moře vody* (SYN2010)

Zbylých 300 výsledků, tj. 97,72% bylo metaforických (např. se jednalo o (60) - (63)).

(60) *moře šampaňského* (SYN2010)

(61) *moře času* (SYN2010)

(62) jeden cvok v *moři cvoků* (SYN2010)

(63) Taková byla její povaha - plout v *moři nerozhodnosti*[.] (SYN2010)

Tvrzení b) tedy platí jak pro angličtinu, tak pro češtinu.

To, že se v obou jazycích používá stejná metafora, ještě nemusí znamenat, že se používá ve zcela stejných situacích. Vyhledal jsem proto “moře + substantivum v genitivu” ještě v paralelním korpusu InterCorp v8. Použil jsem dvě varianty dotazu. V první jsem prostě zadal stejný dotaz jako v předchozím případě a přidal zarovnaný korpus v angličtině. Ve druhé variantě jsem použil jen texty, které nepochází ani z jednoho ze zkoumaných jazyků a vyloučil z vyhledávání subkorpus zápisů z europarlamentu.

Tím jsem se pokusil zjistit, zda není míra shody metafory v českém a anglickém jazyce ovlivněna příliš doslovným překladem na úkor stylistiky (což bych očekával v zápisech europarlamentu, kvůli jejich technickému charakteru). Metaforičnost jsem tentokrát nehodnotil a předpokládal, že je 100%, pouze jsem počítal případy, kdy věty obsahují jak “moře”, tak “sea”.

V případě ovlivnění příliš doslovným překladem by byla hodnota shody ve druhé variantě dotazu menší, to se však nepotvrdilo a shoda byla u druhé variantu dotazu naopak o něco větší. Výsledky jsou uvedeny v tabulce 3.

	varianta 1	varianta 2
moře (InterCorp v8, cz)	383	87
sea (InterCorp v8, en)	257	61
shoda	67,10%	70,11%

Tab.3: Shoda užití metafory typu “sea of [NP]” na stejných textech v češtině a angličtině

Vidíme tedy, že i když je metafora typu “moře něčeho” v obou zkoumaných jazycích frekventovaná, nepoužívá se vždy ve stejných kontextech a ve více než 30% případů mluvčí angličtiny stejnou metaforu nepoužije.

K podobně na metafory bohatému schématu se v češtině dostaneme, i když moře nahradíme džunglí, tedy vyhledáním konstrukce ve tvaru: “džungle + substantivum v genitivu”.⁵⁶

Schéma má menší absolutní frekvenci (v SYN2010 bylo nalezeno 51 výskytů), a častěji se používá doslovně jako ekvivalent pro konstrukci “adjektivum z místního názvu + džungle” (např. queenslandská džungle/džungle Queenslandu). Přesto ze 45 výskytů (6 bylo vyřazeno z obdobných důvodů jako v předchozím případě) bylo 35 metaforických, to je 77,78%. Jistě se dá přijít i na další podobná schémata.

Využití zdrojů dostupných pro češtinu

V českém jazyce není momentálně dostupný žádný zdroj přímo určený pro analýzu metafory, máme však alespoň k dispozici bohatou základnu dat pro výzkum. Jde především o Český národní korpus, konkrétně jeho řady SYN (žánrově vyvážený korpus psané češtiny) a InterCorp (paralelní korpus češtiny a mnoha dalších jazyků), a Pražský

⁵⁶ Dotaz v cql: [lemma = "džungle"][tag = "N...2.*"]

závislostní korpus (PDT; Prague Dependency Treebank; Hajič a kol., 2006). PDT obsahuje čtyři úrovně anotace. Jeho stromová struktura obsahuje v nejvyšší, tzv. *tektogramatické* rovině například i uzly pro nevyjádřené větné členy.

Velký potenciál pro využití v analýze metafory je pak možné vidět ve dvou projektech: Lexikální databázi slovesných valenčních rámců *Verbalex* (Hlaváčková, 2007) a mělké ontologii *Sholva* (Grác & Rambousek, 2012; Grác, 2013).⁵⁷ Ty popisují v samostatných podkapitolách, přičemž se věnují především těm aspektům, které jsou relevantní pro výzkum metafory a popisu jejich využití.

Valenční lexikon VerbaLex

VerbaLex je v současnosti nejrozsáhlejší databází valenčních rámců sloves v češtině. Slovesa jsou řazena systémem synonymických řad a jejich valence je určena obsahem tzv. *slotů*. V nich jsou uloženy syntaktické a sémantické informace. Sémantické informace sestávají z určení sémantické role (např. agent) a přiřazení k určitému stromu v princetonském Wordnetu.

Hesla se v lexikonu zobrazují ve dvou druzích zápisu, v základním a komplexním valenčním rámci. Komplexní rámec přitom obsahuje také atribut *use*. Ten určuje způsob užití slovesa v daném kontextu a může nabývat hodnot: *prim* (doslovné užití), *fig* (obrazné užití - např. metaforické) a *idiom* (součást idiomu). Máme tedy zdroj anotovaný na metaforická užití (alespoň některých) sloves. Na obrázku 1 je zobrazen komplexní valenční rámec obrazného užití slovesa "šetřit".

⁵⁷ Jelikož je nástroj Sholva stále ve vývoji, dva zmíněné zdroje již neobsahují aktuální informace (např. o rozdělení tříd v ontologii). Nejaktuálnějším zdrojem, který se mi podařilo nalézt je bakalářská diplomová práce Zuzany Pelikánové (2015), používám proto údaje z ní.

šetřit₂^{impf} **uspořít**₁^{pf} **vyšetřit**_{n3}^{pf}

definition: úsporou získat
class: bill-54.5
passive: yes

1 šetřit₂, uspořít₁, vyšetřit_{n3} ≈
-frame: **AG**<person:1>_{obl kdo1} **VERB**_{obl} **TIME**<time:2>_{obl co4} **AG**<person:1>_{obl pro+koho4, na+koho4}
-example: vyšetřila si čas pro sebe (pf)
-synonym:
-use: fig
-reflexivity: no

2 šetřit₂, uspořít₁, vyšetřit_{n3} ≈
-frame: **AG**<person:1>_{obl kdo1} **VERB**_{obl} **TIME**<time:2>_{obl co4} **ACT**<act:2>_{opt pro+co4, na+co4, k+čemu3}
-example: vyšetřila si volnou chvíli pro četbu (pf)
-synonym:
-use: fig
-reflexivity: no

3 šetřit₂, uspořít₁, vyšetřit_{n3} ≈
-frame: **AG**<person:1>_{obl kdo1} **VERB**_{obl} **TIME**<time:2>|**EXT**<extent:1>_{obl co4} **ACT**<act:2>_{opt čím7}
-example: uspořil místo vhodným rozestavením nábytku (pf)
-synonym:
-use: fig
-reflexivity: no

4 šetřit₂, uspořít₁, vyšetřit_{n3} ≈
-frame: **INS**<machine:1>|**ACT**<process:1>_{obl co1} **VERB**_{obl} **REC**<person:1>_{opt komu3} **TIME**<time:2>|**ACT**<act:2>_{obl co4}
-example: myčka na nádobí šetří mamince čas (impf)
-example: nový výrobní postup uspoří člověku práci (pf)
-synonym:
-use: fig
-reflexivity: no

Obr. 1: Komplexní valenční rámec synsetu “šetřit, uspořít, vyšetřit”

VerbaLex tak poskytuje možnost identifikovat figurativní užití sloves (z nichž metafora bude tvořit signifikantní část) podle těchto manuálně vytvořených značek, za předpokladu, že použijeme vzorek, který je anotován na dostatečný počet atributů obsažených v lexikonu. Takovým vzorkem mohou být texty PDT v kombinaci s namapováním argumentů slovesa na synsety WordNetu.⁵⁸

Ještě zajímavější je možnost využít sémantických anotací jednotlivých slotů. Ty obsahují informaci o selekčním omezení a detekci jejich porušení lze tedy identifikovat věty s obrazně užitým slovesem a anomální případy. Tuto metodu by šlo opět aplikovat na anotovaný vzorek, jako je PDT nebo jiný korpus, nebo implementovat jako modul do

⁵⁸ Není jasné, zda je v tomto případě vhodnější český WordNet, obsahující podstatně méně dat s nerovnoměrným pokrytím znalostí, nebo překlad argumentu do angličtiny a využití princetonského WordNetu, což je metoda, jejíž nedostatky jsou vlastně příčinou nedostatků českého WN.

programu, který nejprve text automaticky označuje a příslušné argumenty spáruje s významy ve WordNetu.

Ontologie Sholva

Sholva je mělká ontologie vyvíjená v Centru zpracování přirozeného jazyka na Masarykově univerzitě. Od počátku je koncipována jako zdroj pro využití v NLP s nízkými náklady na vývoj.

Nízké finanční a časové náročnosti je docíleno zapojením crowdsourcingu. Anotátory jsou studenti MU, z nichž někteří nemají odborné jazykovědné znalosti ani zkušenosti s anotací. Anotační schéma je proto navrženo co nejjednodušeji a využívá opakované anotace stejných dat, kde pak rozhoduje souhlas většiny.

To může být velmi efektivním způsobem zpracování ontologie. Když uvážíme nesnadnost a z toho vyplývající malou spolehlivost zpracování sémantických úloh s velkou měrou subjektivních odchylek, zdá se sázka na kvantitu namísto nejisté kvality dobrou volbou.

Sholva je založena na *předpokladu otevřeného světa* (Open-world assumption; OWA), to znamená, že neznámé informace jsou vyhodnoceny jako potenciálně pravdivé nebo nepravdivé. Oproti tomu při předpokladu uzavřeného světa (Closed-world assumption; CWA) je informace, která není explicitně uvedena jako pravdivá, vyhodnocena automaticky jako nepravdivá.

OWA je tudíž vhodnější pro práci s lingvistickými daty, kde se předpokládá objevování nových jevů, a obecně pro zdroje znalostí o reálném světě. Sholva se od mnoha dalších znalostníchází předpokládajících otevřený svět liší tím, že obsahuje i explicitně negativní propozice (ve tvaru: A can-not-be-used-as B). Pokud se tedy informace dá odvodit z negativní propozice je vyhodnocena jako nepravdivá (Nevěřilová, 2014, 23).

Mělkost ontologie Sholva je dána tím, že je budována se snahou o co nejmenší počet pater v hierarchii a také co nejmenší počet tříd. Obsahuje čtyři kategorie: *_person*, *_event*, *_substance* a *_pointable*. Při tvorbě těchto kategorií se vycházelo ze základních

WordNetovských synsetů, podle nichž byly zpracovány tématické role ve VerbaLexu. Tím spolu oba projekty úzce souvisí.

Protože se však Sholva zabývá *použitím* slov spíše než jejich *významem*, její definice tříd se mírně odlišují, např. do *_person* jsou řazeny i instituce, které mohou osobu zastupovat, jako je “banka”, “radnice”, “ODS”, apod. (*_person* tedy sdružuje i metonymická užití; Pelikánová, 2015).

Využití sémantické sítě Sholva v analýze metafory může být několik. Například díky tomu, že řadí metonymie do kategorie podle užití, mohla by posloužit k jejich vyfiltrování z množiny metonymií, metafor a anomálií, získané po detekci porušení selekčních omezení. Také by se dala využít pro aplikaci výše popsané metody měření metaforičnosti (kapitola 2.8.2; Dunn, 2011), avšak modifikovanou a provedenou na češtině. Jelikož se třídy ontologie Sholva neshodují se SUMO ontologií, kterou používá Dunn, bylo by zajímavé zjistit, jaké výsledky bude poskytovat odlišné rozdělení domén a prozkoumat, jak se budou shodovat s intuicemi mluvčích.

Převedení existujících systémů do češtiny

V kapitole 4.1 této práce jsou vyjmenovány specifické nevýhody, které znesnadňují počítačové zpracování metafory v českém jazyce. Výhodu na druhou stranu může představovat fakt, že - obrazně řečeno - naskakujeme do rozjetého vlaku. Můžeme se tedy při tvorbě nástroje pro automatickou identifikaci metafory poučít z již existujících systémů. Bylo by ostatně neefektivní pokoušet se znovu vynalézat již jednou objevené, navíc transformací systému na jiný jazyk se daný algoritmus a také teoretická východiska, na kterých je vystaven, dočkají dodatečné evaluace.

Anotace korpusu pro trénink a evaluaci

Právě možnost evaluace je jedním ze základních požadavků, které je třeba splnit pro úspěšný vývoj jakéhokoli lingvistického softwaru. U složitých sémantických a pragmatických úloh je v současnosti stále nejspolehlivějším způsobem vytvoření zlatého

standardu. To u identifikace metafory znamená nechat manuálně anotovat dostatečně velký korpus textu na přítomnost metafory.

Pro anotaci metafory existuje zavedený manuál MIP od Praglejazz Group (2007) a byly již použity jeho upravené verze pro různé účely (Shutova & Teufel, 2010) i jazyky (Barbaczy a kol., 2010). Je tedy možnost vyjít z téhož postupu, případně i z jeho reflexe v NLP. Je totiž známo (Dunn, 2014), že anotace metodou MIP označuje za metaforické i silně lexikalizované případy. Ty se nehodí pro automatickou analýzu metafory a mohou být zpracovány běžnými algoritmy pro WSD. Při návrhu anotačního schématu pro češtinu se lze této nesnáze vyvarovat.

I když takový postup samozřejmě není nutný a je možné vytvářet malé anotované vzorky pro jednotlivá řešení úloh, je výhodnější postupovat systematicky a mít k dispozici jednotný zdroj s jasně stanovenou metodologií anotace, který daleko lépe ukáže výhody a nedostatky různých přístupů, které by v různorodých vzorcích nebyly patrné.

Rozsáhlý anotovaný vzorek užití metafory je také výhodný pro strojové učení. Mohl by být využit např. pro experimentování se *seed setem* anotovaných metaforických mapování, ze kterého může statistický systém odvozovat analogická mapování, která předtím nebyla známá, jako to dělá systém (Shutova a kol., 2010). Případně podrobit experimentálnímu testování jinou metodu původní, či replikovanou.

Replikace přístupu založeného na abstraktnosti kontextu

Vhodným kandidátem na převedení do češtiny může být metoda identifikace metafory v (Turney a kol., 2011; stručně popsána v kapitole 3.5.8). Má dostatečně vysokou míru úspěšnosti a je teoreticky dobře ukotvena, neboť myšlenka, že metafory nám slouží k převedení znalostí z jednoduché dobře známé oblasti do složité neznámé oblasti, je jedním ze základních tvrzení Lakoffovy teorie konceptuální metafory i Gentnerové teorie strukturálního mapování.

Základní komponentou systému je soubor čtyřiceti *paradigmatických* slov, pro jehož tvorbu Turney a kolektiv použili rozsáhlý slovník MRC Psycholinguistic Database Machine Usable Dictionary (Coltheart, 1981), obsahující 4295 slov ohodnocených podle míry

abstraktnosti mluvčími při psycholingvistických pokusech, a poměrně složitý statistický aparát. Takový zdroj v češtině dostupný není, tento problém by se však dal vyřešit minimálně dvěma způsoby.

Prvním je downgradování metody výběru paradigmatických slov. Přístup Turneye a kol. (2011) z podstatné části vychází ze staršího systému (Turney & Littman, 2003). Šlo o metodu pro automatické získávání informace o sémantické orientaci slov. Tu systém určoval podle sémantické shody se souborem 14 paradigmatických slov (7 pozitivních, 7 negativních), která byly vybrána manuálně.

Můžeme tedy zvolit stejný postup. Vodítkem při jejich výběru mohou být psychologické experimenty. Abstraktností a konkrétností se v češtině zabýval psycholog Bohumír Chalupa (2011). Metodou mentální chronometrie zkoumal dobu poznávání konkrétní a abstraktní, za použití dvou souborů slov, 20 konkrétní a 20 abstraktní s vysokou frekvencí a 20 konkrétní a 20 abstraktní s nízkou frekvencí.⁵⁹ V experimentu vyšlo najevo, že reakční doba je u konkrétní podstatně menší a korelace mezi délkou trvání poznání slov a abstraktností dosahovala od 0,833 do 0,882, přitom u dvou skupin probandů (dvou III. ročníků gymnázia) se shoda pohybovala mezi 0,904 a 0,921. Reakční čas je tedy extrémně spolehlivou metodou určování abstraktnosti.⁶⁰

Pro výběr paradigmatických slov by se dali využít nejlepší kandidáti (slova s nejkratší a nejdelší dobou poznání) z Chalupova vzorku, nebo i provést nové experimenty (například s menším počtem probandů a větším počtem slov) a použít jejich výsledky.

Druhým způsobem, jak získat soubor paradigmatických slov, je využít zdroje, který Turney a kol. (2011) vytvořili při tvorbě svého systému pro identifikaci metafory. Původní slovník o přibližně čtyřech tisícovkách slov automaticky rozšířili na 114501 slov (při testování oproti části původního slovníku měla jejich metoda přesnost 0,81). Tento velký slovník by se dal automaticky přeložit do češtiny (je pravděpodobné, že slova budou mít v obou jazycích podobnou míru abstraktnosti) a pak získat paradigmatická slova stejnou počítačovou metodou jako Turney a jeho kolegové.

⁵⁹ Ukázalo se však, že frekvence neměla na dobu poznání slov vliv

⁶⁰ V zásadě by se dala abstraktnost slov, namísto slovníkových definic, definovat přímo reakčním časem.

Celý systém by se dal dále modifikovat tím způsobem, že by bral v úvahu i hodnotu abstraktnosti adjektiv a sloves, testovaných na metaforičnost. Mohl by dosáhnout lepších výsledků vyřazením vysoce abstraktních slov, u kterých je mizivá možnost, že budou použita metaforicky, z kandidátů na metaforu.⁶¹

Shrnutí

V češtině tedy existují způsoby, jak se s použitím zdrojů, které jsou k dispozici pokoušet o automatické zpracování metafor, a lze realizovat několik již vymyšlených způsobů vhodných k testování (z těch, které jsem v této kapitole nezmínil ještě třeba jazykově nezávislá metoda Yulie Tsvetkovové (2013) založená na překladu). Dalším krokem by mělo být uvedení některé metody do praxe, a také výběr a příprava korpusového materiálu k použití při její evaluaci.

⁶¹ Na neschopnost vysoce abstraktních slov být užita metaforicky upozorňuje např. (Hanks, 2006) a (Shutova a kol., 2010)

Závěr

V této práci jsem se snažil uchopit poměrně široké téma automatického zpracování metafor od jeho teoretických základů až k praktickým aplikacím v češtině. Velká obecnost tématu byla možná také proto, že téma analýzy metafor není v českém prostředí příliš populární. Práce by tak mohla sloužit zájemcům o strojové zpracování obrazného vyjadřování jako úvod do problematiky a návod, jakým způsobem začít s vývojem prvních specializovaných nástrojů pro tuto úlohu v češtině.

Pokoušel jsem se, aby bylo jasně patrné propojení mezi popsanými teoriemi a jejich praktickými aplikacemi. Realizoval jsem také dva malé korpusové experimenty, jejichž výsledky byly nepřekvapivé a nasvědčují tomu, že používání metafor vykazuje značné podobnosti napříč různými jazyky.

Co se týče praktického užití automatické identifikace metafor, reálnější se v současnosti zdají být využití pro výzkum než například pro strojové překladače, což však hodnotu této úlohy nijak nesnižuje. Využívat pro výzkum metafor ve větší míře kvantitativní metody je naopak žádoucím cílem.

Je pravděpodobné, že při dalším vývoji disciplíny budou nejúspěšnější systémy čím dál častěji založené na několika modulech, z nichž každý bude identifikovat metaforu jinou metodou. Tím se využijí nejlepší vlastnosti každé metody, zatímco jejich slabiny nebudou mít na kvalitu detekce vliv. V současnosti již s takovým přístupem pracuje Jonathan Dunn a kolektiv (2014).

Je ostatně možné, že rozeznávání různých typů metafor, či jiných projevů figurativního jazyka, je u lidí také realizováno několika systémy, kdy ten nejefektivnější přebírá kontrolu, podobně jako u *multiple drafts modelu* vědomí amerického filozofa myslí Daniela Denneta (1991).

Použité zdroje:

Aristotelés. (1993). *Poetika*. Praha: Gryf. ISBN 80-85829-01-0.

Asch, S.E., Nerlove, H. (1960). The development of double function terms in children: An exploratory investigation. *Perspectives in Psychological Theory*. New York: International Universities Press, Inc.

Alverson, H. (1994). *Semantics and experience: Universal metaphors of time in English, Mandarin, Hindi, and Sesotho*. Baltimore: Johns Hopkins University Press. ISBN 0801848113.

Ayoob, E. (2007). Black & Davidson on Metaphor. *Macalester Journal of Philosophy*. Vol. 16, no. 1, pp. 56 - 64.

Babarczy, A., Simon, E., Sereny, A. (2010) *The Automatic Identification of Conceptual Metaphors in Hungarian Texts: A Corpus-Based Analysis*. In Bel, N., Daille, B., Vasiljevs, A. (2010). *Proceedings Methods for the automatic acquisition of Language Resources and their evaluation methods*. ISBN 2-9517408-6-7.

Barnden, J. (2007). Metaphor, semantic preferences and context-sensitivity. *Words and Intelligence I*. Vol. 36, no. 2, pp. 39 – 62.

Barnden, J., Glasbey, S., Lee, M., Wallington, A. (2002). Reasoning in Metaphor Understanding: The ATT-Meta Approach and System. *Proceeding COLING '02 Proceedings of the 19th international conference on Computational linguistics*. Vol. 2, pp. 1 – 5.

Birke, J., Sarkar, A. (2006) *A clustering approach for the nearly unsupervised recognition of nonliteral language*. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.

Black, M. (1955). *Metaphor. Proceedings of the Aristotelian Society, New Series*. Blackwell Publishing. Hoboken, New Jersey.

Black, M. (1979). *More about metaphor*. In Ortony, A. (1993). *Metaphor and thought*. Cambridge University Press. Cambridge, United Kingdom. ISBN 9780521405614.

Black, M. (1962). *Models and Metaphors*. Cornell University Press. Ithaca, New York. ISBN 0801400414.

Bojar O., Cinková S., Hajič J., Hladká B., Kuboň V., Mírovský J., Panevová J., Peterek N., Spoustová J., Žabokrtský Z. (2012). *The Czech Language in the Digital Age*. Springer-Verlag Berlin Heidelberg, Berlin, Germany, ISBN 978-3-642-30705-8, 79 pp.,

Bollegala, D., Shutova, E. (2013). Metaphor interpretation using paraphrases extracted from the web. *Plosone*. Vol. 8, no. 9, pp. 1 – 10.

Bratož, S. (2004). A comparative study of metaphor in English and Slovene popular economic discourse. *Managing Global Transitions*. Vol. 2, no. 2, pp. 179 – 196.

Bratož, S. (2012). Critical Approaches to Discourse Analysis across Disciplines. *Slovenian and US Elections in Metaphors: A Cross-linguistic Perspective*. Vol. 5, no. 2, pp. 120 – 136.

Bratož, S. (2013). The Anthropomorphic Metaphor in Slovene and English Wine Tasting Discourses. *English Language Overseas. Perspectives and Enquiries*. Vol 10, pp. 23-35.

Cameron, L. (2003). *Metaphor in educational discourse*. London: Continuum. ISBN 0-8264-4939-5.

Casasanto, D. (2008). Who is afraid of the Big Bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*. Vol. 58, pp. 63–79.

Casasanto, D. (2013). Kdopak by se Whorfa bál? Mezijazykové rozdíly v jazykové a kognitivní reprezentaci času. *Studie aplikované lingvistiky*. Vo. 1, pp. 77 – 92.

Cinková, S., Holub, M, Rychlý, P., Smejkalová, L., Šindlerová, J. (2010) *Can Corpus Pattern Analysis Be Used in NLP?* In: Sojka et al. (Ed.): TSD 2010, LNAI 6231, pp. 67–74, 2010.
Springer-Verlag Berlin Heidelberg

Max Coltheart. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A(4):497–505.

Davidson, D. (1978). What Metaphors Mean. *Critical Inquiry*. Vol. 5, No. 1, Autumn.

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown & Co.

Dunn, J. (2011). Gradient Semantic Intuitions of Metaphoric Expressions. *Metaphor and Symbol*. Vol. 26, pp. 53–67.

Dunn, J. (2013). *What metaphor identification systems can tell us about metaphor-in-language*. Proceedings of the First Workshop on Metaphor in NLP. ISBN 978-1-937284-47-3.g

Dunn, J. (2014). *Measuring metaphoricity*. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. ISBN 978-1-941643-00-6.

Dunn, J. et al. (2014). *Language-Independent Ensemble Approaches to Metaphor Identification*. Association for the Advancement of Artificial Intelligence

Falkenhainer, B., Forbus, K. D., Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*. Vol.41, pp. 1-63.

Fass, D. (1991). A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*. Vol. 17, no. 1, pp. 49 – 90.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge University Press. Cambridge, United Kingdom. ISBN 0521599539.

Fauconnier, G., Turner, M., (2002). *The Way We Think*. Basic Books, New York. ISBN 0465087868.

Fauconnier, G., Turner, M. (2003). *Metaphor, Metonymy, and Binding*. In Barcelona, A. *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Walter de Gruyter, Berlin and New York. ISBN 3110175568.

Fellbaum, Ch. (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press. ISBN 026206197X.

Feldman, J. (2006). *From Molecule to Metaphor*. Cambridge, MA: MIT Press. ISBN 9780262562355.

Feldman, J., Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*. No. 89, vol. 2, pp.385–392.

Feldman, J. A., Bergen, B. (2006) *It's the Body, Stupid: Concept Learning According to Cognitive Science*. International Computer Science Institute, Tech. Rep. ICSI TR-06-002

Fillmore, Ch. J. (1967). *The Case for Case*. In Cole, P., Sadock, J. M. (1977) *Syntax and Semantics, vol. 8: Grammatical Relations*. Waltham, Massachusetts: Academic Press. ISBN 0126135088.

Fillmore, Ch. J. (1975). An Alternative to Checklist Theories of Meaning. In Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, pp. 123-131.

Fillmore, Ch. J., Johnson, Ch., Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*. Vol. 16, no. 3, pp. 235-250.

French, R. M. (1990) Subcognition and the limits of the Turing test. *Mind*. Vol. 99, pp. 53 – 66.

Forrester, S. (2010) *Theories of Metaphor in Seventeenth and Eighteenth-Century British Philosophy*. Literature Compass, Volume 7, Issue 8, pages 610–625.

Geary, J. (2012). *I Is an Other: The Secret Life of Metaphor and How It Shapes the Way We See the World*. New York: Harper Perennial. ISBN0061710296.

Gedigian, M., Bryant, J., Narayanan, S., Ciric, B. (2006) *Catching Metaphors*. Proceedings of the 3rd Workshop on Scalable Natural Language Understanding. Association for Computational Linguistics.

Gentner, D. (1977). If a tree had a knee, where would it be? Children's performance on simple spatial metaphors. *Papers and Reports on Child Language Development*. Vol. 13, pp. 157 – 164.

Gentner, D. (1982). Are Scientific Analogies Metaphors? In Miall, D. (1982). *Metaphor: Problems and Perspectives*. Hemel Hempstead: Humanities Press. ISBN 0710800339.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*. Vol. 7, no. 2, pp. 155 – 170.

Gentner, D. (1986). Evidence for structure - mapping in analogy and metaphor. Department of computer science. University of Illinois, Urbana, Illinois.

Gentner, D., Jeziorski, M. (1993). The shift from metaphor to analogy in Western science. In Ortony, A. (1993). *Metaphor and Thought*. Cambridge: Cambridge University Press. ISBN 0-521-40547-5.

Gentner, D., Bowdle, B. F., Wolff, P., Boronat, C. (2001). *Metaphor Is Like Analogy*. In Gentner, D., Holyoak, K. J., Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. MIT Press, Cambridge. ISBN 9780262571395.

Gibbs, R. W. (1984) Literal Meaning and Psychological Theory. *Cognitive science*. Vol. 8, no. 3, pp. 275–304.

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Reyes, A., Barnden, J. (2015) *Sentiment Analysis of Figurative Language in Twitter* (konference). URL: <http://alt.qcri.org/semEval2015/task11/>.

Goatly, A. (1997). *The Language of Metaphors*. Routledge, London. ISBN 0415128765.

Glucksberg, S., Keysar, B. (1990). Understanding Metaphorical Comparisons: Beyond Similarity. *Psychological Review*. Vol. 97, No. 1, pp. 3 – 18.

Grác, M., Rambousek, A. (2012). Low-cost ontology development. In: 6th International Global Wordnet Conference Proceedings. Matsue: Toyohashi University of Technology, 2012, s. 299–304. ISBN 978-80-263-0244-5.

Grác, M. (2013). *Rapid Development of Language Resources*. Brno. Dizertační práce.

Grady, J. (1997). *Foundations of Meaning: Primary Metaphors and Primary Stress*. Disertační práce. Department of Linguistics University of California, Berkeley.

Grady, J. (1998). The Conduit Metaphor revisited: a reassessment of metaphors for communication. In Koenig, J. P. (1998). *Discourse and Cognition: Bridging the Gap*. Stanford: CSLI. ISBN 1575861151.

Grice, H. P. (1975). Logic and conversation. In Cole, P., Morgan, J. L. (1975). *Syntax and semantics 3: Speech arts*. New York: Academic. ISBN 0127854231.

Hajič, Jan a kol. (2006) Prague Dependency Treebank 2.0. Dostupné online na <http://ufal.mff.cuni.cz/pdt2.0>.

Hanks, P. (2004a). Corpus Pattern Analysis. In *Euralex Proceedings*. Lorient, France: Université de Bretagne-Sud. Vol. I, pp. 87 – 98. ISBN 2952245703.

Hanks, P. (2004b). The Syntagmatics of Metaphor and Idiom. *International Journal of Lexicography*. Vol.17, no. 3, pp. 245 – 274.

Hanks, P. (2006). *Metaphoricity is Gradable*. in Stefanowitsch, A., Gries, S. (2006). *Corpora in Cognitive Linguistics*. Vol. 1: Metaphor and Metonymy. Berlin and New York: Mouton de Gruyter. ISBN 978 – 3 – 11 – 019826 – 3.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge: The MIT Press. ISBN 0262018578.

Hanks, P., Church, K. W. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*. Vol. 16, No. 1, pp. 22 – 29.

Hanks, P., Pustejovsky, J. (2005). A Pattern Dictionary for Natural Language Processing. *Revue Francaise de Langue Appliquée*. Vol. 10, no. 2, pp. 1 – 19.

Hesse, M. (1966). *Models and Analogies in Science*. Notre Dame University Press: Notre Dame, Indiana. ISBN 0268001820.

Hill, J. H, Mannheim, B. (1992). Language and World view. *Annual Review of Anthropology*. Vol. 21, pp. 381–406.

Hlaváčková, D. (2007). Databáze slovesných valenčních rámců Verbalex. PhD thesis, Masaryk university, Brno

Hoijer, H. (1954). *Language in culture: Conference on the interrelations of language and other aspects of culture*. In Goodenough, W. H. (1955). *Language*. *Linguistic Society of America*. Vol. 31, No. 2, pp. 241-245.

Chalmers, D. J., French, R. M., Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*. Vol. 4, no. 3, pp. 185–211.

Charteris-Black, J., Ennis, T. (2001). A comparative study of metaphor in Spanish and English financial reporting. *English for Specific Purposes*. Vol. 20, pp. 249-266.

Indurkha, B. (1986). Constrained semantic transference: a formal theory of metaphors. In Frieditis, A., Kaufmann, M. (1987) *Analogica: Proceedings of the First Workshop on Analogical Reasoning*. Pitman Publishing: London. ISBN 0273087800.

Katz, J. J., Fodor, J. A. (1963). The Structure of a Semantic Theory. *Language*. Vol. 39, No. 2, pp. 170 – 210.

Karov, Y., Edelman, S. (1998) Similarity-based word sense disambiguation. *Computational Linguistics*. Vol. 24, no. 1, pp. 41 – 59.

Keysar, B., Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*. Vol. 34, no. 1, pp.89 – 109.

Kingsbury, P., Palmer, M, (2002) *From TreeBank to PropBank*. Third International Conference on Language Resources and Evaluation, LREC-02.

Korkontzelos, I., Torsten Z., Zanzotto, F. M., Biemann, Ch. (2013). *Semeval-2013 task 5: Evaluating phrasal semantics*.

Kövecses, Z. (2010). *Metaphor: A Practical Introduction*. Oxford University Press: Oxford. ISBN 978-0-19-537494-0.

Krishnakumaran, S., Zhu, X. (2007) *Hunting Elusive Metaphors Using Lexical Resources*. In Proceedings of the Workshop on Computational Approaches to Figurative Language. Pp. 13 – 20.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press. ISBN 9780226458113.

Maarouf, I. E., Baisa, V., Bradbury, J., Hanks, P. (2014) Disambiguating Verbs by Collocation: Corpus Lexicography meets Natural Language Processing. In LREC (pp. 1001-1006).

Mason, Z. J. (2004) CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System. *Computational Linguistics* Vol. 30, No. 1, pp. 23 – 44.

Martin, J. M. (1988). *Representing Regularities in the Metaphoric Lexicon*. Proceedings of the 12th International Conference on Computational Linguistics.

Martin, J. M. (1990). *A Computational Model of Metaphor Interpretation*. Waltham, Massachusetts: Academic Press Professional. ISBN:0-12-474730-2.

Martin, J. M. (1994). Metabank: A knowledge-base of metaphoric language conventions, *Computational Intelligence*. Vol. 10, no. 2, pp. 134 – 149.

Martin, J. M. (1996) Computational Approaches to Figurative Language. *Journal of Metaphor and Symbolic Activity*. Vol. 11, no. 1, pp. 85 – 100.

Martin, J. M. (2006). *A Rational Analysis of the Context Effect on Metaphor Processing*. In Gries, S., Stefan, T., Stefanowitsch, A. (2006) *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin / New York: Mouton de Gruyter. ISBN 978-3-11-019827-0.

Narayanan, S. S. (1997). Knowledge-based Action Representations for Metaphor and Aspect. Disertační práce. Computer Science. University of California, Berkeley.

Narayanan, S. (1999). Moving Right Along: A Computational Model of Metaphoric Reasoning about Events. Proceedings of the 16th national conference on artificial intelligence. ISBN 978-0-262-51106-3.

Nevěřilová, Z. (2014). *Paraphrase and Textual Entailment Generation in Czech*. Masaryk University. Brno. PhD thesis.

Niles, I., Pease, A. (2001) Towards a Standard Upper Ontology. Proceedings of FOIS-2001.

Nirenburg, S., Raskin, V. (2004). *Ontological semantics*. Cambridge, MA.: MIT Press. ISBN: 0262140861.

Núñez, R. E., Sweetser, E. (2006) With the Future Behind Them: Convergent Evidence From Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time. *Cognitive Science*. Vol. 30, no. 3, pp. 401 – 50.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. Chicago University Press: Chicago. ISBN 0226468011.

Lakoff G. (1987) The Death of Dead Metaphor. *Metaphor and Symbolic Activity*. Vol, 2, no. 2, pp. 143 – 147.

Lakoff, G. (1992). *The contemporary theory of metaphor*. In Ortony, A. (1993). *Metaphor and Thought*. Cambridge: Cambridge University Press. ISBN 0-521-40547-5.

Lakoff, G., Johnson, M. (2014). *Metafory, kterými žijeme*. Brno: Host. ISBN LAKOFF. ISBN 978-80-7491-152-1.

Lakoff, G., Espenson, J., Schwartz, A. (1991). *Master Metaphor List*. Second draft copy. Technical report, Cognitive Linguistics Group, University of California Berkeley.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford University Press: Redwood City, California. ISBN 9780804738514.

Lönneker, B. (2004). Lexical databases as resources for linguistic creativity: Focus on metaphor. Proceedings of the LREC 2004 Satellite Workshop on Language Resources and Evaluation: Language Resources for Linguistic Creativity. ISBN 2-9517408-1-6.

Lönneker-Rodman, B. (2008). The hamburg metaphor database project: issues in resource creation. *Language Resources and Evaluation*. Vol. 42, no. 3, pp. 293–318.

Lönneker-Rodman, B., Narayanan, S. (2008). Computational Approaches to Figurative Language. *Cambridge Encyclopedia of Psycholinguistics*. Cambridge: Cambridge University Press.

Ovchinnikova, E., Israel, R., Wertheim, S., Zaytsev, V., Montazeri, N., Hobbs, J. (2014). *Abductive inference for interpretation of metaphors*. Proceedings of the Second Workshop on Metaphor in NLP.

Pala K., Čapek T., Zajíčková B., Bartůšková D., Kulková K., Hlaváčková D., Hoffmannová P., Bejček E., Straňák P., Hajič J. (2010). Český WordNet 1.9 PDT. Data/software, UFAL MFF UK, <http://hdl.handle.net/11858/00-097C-0000-0001-4880-3>, Dec 2010.

Pelikánová, Z. (2015) *Rozšíření sémantické sítě s využitím korpusu*. Masarykova univerzita. Brno. Bakalářská práce

Peters, W., Peters, I. (2000). Lexicalised Systematic Polysemy in WordNet. Proceedings of LREC-2000, Athens, Greece.

Pinker, S. (2008). *The Stuff of Thought: Language as a Window into Human Nature*. Penguin Books: London. ISBN 0143114247.

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*. Vol. 2, no. 1, pp. 37–63.

Pustejovsky, J. (1991). The Generative Lexicon. *Computational Linguistics*. Vol. 17, No. 4, pp. 409 – 441.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge:MIT Press. ISBN 0262161583.

Pustejovsky, J., Rumshisky, A., Hanks, P. (2004). *Automated Induction of Sense in Context*. Geneva: COLING Proceedings. Stroudsburg, Pa. : Association for computational linguistics. ISBN 9781634396783.

Reddy, M. J. (1979). *The conduit metaphor: A case of frame conflict in our language about language*. In Ortony, A.(1993) *Metaphor and Thought*. Cambridge University Press:Cambridge. ISBN 0-521-29626-9.

Reining, A., Loenneker-Rodman, B, (2007). Corpus-driven metaphor harvesting. Proceedings of the HLT/NAACL-07 Workshop on Computational Approaches to Figurative Language.

Resnik, P. (1993). Selection and information: A class-based approach to lexical relationships. A dissertation in Computer and Information Science. University of Pennsylvania.

Searle, J. (1979): Metaphor. In Ortony, A. (1993) *Metaphor and Thought*. Cambridge University Press:Cambridge. ISBN 0-521-29626-9.

Shalizi, C. R. (2003). Analogy and Metaphor. URL <http://www.cscs.umich.edu/~crshalizi/notabene/analogy.html>.

Shutova, E. (2010a). Automatic Metaphor Interpretation as a Paraphrasing Task. In Proceedings of NAACL 2010, Los Angeles, USA.

Shutova, E. (2010b). Models of Metaphor in NLP. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Pp. 688–697.

Shutova, E. (2011). Computational approaches to figurative language. PhD thesis, Computer Laboratory, University of Cambridge, UK.

Shutova, E. (2015). Design and Evaluation of Metaphor Processing Systems. *Computational Linguistics*. Vol. 41, no. 4, pp. 1 – 46.

Shutova, E., Teufel, S. (2010). Metaphor corpus annotated for source - target domain mappings. In Proceedings of LREC 2010. ISBN 2-9517408-6-7.

Shutova, E., Sun, L., Korhonen, A. (2010). Metaphor Identification Using Verb and Noun Clustering. In Proceedings of COLING 2010, Beijing, China.

Shutova, E., Teufel, S., Korhonen, A. (2012). Statistical Metaphor Processing, *Computational Linguistics*, 39(2).

Shutova, E., Sun, L. (2013). Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering. *Association for Computational Linguistics. Proceedings of NAACL-HLT 2013*, pp. 978–988.

Skorczynska, H., Piqué-Angordans, S. J. (2005). A corpus-based description of metaphorical marking patterns in scientific and popular business discourse. *Metaphoric. de 9*, pp. 112 – 129.

Skorczyńska, H., Ahrens, K. (2015). A corpus-based study of metaphor signaling variations in three genres. *Text & Talk*. Vol. 35, No. 3, Pp. 359–381.

Sperber, D., Wilson, D. (1995). *Relevance: Communication and Cognition*, Oxford/Cambridge: Blackwell Publishers, ISBN 978-0-631-19878-9.

Sperber, D., Wilson, D. (2008). *A deflationary account of metaphors*. In Gibbs, R. W. (2008). *The Cambridge Handbook of Metaphor and Thought*. Cambridge University Press: Cambridge, United Kingdom. ISBN 9780511816802.

Stachová, J. (1992). Metafora podle J. R. Searla a Maxe Blacka. *Slovo a slovesnost*. Ročník 53, číslo 4, str. 283 – 292.

Steen, G. J., Dorst, A. G., Herrman, B., Kaal, A. A., Krennmayr, T. (2010) Metaphor in usage. *Cognitive Linguistics*. Vol. 21, no. 4, pp. 757–788.

Stefanowitsch, A. (2004). *HAPPINESS in English and German: A metaphorical-pattern analysis*. In Achard, M., Kemmer, S. (2004). *Language, Culture, and Mind*. Stanford: CSLIPublications. ISBN 9781575864648.

Stefanowitsch, A. (2006). *Words and their metaphors: A corpus-based approach*. In Gries, S., Stefan, T., Stefanowitsch, A. (2006) *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin / New York: Mouton de Gruyter. ISBN 978-3-11-019827-0.

Sun, L., Korhonen, A. (2009). *Improving Verb Clustering with Automatically Acquired Selectional Preferences*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).

Thyer, B. A., Pignotti, M. G. (2015). *Science and Pseudoscience in Social Work Practice*. Springer Publishing Company. ISBN 9780826177698.

Tourangeau, R., Sternberg, R. (1982). Understanding and appreciating metaphors. *Cognition*, vol. 11, no. 3, pp. 203–244.

Tourangeau, R., Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology* Vol. 13, no. 1, Pp. 27-55.

Tsvetkov, Y., Mukomel, E., Gershman, A. (2013) Cross-Lingual Metaphor Detection Using Common Semantic Features. Proceedings of the First Workshop on Metaphor in NLP, Atlanta, Georgia, pp 45-51.

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., Dyer, C. (2014) Metaphor Detection with Cross-Lingual Model Transfer. ACL 2014, pp 248--258.

Tufiş, D., Cristea, D., Stamou, S. (2004). BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal of Information Science and Technology*. Vol. 7, No. 1-2, pp. 9 – 43.

Turney, P. D., Littman, M. L. (2003). *Measuring praise and criticism: Inference of semantic orientation from association*. ACM Transactions on Information Systems, 21(4):315–346.

Turney, P. D. (2008). The Latent Relation Mapping Engine: Algorithm and Experiments. *Journal of Artificial Intelligence Research*. Vol. 33, pp. 615 - 655.

Turney, P. D., Neuman, Y., Assaf, D., Cohen, Y. (2011) Literal and Metaphorical Sense Identification through Concrete and Abstract Context. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.

Vaňková, I. (2002). *Lingvistika mysli a těla*. Vesmír. číslo 81, pp. 627-629.

Vaňková, I., Nebeská, I., Saicová Římalová, L., Šlédrová, J. (2005) *Co na srdci, to na jazyku*. Praha: Karolinum. ISBN 80-246-0919-3.

Veale, T., Hao, Y. (2007) *Comprehending and Generating Apt Metaphors: A Web-driven, Case-based Approach to Figurative Language*. Proceedings of AAAI.

Veale, T., Hao, Y. (2008) *A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors*. Proceedings of the 22nd International Conference on Computational Linguistics.

Vosniadou, S., (1986). *Children and metaphors*. Technical Report No. 370. University of Illinois, Urbana-Champaign.

Vossen, P. (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer, Dordrecht, The Netherlands. ISBN 978-94-017-1491-4.

Way, E. C. (1993). Review of "Metaphor and cognition: an interactionist approach" by Bipin Indurkha. *Journal of Computational Linguistics*, Vol. 19, no. 3, pp. 553-556.

Wilks, Y (1975). A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*. Vol. 6, pp. 53 – 74.

Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence*. Vol. 11, no. 3, pp. 197 – 223.

Wilson, D., Carston, R. (2007). Metaphor and the 'Emergent Property' Problem: A Relevance-Theoretic Approach. *Baltic International Yearbook of Cognition, Logic and Communication*. Vol. 3, pp. 1 – 40.

Použité korpusy:

Křen, M. – Bartoň, T. – Cvrček, V. – Hnátková, M. – Jelínek, T. – Kocek, J. – Novotná, R. – Petkevič, V. – Procházka, P. – Schmiedtová, V. – Skoumalová, H.: SYN2010: žánrově vyvážený korpus psané češtiny. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <http://www.korpus.cz>

Rosen, A. – Vavříň, M.: Korpus InterCorp – čeština, verze 8 z 4. 6. 2015. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>