



IMPLICITNÍ DISKURZNÍ VZTAHY V ČEŠTINĚ

Šárka Zikánová



ÚSTAV FORMÁLNÍ
A APLIKOVANÉ LINGVISTIKY

 **STUDIES IN COMPUTATIONAL
AND THEORETICAL LINGUISTICS**

Šárka Zikánová

IMPLICITNÍ DISKURZNÍ VZTAHY V ČEŠTINĚ

Published by the Institute of Formal and Applied Linguistics
as the 22nd publication in the series
Studies in Computational and Theoretical Linguistics.

Editor-in-chief: Jan Hajič

Editorial board: Nicoletta Calzolari, Mirjam Fried, Eva Hajičová,
Petr Karlík, Joakim Nivre, Jarmila Panevová,
Patrice Pognan, Pavel Straňák, and Hans Uszkoreit

Reviewers: Pavel Kosek, Masaryk University, Brno
Barbora Skarabela, University of Edinburgh

The author gratefully acknowledges the support of the grant project GA 17-03461S *Implicit relations in text coherence* of the Czech Science Foundation (GA ČR) and of the institutional funds of Charles University.

Printed by MatfyzPress.

Copyright © Institute of Formal and Applied Linguistics, 2021

ISBN 978-80-88132-12-7

Obsah

1	Úvod	1
1.1	Koherence textu	2
I	Teoretické východisko	5
2	Základní pojmy	9
2.1	Text a diskurz	9
2.2	Diskurzní argument	10
2.3	Diskurzní konektor	11
2.4	Diskurzní vztah	12
3	Kontext výzkumu	15
3.1	Teoretické rámce formálního popisu diskurzních vztahů	15
3.2	Implicitní diskurzní vztahy a elektronické korpusy	16
3.3	Čeština a výstavba textu	16
3.4	Výstavba textu v psycholingvistickém výzkumu	18
4	Koncepce výzkumu diskurzních vztahů	19
4.1	Penn Discourse Treebank 2.0	19
4.1.1	Významová klasifikace vztahů v Penn Discourse Treebank 2.0	21
4.2	Funkční generativní popis a tektogramatická rovina	27
4.2.1	Významová klasifikace tektogramatických vztahů v Pražském závislostním korpusu (funktory)	29
4.3	Pražský diskurzní korpus a navazující výzkum diskurzních vztahů	36
4.3.1	Významová klasifikace diskurzních vztahů v Pražském diskurzním korpusu	38

4.3.2	Od Pražského závislostního korpusu po Rozšířenou anotaci diskurzních vztahů (PDiT-EDA 1.0)	48
4.4	Korpus PDiT-EDA 1.0 (Rozšířená anotace diskurzních vztahů)	50
4.4.1	Implicitní diskurzní vztahy	51
4.4.2	Další druhy vztahů	53
4.4.3	Mezianotátorská shoda	58
4.4.4	Typické příčiny neshody na existenci implicitního diskurzního vztahu	60
4.4.5	Typické příčiny neshody na významovém popisu diskurzního vztahu	65
4.4.6	Mezianotátorská shoda u implicitních diskurzních vztahů ve srovnání s jinými druhy textových vztahů	74
II	Analýza	77
5	Výsledky anotace v korpusu PDiT-EDA 1.0	81
5.1	Implicitní a explicitní diskurzní vztahy	81
5.2	Implicitnost a sémantické typy diskurzních vztahů	83
5.3	Způsoby vyjádření různých sémantických typů diskurzních vztahů	84
5.4	Souvislost mezi významem diskurzních vztahů a jejich vyjádřením	93
5.5	Distribuce z hlediska formy	94
5.6	Distribuce z hlediska významu	94
5.7	Závěr	96
6	Implicitní vztahy v psycholingvistických experimentech	97
6.1	Metodologie experimentů	97
6.2	Výsledky experimentů	103
6.2.1	Konfrontace	103
6.2.2	Přípustka	105
6.2.3	Současnost	106
6.2.4	Nesoučasnost	107
6.3	Závěr a diskuse	108
7	Shrnutí	113
7.1	Korpus PDiT-EDA 1.0	113
7.2	Mezianotátorská shoda	114

7.3	Výsledky anotace v korpusu PDiT-EDA 1.0	116
7.4	Implicitní vztahy v psycholingvistických experimentech	117
7.5	Závěr	118
8	Summary: Implicit Discourse Relations in Czech	121
8.1	Basic terms	122
8.2	State of the art	123
8.3	Methodology	125
8.4	Data	126
8.5	Results of the annotation of implicit discourse relations	131
8.6	Implicit discourse relations in psycholinguistic experiments	135
8.7	Conclusion	139
	Seznam obrázků	141
	Seznam tabulek	143
	Seznam zkratek	145
	Literatura	147
	Rejstřík	157
A	Výchozí sada dokladů pro experiment s kontrastivními vztahy	159
A.1	Přípustka	159
A.2	Konfrontace	163
A.3	Výplňové položky 1	167
B	Výchozí sada dokladů pro experiment s časovými vztahy	177
B.1	Současnost	177
B.2	Nesoučasnost	181
B.3	Výplňové položky 2	185

Poděkování

Děkuji milým kolegyním a kolegům v Ústavu formální a aplikované lingvistiky za vytvoření podnětného prostředí, v němž probíhal výzkum implicitních vztahů.

Dr. Jiřímu Mírovskému jsem vděčna za soustavnou odbornou zpětnou vazbu a všestrannou technickou podporu.

Dr. Pavlíně Synkové patří dík za spolupráci při vytváření korpusu PDiT-EDA 1.0.

Doc. Filipu Smolíkovi z laboratoře LABELS (FF UK a Psychologický ústav AV ČR) děkuji za spolupráci na psycholingvistických experimentech.

Za podporu ve finální fázi výzkumu jsem zavázána prof. Janu Hajičovi a dr. Lucii Polákové.

Zvláštní dík patří prof. Evě Hajičové, mé životní učitelce, za odborné i lidské vedení a za laskavou podporu mé vědecké práce.

Za pochopení, pomoc a samostatnost děkuji svým skvělým milovaným dětem.

Motto:

*Nemohu si pomoci: bez účasti mocností tak málo vědeckých, jako jsou zdravý lidský rozmysl
a lidské svědomí, se zřejmě neobejde nic.*

— Václav Havel

Motto 2:

K tématu dále uvádíme?	<i>Široký hluboký ty vltavský tůň</i>
A i přesto?	
Konkrétně?	
Kdežto?	<i>Měl jsem já panenku</i>
V důsledku čehož?	
Neboli?	
A přesto?	<i>Ted' mi srdce stůně</i>

1

Úvod

Přenos informace mezi lidmi při jazykové komunikaci je nesmírně zajímavý a složitý proces, jemuž se věnuje řada vědních odvětví. V této práci se zabýváme strukturou textu z pohledu textové lingvistiky, sledujeme, jak jsou jednotky v textu navzájem propojeny a jakými výrazovými prostředky se toto propojení signalizuje.

K zajištění spojitosti (koherence) textu slouží řada jazykových subsystémů, jako jsou např. koreferenční řetězce, vztahy asociační anafory, aktuální členění či vztah mezi otázkou a odpovědí. Tento výzkum je zaměřen na významové vztahy mezi vyššími jednotkami v textu (tzv. vztahy diskurzí, viz podkap. 2.4). Příkladem takového vztahu je jednoduché spojení vět *Musíme na nákup. Všechno jsme totiž snědli*. Mezi těmito větami je vztah příčiny, vyjádřený výrazem *totiž*. Obecněji řečeno, vyšší jednotky v textu jsou např. věty či vícevětne úseky; vztahy mezi nimi pak mívají různý význam. Jeden vyjadřuje např. exemplifikaci pro druhý, nebo opravu jeho tvrzení (rektifikace), podmínku, generalizaci apod.

V uvedeném příkladu je zajímavá role spojovacího výrazu *totiž*. Vidíme, že obecně informace sdělovaná pomocí jazyka má určitou strukturu. Existují signály, jimiž vyjadřujeme např. začátek a konec sdělování, ohraničení různých jednotek v naší řeči (intonací vymezujeme větu) a také vztahy mezi těmito jednotkami (např. pomocí koncovek a slovosledu vyjadřujeme větnou stavbu). Ne všechny signály ale musejí být nutně vždy přítomny. Některé se mohou v určitých kontextech vypouštět, a některé se z důvodů jazykové ekonomie vypouštějí dokonce pravidelně (např. na otázky neodpovídáme celou větou). To znamená, že určitou část jazykové struktury potom jako posluchači či čtenáři rekonstruujeme tak, že jsme schopni přicházející informaci porozumět.

Otázkou je, co všechno je možné z textu vypustit, aby byl text ještě srozumitelný. Kde leží ta hranice? V případě diskurzí vztahů vidíme, že není nutné každou dvojici vyšších jednotek spojit zvláštním výrazem, a přesto jsme schopni porozumět významovému vztahu mezi nimi. Kdybychom v příkladu seshora vypustili slovo *totiž*, stále bude zřejmý příčinný vztah mezi oběma větami: *Musíme na nákup. Všechno jsme snědli*. Na druhou stranu u jiných významových vztahů by vypuštění spojovacího výrazu mohlo vést k nepochopení: *Potřebuju se naučit řídit auto. Nepřihlásím se do autoškoly*. v tomto spojení vět je zjevný – a matoucí – rozpor, nejsme schopni jeho strukturu rekonstruovat, můžeme se proto domnívat, že jsme např. špatně slyšeli. Teprve po doplnění spojovacího výrazu *přesto* (*Potřebuju se naučit řídit. Přesto se nepřihlásím do autoškoly*.) získáváme potvrzení, že mezi obsahy těchto vět je skutečně rozpor

(konkrétně vztah přípustky), nedošlo tedy k chybě v našem porozumění, naopak můžeme očekávat vysvětlení tohoto rozporu v okolním kontextu.

V této práci zkoumáme právě možnosti vypouštění spojovacích výrazů mezi vyššími jednotkami v textu, a to v češtině. Zabýváme se následujícími otázkami: ve kterých kontextech lze vypouštět spojovací výrazy a ve kterých nikoli? Existuje nějaká typická sémantika vztahu mezi vyššími jednotkami, kterou nedokážeme z kontextu rekonstruovat? Může vypouštění spojovacího prostředku narušit spojitost textu? Jsou tato narušení srovnatelná, nebo některá vadí pro porozumění víc a některá méně?

1.1 Koherence textu

Tyto otázky se opakovaně vracejí k pojmu *spojitost* neboli *koherence* textu. Koherence bývá považována za jeden ze základních znaků textu (de Beaugrande – Dressler 1981), rozumí se jí vnitřní tematická a funkční soudržnost. Koherence se dosahuje pomocí tzv. *kohezních prostředků*, tedy konkrétních výrazových prostředků, které text propojují. Patří k nim např. spojovací výrazy, zájmena, opakování týchž slov, slovosledné modifikace, v některých jazycích určité a neurčité členy. Je na místě položit si otázku, zda a proč je koherence textu vlastně důležitá. Proč bychom se nemohli vyjadřovat nespojitě, nahodile? Je koherence skutečně nutným rysem textu? Někdy se vyjadřujeme nekoherentně, např. v krátkých neformálních nepřipravených projevech, kde teprve hledáme svoji myšlenku. Naši partneři v řeči si pak dodatečně ujasňují, co přesně máme na mysli, a my sami cítíme potřebu dovysvětlit, jak jsme to vlastně mysleli. Tento proces ukazuje na fakt, že jak my, tak naši partneři chápeme text jako předání určitého smyslu. Proces upřesňování při neporozumění pak ukazuje, že nedostatek koherence může narušit předání smyslu, tj. vlastní funkci textu. Koherence je tedy podstatná, napomáhá recipientovi pochopit smysl textu. A mluvčímu zase pomáhá jasně si formulovat a uspořádat myšlenky.

U textu předem očekáváme, že bude mít smysl,¹ a také u něj předpokládáme koherenci. To je důvod, proč to, co slyšíme, interpretujeme přednostně jako koherentní. Při možnosti více výkladů chápeme to, co k nám přichází, takovým způsobem, aby to pro nás dobře navazovalo na předchozí text. Naopak jako mluvčí jsme naučení podávat signály koherence a někdy to děláme ze zvyku dokonce i nadbytečně (srov. systematický pokles spojovacích výrazů v částice, např. u slova *takže* na počátku promluvy).

Výzkum vypouštění spojovacích prostředků mezi vyššími jednotkami v textu pomáhá popsat, kde může docházet k nežádoucímu porušení koherence textu a kde naopak jsme schopni smysl dobře rekonstruovat. Jeho výsledky pak mají dvojí uplatnění. První, praktické, je více nasnadě: takovýto výzkum poskytuje podklady pro různé druhy úpravy textu. Např. v automatické sumarizaci může ukázat, které spojovací výrazy a na kterých místech je dobré zachovat, aby nebyl narušen smysl zkrácené-

¹ Výjimečně se setkáváme s texty beze smyslu. Někdy nemají smysl právě pro nás (text v jazyce, kterému nerozumíme), jindy nemají smysl obecně a slouží např. pouze jako zvuková či grafická výplň (slabiky beze smyslu v textu písni nebo psaný pseudolatinský text *lorem ipsum* využívaný pro grafický design).

ho textu. Podobně jej lze využít u překladu, kde opět dává vodítko, které spojovací výrazy mají podstatný význam, jenž by měl být převeden do cílové řeči.

Druhé uplatnění je teoretické. Odpovědi na otázky uvedené výše pomáhají lingvistům ujasnit si představu o přesnosti jazyka. Přirozený jazyk totiž není a ani nepotřebuje být dokonale přesný a je dobré znát přibližné meze, jaká míra nepřesnosti je už přílišná, jaká je přijatelná a běžná a naopak nakolik přespříliš vysoká míra přesnosti je třeba až na obtíž porozumění. Při sémantickém popisu jazyka je totiž potřeba počítat také s tím, že význam jistého úseku textu nemusí obsahovat všechny (byť nevyjádřené) podrobnosti, které by obsahovat mohl.²

V naší práci zkoumáme možnosti vypouštění spojovacích prostředků dvěma způsoby, jednak analyzujeme data existujících českých textů v jazykovém korpusu PDiT-EDA 1.0, jednak jsme vytvořili pečlivě konstruované texty, které jsme předkládali respondentům v psycholinguvistických experimentech, jež umožňují osvětlit vliv vyjádřenosti spojovacích výrazů na porozumění jazyku. V teoretické části definujeme základní pojmy pro výzkum diskurzních vztahů (kap. 2), uvádíme zde přehled vědeckého zkoumání výstavby textu v české a světové lingvistice (kap. 3) a popisujeme naši metodologii a dále data a jejich zdroje (kap. 4).

V analytické části potom v kap. 5 shrnujeme výsledky pozorování na datech jazykového korpusu PDiT-EDA 1.0: zabýváme se zde anotátorskou shodou a počty různých druhů vztahů, které byly v tomto korpusu zachyceny. Kap. 5.2 u těchto dat podrobně rozebírá souvislost mezi významem vztahů mezi vyššími jednotkami v textu a možností neužívat u nich spojovacího výrazu. Vybrané závěry této kapitoly jsou pak ověřeny v psycholinguvistických experimentech v kap. 6, zaměřených na případnou nesrozumitelnost některých vztahů, u nichž není přítomen spojovací výraz. V příloze jsou připojeny podklady, které byly použity v psycholinguvistických experimentech.

² Srov. Štícha 2007, Hoffmannová 2008.

I

Teoretické východisko

V části Teoretické východisko se nejprve zabýváme základními pojmy, které se vztahují k tématu diskurzních vztahů (kap. 2). Následně popisujeme kontext výzkumu diskurzních vztahů: zachycujeme vybrané otázky stavby textu, jimž se věnovala či věnuje zahraniční a česká lingvistika (kap. 3). Přitom se soustředíme zejména na vědecké okruhy, které jsou spojeny s korpusovou lingvistikou a s psycholingvistikou. V další části (kap. 4) jsou představeny dva hlavní zdroje našeho přístupu, totiž funkční generativní popis a lexikální stromová gramatika (Lexical Tree-Adjoining Grammar), jak je aplikována v Penn Discourse Treebank. Na jejich základě vzniklo anotační schéma pražských diskurzních korpusů, které je popsáno v podkap. 4.3. V podkap. 4.4 je pak popsán způsob anotace implicitních diskurzních vztahů v korpusu PDiT-EDA 1.0, který sloužil jako základ pro náš výzkum.

2

Základní pojmy

2.1 Text a diskurz

Význam termínů *text* a *diskurz* se v lingvistice do značné míry překrývá. V tradiční evropské lingvistice (a zejména lingvistice německy psané) se při popisu soudržnosti jazykového útvaru používá spíše termínu *text* – srovnejme pojem *textová lingvistika* (*Textlinguistik*), jehož užívá W. Dressler (Dressler 1972); v české tradici pak např. spojení *textová syntax*, užívané J. Hrbáčkem (Hrbáček 1994).

Vedle tohoto významu se slova *text* používá také v kontextu dalších věd a oborů, jako je např. literární věda, textologie, drama, překladatelství ad. Pojem *text* často označuje jeden komunikát, a to psaný; obvykle u něj předpokládáme jednoho mluvčího a jednu hlavní komunikační funkci (Karlík et al. 2016b, s. 1866–1868).

Termín *diskurz* se v souvislosti s rysy soudržnosti jazykového útvaru užívá spíše v americké lingvistické tradici, srov. např. Harrisův termín *discourse analysis* (Harris 1952). Užívá se také v korpusové a počítačové lingvistice zabývající se koherencí, viz zejména Penn Discourse Treebank (PDTB; Prasad et al. 2008a) jako jeden z prvních korpusů zaměřených výhradně na problematiku koherence. Po jeho vzoru vznikla řada dalších korpusů, které leckdy v terminologii či přímo v názvu navazují na PDTB, například Turkish Discourse Treebank (Zeyrek et al. 2013) či Chinese Discourse Treebank (Zhou et al. 2014). Protože pro naši práci představuje Penn Discourse Treebank jeden z důležitých inspiračních zdrojů, navazujeme také na jeho názvosloví a přejímáme pojem *diskurz* jak do názvu našich korpusů (Prague Discourse Treebank, M. Rysová et al. 2016), tak i do anglicky i česky psaných prací, které ve svém výzkumu publikujeme. Hovoříme proto o *diskurzních vztazích*, *diskurzních konektorech*, *diskurzních argumentech* apod. Neznamena to však, že bychom popisovali jevy, jimiž se jazykovědná bohemistika doposud nezabývala; zpracování různých aspektů této problematiky má v bohemistice bohatou tradici, jen s užitím jiné terminologie (viz podkap. 3.3).

Pojem *diskurz* se obecně však užívá také v jiných vědeckých souvislostech. Může být spojován s dialogičností, někdy s mluvenými jazykovými projevy. (Rozlišuje se pak *text* jako psaná forma a *diskurz* jako forma mluvená.) *Kritická analýza diskurzu* (Fairclough 1995) chápe diskurz jako sociální jev, jde o ustálený způsob vypovídání v jistém společenském okruhu nebo o jistém tématu (odborný diskurz, mediální diskurz o migraci Romů; viz Karlík et al. 2016a, s. 350–353).

Celkově při užívání pojmu *diskurz* vystupují tři podstatné významové rysy, z nichž každý je zdůrazněn v jiném oboru. Prvním z těchto významů jsou „strukturální nad-

větné jevy projevující se zejména v soudržnosti jazykových projevů“, druhým „jazyk v konkrétním užití, jako projev parole“, třetím pak „pragmatické aspekty užívání jazyka“ (srov. Schiffrin 1994, s. 23–43 a Karlík et al. 2016a, s. 351).

Náš výzkum patří do prvního okruhu: zabýváme se strukturou jazykových projevů v nadvětné perspektivě. Pro tento výzkum není ani tak podstatný sám pojem *diskurz* jako spojení *diskurz ní vztahy* (*discourse relations*). Právě ty nás zajímají jako jeden ze způsobů, jak zajistit koherenci textu, vedle koreferenčních řetězců, aktuálního členění apod. V neterminologickém, obecném smyslu užíváme v této práci také slova *text*.

2.2 Diskurz ní argument

V kapitole 1 jsme uvedli příklad diskurz ní vztahu s významem příčiny: *Musíme na nákup. Všechno jsme totiž snědli*. Diskurz ní vztah je relací mezi jistými jednotkami v textu, tzv. *diskurz ními argumenty*; v uvedeném příkladě jsou to dvě věty jednoduché. Po významové stránce diskurz ní argument vyjadřuje jistý samostatný ucelený děj, stav, připsání vlastnosti apod., je tedy vyjádřením propozice (MČ 3, s. 9).¹ Jádrovou formou diskurz ního argumentu, s níž pracujeme, je významově samostatná propozice vyjádřená větou (klauzí) s určitým slovesným tvarem. Po formální stránce může být taková propozice realizována jako věta jednoduchá nebo jako věta závislá, pokud je její význam relativně nezávislý na obsahu věty řídící (tj. jako diskurz ní argument pojmáme např. větu vedlejší podmínkovou, nikoli ale větu vedlejší předmětnou, jež je součástí valenčního rámce přísudku ve větě řídící). Diskurz ní argument může mít také složitější stavbu, viz níže.

Propozice může být vyjádřena také nevětnou formou (např. *babiččin příchod*), potom je však její zapojení do textu řízeno především syntaktickými pravidly, prostředky nadvětné koherence se zde mohou uplatnit jen okrajově. Nevětné vyjádření propozice jsme proto v tomto výzkumu ponechali stranou, nepovažujeme je za diskurz ní argument.

Někdy bývá v textu část věty vypuštěna (tzv. elipsa), je však možné ji rekonstruovat. My jsme pracovali s daty Pražského diskurz ního korpusu, kde jsou elipsy rekonstruovány podle jednotných pravidel. Pokud tedy byla eliptická konstrukce rekonstruována jako celá věta, pracovali jsme s ní jako s úplným diskurz ní argumentem.

Pražský diskurz ní korpus obsahuje psané texty (novinové články), při vymezování diskurz ní argumentů pro sledování nadvětných jevů jsme proto jako vodičko mohli využít také grafického členění vět, zejména koncových interpunkčních znamének. Tato znaménka oddělují nejen věty obsahující určitý slovesný tvar, ale také věty jednoduchenné neslovesné (např. název článku, jméno autora). V anotaci implicitních diskurz ní vztahů jsme těmto strukturám věnovali pozornost: většinou jsme je označovali odlišnými značkami, aby byla jasná jejich role ve stavbě textu (např. právě název článku). Na druhou stranu v některých (velmi ojedinělých) případech bylo zřejmé, že

¹ Penn Discourse Treebank (Prasad et al. 2007) vymezuje diskurz ní argument podobně, jako tzv. „abstract object“ dle N. Ashera (Asher 1993).

i neslovesné jednočlenné věty v uvedeném kontextu vstupují do diskurzních vztahů; v tom případě jsme je zařadili mezi diskurzní argumenty.²

Do diskurzních vztahů vstupují nejen jednoduché jednotky, jako je např. větná klauze s jednou vyjádřenou predikací, ale také jednotky sdružené do většího argumentu, mezi nimiž existují další koherenční (např. diskurzní) vztahy. Diskurzním argumentem tedy může být vedle věty jednoduché či jedné věty v souvětí také souvětí jako celek, několik vět za sebou, celý odstavec apod.

Shrňme tedy ještě jednou typické formy diskurzních argumentů:

- věta jednoduchá slovesná
- závislá větná klauze, pokud má samostatný význam vzhledem k větě řídící
- věta jednoduchá neslovesná
- kombinace uvedených forem (souvětí souřadné, několik vět za sebou, souvětí spolu s částí následujícího souvětí apod.)

Při vymezování diskurzních argumentů v textu jsme uplatňovali tzv. princip minimality, tj. zejména u souvětí se složitější syntaktickou stavbou jsme přihlíželi k tomu, abychom do diskurzního argumentu vřazovali pouze ty jeho části, které skutečně po významové stránce vstupují do diskurzního vztahu. Bylo tedy možné podle kontextu nezařadit do diskurzního vztahu např. vsuvky, některé věty závislé či polovinu souvětí souřadného.

2.3 Diskurzní konektor

Diskurzní konektor je spojovací výraz, který formálně spojuje diskurzní argumenty v diskurzním vztahu. V ilustračním dokladu *Musíme na nákup. Všechno jsme totiž snědli*. je jím výraz *totiž*. Jako diskurzní konektory slouží výrazy s různou slovnědruhovou příslušností (srov. Zikánová et al. 2015a, s. 15–17):

- spojky podřadící (např. *protože, pokud, když*)
- spojky souřadící (např. *a, ale, takže, proto*)
- částice (*zkrátka, například*)
- vybraná příslovce (*potom, současně*)
- prostředky uvozující položky výčtu (*za první, za druhé; 1., 2.; a, b*)
- interpunkční znaménka (středník, dvojtečka)

Vedle jednoslovných, tzv. *primárních diskurzních konektorů* existují také *diskurzní konektory sekundární, víceslovné*, např. *pod podmínkou, že; z tohoto důvodu* apod. Jejich význam bývá synonymní s významem konektorů primárních (srov. *proto – z toho dů-*

² Jde zejména o vztahy v různých přehledech a ustálených textových žánrech, pro něž je běžné zkratkovitě vyjadřování. Srov. vztah konjunkce mezi větami *V 1 000 m na horách kolem 4 st. C. Slabý proměnlivý vítr;* podobně vztah konjunkce mezi větami *Inscenace Petr Léb, scéna WN, kostýmy Kateřina Štefková, dramaturgie Vlasta Smoláková, pohybová spolupráce Antonín Hodek, hudební spolupráce Miki Jelínek. Premiéra 22. února.* (PDiT-EDA 1.0)

vodu), i když rozvíetí v sekundárním konektoru umožňují význam zpřesňovat (z *tohoto jediného důvodu*).³

Zatímco některé výrazy jsou diskurzním konektorem pravidelně (např. *zatímco*), jiné mohou mít funkci diskurzního konektoru jen v některých kontextech (např. *současně*). Kritériem pro odlišení diskurzního konektoru je schopnost výrazu v daném kontextu otevřít pozice pro dva diskurzní argumenty.

Na druhou stranu fakt, že určitý spojovací výraz otevírá dvě pozice, ještě neznamená, že je v daném kontextu diskurzním konektorem. Důležité totiž také je, aby spojoval skutečné diskurzní argumenty, tj. vyjádření propozic vymezené výše. Toto rozlišení je nutné provádět zejména u spojek: pokud spojka vyjadřuje vztah mezi dvěma diskurzními argumenty, považujeme ji za diskurzní konektor (např. *Petr přišel a Pavel odešel.*). Pokud však spojuje pouze větné členy nevyjadřující propozice, za konektor ji nepovažujeme (*Petr a Pavel*).

Diskurzní konektor má svůj vlastní lexikální význam, který ukazuje na význam vztahu mezi diskurzními argumenty. Význam diskurzního vztahu však vyplývá také z obsahů diskurzních argumentů a z předchozího kontextu. Může proto dojít k situaci, kdy např. diskurzní konektor vyjadřuje význam diskurzního vztahu pouze vágně a jeho zpřesnění vyvozujeme ze vztahu mezi obsahy argumentů (srov. spojka *a* v odporovacím spojení: *Potřebuju si něco nechat vytisknout, a teď mají v copy centru zavřeno.*) Jindy je konektor nezbytný, abychom dokázali vztah mezi diskurzními argumenty správně pochopit, srov. příklad přípustky z úvodní kapitoly: *Potřebuju se naučit řídit auto. Přesto se nepřihlásím do autoškoly.* A v dalších případech jsme schopni význam diskurzního vztahu vyvodit i bez konektoru, např. příčinný vztah ve spojení *Vyplavilo nám to sklep, hrozně přšelo.*

2.4 Diskurzní vztah

Diskurzním vztahem rozumíme významovou vazbu mezi dvěma ucelenými diskurzními argumenty. Sémantické typy diskurzních vztahů se v různých klasifikacích liší; klasifikace užívaná v Pražském diskurzním korpusu je popsána v části 4.3.1.

Formálně se diskurzní vztahy vyjadřují zejména pomocí diskurzních konektorů – tuto formu diskurzních vztahů nazýváme *explicitní*. Pokud vztah není vyjádřen diskurzním konektorem, ať už primárním, či sekundárním, avšak do vztahu je možné (při případné malé reformulaci) konektor dosadit, aniž se tím zkrslí význam textové návaznosti, má diskurzní vztah formu *implicitní*. Srov. např. implicitní vztah rektifikace ve dvojici argumentů *Nešel do obchodu, (nůbrž) zůstal doma.*⁴

³ Podrobněji viz M. Rysová 2018.

⁴ Implicitní diskurzní vztah je třeba odlišovat od implicitního argumentu. Implicitní diskurzní vztah má nevyjádřený diskurzní konektor, ale jeho argumenty jsou vyjádřené. Naproti tomu implicitní diskurzní argument představuje myšlenku, s níž se v textu pracuje, avšak přímo vyjádřená není, např. ve spojení [*Myslím, že*] *Šel do práce, protože nemá před domem auto.*

Diskurzní vztahy je třeba odlišit od dvou velkých skupin vztahů, totiž na jedné straně od vztahů čistě syntaktických, na druhé straně od ostatních vztahů koherenčních. Ke vztahům čistě syntaktickým, které neřadíme mezi vztahy diskurzní, patří vztah věty vedlejší podmětnej, předmětnej, přísudkovej, zpravidla přívlastkovej a příslovecnej k větě řídící. Nepovažujeme je za diskurzní argumenty, protože jejich sémantika nebývá ucelená, nezávislá; často naopak věta závislá je částí valenčního rámce členu věty řídící. Někdy může být sémantika věty závislé samostatná, potom jde o závislost formální, avšak nikoli významovou. Tak tomu bývá v případě některých vět příslovecných (nevalenční věty vedlejší časové, dále věty vedlejší účelové, podmínkové, přípustkové a příčinné) a nepravých vedlejších vět vztazných.⁵ Tyto případy hodnotíme jako diskurzní argumenty vstupující do diskurzního vztahu s větou řídící.

Druhou, poměrně různorodou skupinou vztahů, které neřadíme ke vztahům diskurzním, jsou různé další vztahy koherenční. Patří sem např. vztahy koreferenční (Zikánová et al. 2015a), vztah mezi otázkou a odpovědí, aktuální členění věty (jde o součást vlastního významu jedné věty) a další. Na rozdíl od těchto vztahů má diskurzní vztah nezávislé diskurzní argumenty, které jsme charakterizovali výše, a lze u něj rozpoznat některý z významů sémantické klasifikace. Jedním z rysů, který odlišuje diskurzní vztahy od různých jiných koherenčních vztahů, je také celistvost jeho argumentů. Diskurzní vztahy jsou vztahem celé propozice (případně několika) k celé jiné propozici (několika propozicím); za diskurzní vztah tedy nepovažujeme např. vztah mezi jmennou skupinou v jedné větě a druhou větou.⁶

Diskurzní vztahy, s nimiž pracujeme v Pražském diskurzním korpusu, vymezujeme jako vztahy binární, tj. vztahy právě mezi dvěma argumenty. Některé vztahy v textu by bylo možné interpretovat také tak, že mají více argumentů než dva. Jde např. o některá vícečlenná konjunktivní spojení, jako je *Ryba plave, pták létá a člověk chodí*. Také typická struktura seznamu má více částí než jen dvě: tvoří ji propozice s významem nadřazeným, za níž následuje několik propozic s významem podřazeným (např. *O prázdninách jsem vyzkoušela několik nových sportů. Jednak jsem se přihlásila na kurz badmintonu, jednak jsme v oddíle začali dělat kraula, a pak mě kamarádka vzala na spinning.*) Tyto struktury popisujeme jako soustavu binárních vztahů, rozkládáme je tedy na dvojice argumentů.

⁵ Příkladem takového spojení je souvětí *Nastoupil jsem do soukromé firmy, kde jsem dosáhl značných úspěchů*. Druhá věta zde má samostatný obsah, který vyjadřuje děj časově následující po ději věty předchozí (formálně řídící).

⁶ Příkladem vztahu, který neřadíme mezi vztahy diskurzní, je vazba s významem specifikace v následující dvojici vět: *Mluvil o mnoha problémech. Děti se špatně učí, ...* Druhá věta tu není specifikací celé situace popsané v první větě (tj. že mluvil o mnoha problémech), nýbrž specifikuje pouze povahu těchto problémů, tj. pouze jeden člen první věty jako možného diskurzního argumentu.

3

Kontext výzkumu

Výstavba textu je předmětem zájmu řady lingvistických přístupů. Ve světové lingvistice se od pol. 20. století, kdy byly na toto téma publikovány základní teoretické práce (např. Halliday – Hasan 1976; de Beaugrande – Dressler 1981; van Dijk 1980), rozvíjí výzkum stavby textu v různých směrech, jež se zaměřují např. na formální strukturu textu, na pragmatické aspekty výstavby textu, na vztah mezi textem a společností (sociologie, sociolingvistika) ad.

3.1 Teoretické rámce formálního popisu diskurzních vztahů

Také v rámci výzkumu textové koherence a jejích formálních prostředků se vyvinula řada vědeckých přístupů. Poměrně rozsáhlý tematický okruh představují práce zaměřené na výzkum diskurzních konektorů a dalších prostředků organizujících text (např. Fedriani – Sansò 2017; Fischer 2006, Degand et al. 2013).

Dalším okruhem jsou pak ucelené teorie představující výstavbu textu jako celku; některé z nich se uplatnily při vytváření elektronických korpusů zaměřených na výstavbu textu. Např. teorie rétorické struktury (Rhetorical Structure Theory, Mann – Thompson 1988) popisuje text jako jednu souvislou stromovou strukturu; tento přístup byl použit v jednom z prvních korpusů zaměřených na stavbu textu, v korpusu RST-Treebank (Carlson et al. 2001). Volnější model stavby textu představuje projekt Discourse Graphbank (Wolf – Gibson 2005), který se zaměřuje také na složitější jevy, jako jsou vztahy mezi vzdálenými jednotkami, překrývání jednotek, zahrnutí jedné jednotky do druhé apod. Významným projektem zaměřeným na textovou výstavbu je také Penn Discourse Treebank (PDTB 2.0; Prasad et al. 2008a), který je založen na lexikálním popisu diskurzních konektorů. Metodologie korpusu Penn Discourse Treebank se stala východiskem pro anotace diskurzních vztahů v řadě dalších jazyků vedle angličtiny: s určitými modifikacemi ji převzal Pražský diskurzní korpus (Poláková et al. 2012a) pro češtinu, dále byla uplatněna pro čínštinu (Zhou – Xue 2012), arabštinu (Al-Saif – Markert 2010), němčinu (Stede – Neumann 2014) a další.

V poslední době vznikly také dva návrhy na unifikaci přístupu k zachycování diskurzních vztahů, které by v důsledku umožnily např. porovnávat a navzájem převádět data různých elektronických korpusů, jak v rámci jednoho jazyka, tak i mezijazykově (Sanders et al. 2018, Bunt – Prasad 2016).

3.2 Implicitní diskurzní vztahy a elektronické korpusy

V mezích různých teoretických rámců se předmětem výzkumu korpusové lingvistiky staly také implicitní diskurzní vztahy. Výstupy těchto výzkumů se liší, což je dáno odlišnými přístupy i rozdílností vstupních dat.

Jeden z prvních pokusů o anotaci implicitních diskurzních vztahů v elektronickém korpusu byl proveden v projektu Penn Discourse Treebank 2.0. Anotátoři zde měli za úkol mezi sousední věty, které nebyly explicitně propojeny diskurzním konektorem, vždy příslušný konektor vložit a anotovat sémantiku vztahu. Z analýzy anotace vyplývá, že distribuce explicitních a implicitních vztahů v korpusu je poměrně vyrovnaná (Prasad et al. 2008b).

Podobné výsledky u obou typů diskurzních vztahů vykazuje také Hindi Discourse Treebank (Oza et al. 2009), zatímco v korpusu Biomedical Discourse Relation Bank distribuce implicitních vztahů mírně převažuje nad vztahy explicitními (v poměru 6:5; Prasad et al. 2011). Opačný poměr těchto vztahů je zaznamenán v korpusu LUNA, zaměřeném na mluvené dialogy v italštině, v němž explicitní vztahy výrazně převažují nad vztahy implicitními (Tonelli et al. 2010, s. 2089). Značkování implicitních vztahů ve všech sledovaných korpusech dosahuje poměrně nízké mezianotátorské shody, která je měřena obvykle pouze pomocí hodnoty Cohenovy kappy.

Anotace implicitních diskurzních vztahů se využívá jako podklad pro další automatické zpracování jazyka, srov. např. využití dat korpusu Penn Discourse Treebank pro experimenty s automatickou predikcí významu implicitních diskurzních vztahů (Pitler et al. 2009, Kim et al. 2020, Braud – Denis 2014). V současné době pak znamená prudký rozvoj automatické vyhledávání implicitních diskurzních vztahů s pomocí počítačových neuronových sítí (např. Dai – Huang 2018).

Zároveň se dosavadní anotace implicitních vztahů (nazvěme je prvními verzemi) využívají také pro zlepšení popisu teoretického jazyka obecně, tj. pro adekvátnější modely, srov. např. nedávno postulovanou tezi o koexistenci explicitních a implicitních vztahů mezi týmiž argumenty, tj. o možnosti, že dva diskurzní argumenty jsou mezi sebou spojeny více než jedním diskurzním vztahem (Webber et al. 2018; Rohde et al. 2018). Analyzuje se také vztah mezi implicitností a sémantickými typy diskurzních vztahů (Torabi Asr – Demberg 2012, van der Vliet – Redeker 2014, Speyer – Fetzer 2014, Hoek et al. 2015, Hoek – Zufferey 2015) či přídatné signály, na jejichž základě lze vysoudit existenci a význam implicitního vztahu (Poláková et al. 2017, Taboada 2009, Das – Taboada 2013).

3.3 Čeština a výstavba textu

Česká (československá) jazykověda se věnuje analýze různých aspektů výstavby textu již od počátku 20. století, přičemž podrobně je zpracována zejména stránka formálně strukturní. Velké pozornosti se trvale těší výzkum aktuálního členění věty, který u nás zahájil V. Mathesius (Mathesius 1947) a který dále rozvinuli jeho následovníci

zejména v Brně (Firbas 1992, A. Svoboda 1984) a v Praze (Sgall et al. 1980, Sgall et al. 1986, Hajičová 1993, Uhlířová 1987). Význačným tématem české lingvistiky jsou také tematické posloupnosti popsané F. Danešem (Daneš 1968), který analyzoval vztah mezi výstavbou textu a syntaxí; vlastní ucelený přístup k textové syntaxi pak navrhl J. Hrbáček (Hrbáček 1994). Vztahem mezi stavbou věty a textem se zabývá také J. Tárníková (Tárníková 2012). Dalším polem výzkumu jsou vzájemné vztahy mezi výstavbou textu a stylistikou, představené např. v pracích K. Hausenblase, J. Mistríka či J. Hoffmannové (Hausenblas 1971, Mistrík 1966, Mistrík 1989, Hoffmannová et al. 2016) či mezi výstavbou textu a pragmatikou (Hirschová 2017).

Dlouhou tradici má v bohemistice také výzkum spojovacích výrazů v širokém smyslu (od spojek přes částice po diskurzní markery) z různých pohledů. K těmto přístupům patří např. syntax, lexikologie, etymologie, přístup diachronní, kontrastivní ad. (srov. K. Svoboda 1972, Kopečný et al. 1980, Pešek 2011, Kosek 2003, Čermáková et al. 2019).

Naše práce vznikla v rámci výzkumu větné sémantiky a koherenčních vztahů, který dlouhodobě probíhá v Ústavu formální a aplikované lingvistiky na Matematicko-fyzikální fakultě Univerzity Karlovy. Zdejší výzkumná skupina P. Sgalla vytvořila jako teoretický rámec tzv. *funkční generativní popis jazyka* (Sgall et al., 1969; Sgall et al. 1986, viz podkap. 4.2), na jehož základě byl posléze vybudován Pražský závislostní korpus (v několika verzích, 1995–2013). Předmětem výzkumu založeného na korpusech se pak – vedle dalších témat – stalo aktuální členění věty v češtině (Hajičová 2012, Zikánová – Týnovský 2009), koreference a asociační anafora (Nedoluzhko 2011) a diskurzní vztahy (Zikánová 2015a). Výzkum diskurzních vztahů byl v první fázi zaměřen na vztahy explicitní (Zikánová et al. 2010, Mírovský et al. 2014), hledaly se také způsoby, jak náležitě měřit mezianotátorskou shodu (Mírovský et al. 2010). Zvláštní pozornost byla věnována diskurzním konektorům primárním i sekundárním po teoretické stránce (M. Rysová 2018), později vznikl jednak elektronický slovník českých diskurzních konektorů (CzeDLex; Mírovský et al. 2017, M. Rysová et al. 2018), jednak slovník českých a německých konektorů s anaforickým elementem (GeCzLex; Poláková et al. 2020). S tématem konektorů souvisí také určení jejich funkcí a rozlišení konektorů od fokalizátorů (rematizátorů; Hajičová et al. 2020) a dále jejich možná významová nespecifičnost (Crible et al. 2019). Dalším okruhem výzkumu je výstavba textu a realizace diskurzních vztahů v textech cizinců, kteří si osvojují češtinu, v jehož rámci byl vytvořen také elektronický evaluační nástroj EVALD, který hodnotí koherenční stránku jazyka (K. Rysová et al. 2016, Novák et al. 2019). Výzkum diskurzních vztahů v češtině se v Ústavu formální a aplikované lingvistiky rozvíjí i dalšími směry: vedle analýzy vztahů implicitních, představené v této práci a navazující na předchozí průzkumné sondy (Poláková 2014, Zikánová et al. 2019), se nově zkoumají také vztahy makrostruktury, tj. vztahy mezi rozsáhlejšími úseky v textu (Poláková et al. 2021).

3.4 Výstavba textu v psycholingvistickém výzkumu

Současná psycholingvistika objasňuje řadu aspektů jazykové struktury a jejího fungování v komunikaci (Ferrnándezová – Smithová Cairnssová 2014, Eysenck – Keane 2015, Harley 2014). Často se zaměřuje zejména na osvojování jazyka a na řečové poruchy, její metody se však využívají jako podpůrný prostředek v mnoha dalších odvětvích jazykovědy. Při výzkumu výstavby textu se používají psycholingvistické experimenty např. pro testování porozumění diskurzním konektorům (výzkumná skupina na Utrechtské univerzitě pod vedením Teda Sanderse, srov. Zufferey et al. 2015), při testování namáhavosti zpracování diskurzních konektorů při vnímání textu (výzkumný tým na Heidelberské univerzitě vedený Oscarem Louredou, viz Loureda et al. 2015) či pro ověření souvislostí mezi syntaktickou formou a diskurzními vztahy (viz např. Hoek et al. 2021, o roli vztažných vět jako diskurzních argumentů).

V české lingvistice se psycholingvistickým experimentům věnuje např. laboratoř LABELS (společné pracoviště Psychologického ústavu Akademie věd ČR a Filozofické fakulty Univerzity Karlovy), kde proběhly mj. pokusy zaměřené na český slovosled a aktuální členění (Smolík – Lukavský 2009, Smolík 2015). Pomocí psycholingvistických experimentů se v současné době v češtině zkoumá také např. obecná srozumitelnost textů, a to v odborném a administrativním stylu (Chromý et al. 2021, Bendová – Cinková, v tisku).

V naší práci propojujeme několik přístupů dohromady. Po teoretické stránce vycházíme z rámce amerického korpusu Penn Discourse Treebank (Prasad et al. 2008a) a českého funkčního generativního popisu (Sgall et al. 1986); pokud jde o metody a data, navazujeme na Pražský závislostní korpus (Bejček et al. 2013) a pro češtinu nově využíváme také psycholingvistického přístupu. Blíže je náš postup popsán v podkap. 4.3, psycholingvistické experimenty pak v kap. 6.

4

Koncepce výzkumu diskurzních vztahů

Analýza implicitních diskurzních vztahů v češtině, jíž se věnuje tato práce, navazuje na rozsáhlý výzkum diskurzních vztahů, který probíhá v Ústavu formální a aplikované lingvistiky (Matematicko-fyzikální fakulta Univerzity Karlovy).¹ Tyto studie mají společný teoretický základ, který vychází ze dvou hlavních zdrojů: přístup amerického korpusu Penn Discourse Treebank (Prasad et al. 2008b) a český funkční generativní popis jazyka (Sgall et al. 1969; Sgall et al. 1986).

V této kapitole nejprve představíme obě východiska, totiž Penn Discourse Treebank a funkční generativní popis, jak je aplikován v Pražském závislostním korpusu (podkap. 4.1–4.2). Tato východiska byla zohledněna v konceptu anotace diskurzních vztahů užívaném v rodině korpusů Pražského závislostního korpusu (viz podkap. 4.3). Ve všech těchto přístupech věnujeme pozornost především významové klasifikaci vztahů.

4.1 Penn Discourse Treebank 2.0

Přístup korpusu Penn Discourse Treebank 2.0 (PDTB 2.0, Prasad et al. 2007) je založen na tzv. Lexicalized Tree-Adjoining Grammar (Webber – Joshi 1998), která chápe diskurzní konektory jako lexikální operátory spojující diskurzní argumenty do vyšších jednotek. Tento přístup je tedy lokální: vyhledává pro lexikální operátory argumenty v nejbližším okolí, a postupně, odzdoła, vytváří stavbu celého textu.

Diskurzní argument byl vymezen jako tzv. *abstraktní objekt* dle N. Ashera (Asher 1993), tj. vyjádřená propozice. Jako diskurzní argument byly nejprve chápány pouze klauze s určitými slovesnými tvary, v pozdější variantě korpusu (PDTB 3.0, Prasad et al. 2019) k nim byly přiřazeny také slovesné tvary neurčité (participia, infinitivy). Při vymezování rozsahu diskurzního argumentu byl uplatňován *princip minimality*, tj. součástí argumentu je minimum nutné pro pochopení diskurzního vztahu. Pokud může být pochopení zpřesněno na základě nějaké doplňkové informace uvedené v textu, je tato doplňková informace označena jako tzv. doplnění (supplement).

Pro vymezení, co je a není diskurzní vztah, bylo v PDTB 2.0 užito anotace druhů vztahů, která rozlišuje následující podtypy možných spojení mezi klauzemi (Prasad et al. 2007):

- explicitní diskurzní vztah
- implicitní diskurzní vztah

¹ Srov. např. přehledové práce Zikánová et al. 2015a, Zikánová et al. 2015b.

- alternativní lexikalizace diskurzního konektoru
- vztah mezi entitami
- neexistence vztahu

V průběhu anotace byly nejprve vyznačeny explicitní diskurzní vztahy, na základě přítomnosti konektoru. Dalším krokem pak byla anotace implicitních diskurzních vztahů, tedy míst, kde se vztah, potažmo konektor, pouze vyrozumívá. Při této anotaci se ukázalo, že zatímco do některých spojení vět lze bez problémů vsunout konektor a interpretovat význam celého spojení, na jiných místech to možné není. Dvojice vět, do nichž není možné vsunout konektor, pak byly rozděleny do tří kategorií: (a) mezi argumenty existuje diskurzní vztah, ale je vyjádřen jinými výrazy, doplněný konektor je tu nadbytečný (tzv. alternativní lexikalizace, např. *the reason is*); (b) mezi argumenty existuje vztah, vyskytuje se mezi dvěma entitami zmíněnými v obou argumentech; (c) mezi argumenty žádný vztah není. Všechny tyto typy vztahů jsou v korpusu vyznačeny, navíc je označována také tzv. atribuce, tj. vztah mezi autorskou a přímou řečí a jeho modifikace.

Texty v PDTB jsou tedy anotovány jako souvislé řetězce argumentů, kde každá dvojice sousedících argumentů uvnitř odstavců má mezi sebou vyznačený jistý vztah. Kromě toho jsou v rámci explicitních vztahů anotovány i vztahy mezi vzdálenými argumenty.² Implicitní vztahy byly vyznačovány pouze v rámci jednotlivých odstavců, nikoliv mezi odstavci. Anotace v PDTB je přitom zaměřena pouze na vztahy mezivětné (tj. vztahy mezi syntaktickými celky ukončenými koncovým interpunkčním znaménkem; výjimku tvoří vztahy mezi jednotkami oddělenými dvojtečkou a středníkem, které se řadí mezi vztahy diskurzní).³

Diskurzní vztahy v PDTB jsou chápány jako relace binární. Argumenty ve vztahu jsou označovány čísly. Značka argumentu, tedy Argument1 a Argument2 (zkráceně Arg1 a Arg2) se v PDTB 2.0 přiřazují na základě místa výskytu diskurzního konektoru: argument syntakticky připojený ke konektoru je označen jako Arg2, zbylý argument se označuje jako Arg1. U implicitních vztahů se postupuje analogicky, přičemž se argumenty číslovají podle pozice vsunutého (nevyjádřeného) konektoru.

Označování argumentů se tedy řídí jevem povrchové syntaxe (umístění konektoru), nikoli např. pořadím argumentů v textu, syntaktickými či významovými vztahy. Podle autorů této koncepce umožňuje tento přístup zpřístupnit data PDTB pro různé teoretické rámce, aniž by sám vnášel do popisu dat nadbytečné teoretické interpretace (Prasad et al. 2008b).

Anotace textů v PDTB probíhá přímo na textu jako lineárním řetězci, nevyužívá např. předběžné syntaktické analýzy. To znamená, že anotátoři mají za úkol samo-

² Ačkoliv se PDTB zaměřuje na vztahy lokální, je u některých diskurzních konektorů zřejmé, že nespojují přímo sousedící jednotky. V těchto případech bylo v anotaci povoleno vyznačování vztahů i mezi jednotkami vzdálenými.

³ Vztahy vnitrovětné byly zařazeny do anotace až v následující verzi korpusu, PDTB 3.0, Prasad et al. 2019, viz níže.

statně určit, jak přistupovat k některým syntakticky složitějším jevům, jako jsou např. elipsy, vsuvky, souřadné konstrukce apod.

4.1.1 Významová klasifikace vztahů v Penn Discourse Treebank 2.0

Diskurzní vztahy se pak rozlišují podle významu. Protože pro anotaci diskurzu v pražském korpusu byla významová klasifikace PDTB 2.0 zásadním inspiračním zdrojem, popisujeme ji v této části podrobněji.

Penn Discourse Treebank 2.0 používá významovou klasifikaci diskurzních vztahů s několika stupni obecnosti (Prasad et al. 2007, s. 26–37). Rozlišují se zde čtyři hlavní skupiny vztahů: vztahy časové (temporal), příčinné (contingency), srovnávací (comparison) a rozšiřující (expansion). V jejich rámci se pak rozeznávají další podtypy, z nichž některé mohou mít ještě jednu úroveň vnitřního členění. Úplná významová klasifikace diskurzních vztahů je uvedena v tab.4.1.

Třída časových vztahů zahrnuje vztahy, v nichž situace popsané v diskurzních argumentech jsou navzájem uspořádány časově, přičemž v této nejobecnější rovině není určena jejich současnost či předčasnost. Tato třída se dělí na dva podtypy: *současnost*, u níž konektor vyjadřuje, že obě situace probíhají zároveň, a *nesoučasnost*, při níž z významu konektoru vyplývá, že děje argumentů následují časově po sobě. Podle významu konektorů se pak nesoučasnost dělí na dva podtypy: u *předčasnosti* konektor vyjadřuje, že děj Arg1 předchází před dějem Arg2 (např. u spojky *before – než*), u *následnosti* je pak pořadí dějů argumentů obrácené, děj Arg1 následuje po ději Arg2 (spojka *after – až*).

U třídy příčinných vztahů vyplývá vždy z významu užitých konektorů, že jedna ze situací popsaných v Arg1 a Arg2 příčinně ovlivňuje druhou. Typ *příčina* zahrnuje případy, v nichž se děje Arg1 a Arg2 příčinně ovlivňují, ale jeden není podmínkou druhého. Směr příčinného vztahu není na této úrovni specifikován, je zachycen až v podtypech. U podtypu *vlastní příčina* vyplývá z významu konektoru, že situace v Arg2 je příčinou a situace v Arg1 následek (z popisu lze vyvodit, že takto funguje např. spojka *protože*). Naopak u podtypu *následek* jsou děje v argumentech uspořádány obráceně, situace v Arg2 je výsledkem situace popsané v Arg1 (např. spojka *so – takže*).

Zvláštní typ příčinných vztahů tvoří *pragmatická příčina*. Kategorie tzv. vztahů pragmatických (*pragmatická příčina*, *pragmatická podmínka*, *pragmatický kontrast*, *pragmatická přípustka*) byla v PDTB zavedena pro taková užití konektorů, v nichž nejde o význam spojení přímo obsahů jednotlivých argumentů, ale o spojení významů specifických užití těchto argumentů (Prasad et al. 2007, s. 27), zejména o zvláštní užití v řečových aktech. U typu *pragmatická příčina* s jediným podtypem *zdůvodnění* konektor vyjadřuje, že Arg1 obsahuje jisté tvrzení a Arg2 vyjadřuje zdůvodnění tohoto tvrzení (op. cit., s. 29); v Arg2 se tak např. odůvodňuje volba slov obsažených v Arg1. Typickým konektorem je spojka *because – protože*.

Třída vztahů	Typ	Podtyp
ČASOVÉ	Současnost	
	Nesoučasnost	Předčasnost Následnost
PŘÍČINNÉ	Příčina	Vlastní příčina Následek
	Pragmatická příčina	Zdůvodnění
	Podmínka	Hypotetická Obecná Reálná v přítomnosti Reálná v minulosti Ireálná v přítomnosti Ireálná v minulosti
	Pragmatická podmínka	Relevance Implicitní tvrzení
SROVNÁVACÍ	Kontrast	Juxtapozice Opozice
	Pragmatický kontrast	
	Přípustka	Očekávání Popření očekávání
	Pragmatická přípustka	
ROZŠÍŘUJÍCÍ	Konjunkce	
	Exemplifikace	
	Přeformulování	Specifikace Ekvivalence Generalizace
	Alternativa	Slučovací Vylučovací Vybraná
	Výjimka	
	Seznam	

Tabulka 4.1: Významová klasifikace diskurzních vztahů v PDTB 2.0

Dalším typem ve třídě příčinných vztahů je *podmínka*. Stejně jako u předchozích dvou typů tu existuje jistá příčinná souvislost mezi ději vyjádřenými v obou argumentech, zároveň ale je situace zachycená v Arg2 podmínkou a situace v Arg1 jejím následkem. Na rozdíl od předchozích dvou typů se u *podmínky* nelze vyslovit o pravdivostní hodnotě jednotlivých argumentů bez ohledu na konektor. PDTB 2.0 rozlišuje poměrně

podrobně jednotlivé podtypy podmínky. Protože se tato typologie podmínkových vztahů částečně liší od českých typologií, uvádíme ji s vysvětlením a s příklady z anotačního manuálu PDTB 2.0 (Prasad et al. 2007).

Prvním podtypem je *podmínka hypotetická*. Je charakterizována následovně: pokud platí Arg2, potom platí i Arg1 ve všech možných eventualitách v budoucnosti. Slovesa jsou obvykle ve tvaru přítomného nebo budoucího času.⁴

- (1) *Both sides have agreed* <Arg1: *that the talks will be most successful*>
 <Arg2: [*if* PODMÍNKA HYPOTETICKÁ] *negotiators start by focusing on the areas that can be most easily changed.*> (Prasad et al. 2007)

Obě strany se dohodly, <Arg1: *že rozhovory budou nejúspěšnější,*>

<Arg2: [*pokud* PODMÍNKA HYPOTETICKÁ] *vyjednávači začnou zaměřením na oblasti, které lze nejsnáze změnit.*>

Dalším vztahem je *podmínka obecná*. V tomto případě konektor vyjadřuje, že vždy, když je Arg2 pravdivý, je pravdivý i Arg1. Obvykle se tak vyjadřuje nějaká obecná pravda o světě, případně nějaký pravidelný následek podmínky. Od hypotetické podmínky se obecná podmínka liší tím, že hypotetická podmínka se vyjadřuje o vztahu, který proběhne jednou.

- (2) <Arg1: *They won't buy*>
 <Arg2: [*if* PODMÍNKA OBEČNÁ] *the quality is not there.*> (Prasad et al. 2007)
 <Arg1: *Nebudou to kupovat,*>
 <Arg2: [*pokud* PODMÍNKA OBEČNÁ] *to nebude kvalitní.*>

U *reálné podmínky v přítomnosti* konektor vyjadřuje, že situace v Arg2 je buď podávána jako reálný fakt v předchozím textu, nebo ji jako reálnou chápe někdo jiný než mluvčí.

- (3) *"I've heard that there is \$40 billion in nationwide boiler rooms every year," Mr. McClelland says.*
 "<Arg2: [*If* PODMÍNKA REÁLNÁ V PŘÍTOMNOSTI] *that's true,*>
 <Arg1: *Orange County has to be at least 10% of that.*>" (Prasad et al. 2007)
 „Slyšel jsem, že celostátně to u kotelen činí 40 miliard dolarů ročně," říká McClelland.
 „<Arg2: [*Pokud* PODMÍNKA REÁLNÁ V PŘÍTOMNOSTI] *je to pravda,*>
 <Arg1: *musí na tom oranžské hrabství mít podíl alespoň 10 %.*>"

U *podmínky reálné v minulosti* panuje mezi oběma ději stejný vztah jako v předchozím případě, pouze s tím rozdílem, že děj Arg2 je situován do minulosti. Arg1 pak vyjadřuje výsledek podmínky, stejně jako v předchozích případech.

⁴ V příkladech diskurzních vztahů značíme rozsah jednotlivých argumentů pomocí špičatých závorek, pro označení argumentů používáme zkratky Arg1 a Arg2. Diskurzni konektor je vtištěn tučně; sémantický typ diskurzního vztahu je vyznačen u diskurzního konektoru.

- (4) "<Arg2: [*If* PODMÍNKA REÁLNÁ V MINULOSTI] *they had this much trouble with Chicago & North Western,*>
 <Arg1: *they are going to have an awful time with the rest.*>" (Prasad et al. 2007)
 "<Arg2: [*Pokud* PODMÍNKA REÁLNÁ V MINULOSTI] *měli tolik problémů s Chicagem a Severozápadní provincií,*>
 <Arg1: *budou se s dalšími vysloveně trápit.*>"

Další dva podtypy jsou *podmínky ireálné*. U *podmínky ireálné v přítomnosti* konektor vyjadřuje, že situace popsaná v Arg2 buďto v přítomnosti neplatí, nebo ji někdo považuje za neplatnou.

- (5) *Of course,*
 <Arg2: [*if* PODMÍNKA IREÁLNÁ V PŘÍTOMNOSTI] *the film contained dialogue,*>
 <Arg1: *Mr. Lane's Artist would be called a homeless person.*> (Prasad et al. 2007)
Samozřejmě,
 <Arg2: [*pokud* PODMÍNKA IREÁLNÁ V PŘÍTOMNOSTI] *by ten film obsahoval dialog,*>
 <Arg1: *řikalo by se umělci pana Lanea bezdomovec.*>

Vztah *podmínky ireálné v minulosti* pak spojuje dva děje v podmínkovém vztahu, u nichž děj v Arg2 nenastal v minulosti a Arg1 vyjadřuje, jaký by byl výsledek, kdyby děj Arg2 nastal.

- (6) <Arg2: [*If* PODMÍNKA IREÁLNÁ V MINULOSTI] *I had come into Friday on margin or with very little cash in the portfolios,*>
 <Arg1: *I would not do any buying.*> (Prasad et al. 2007)
 <Arg2: [*Kdybych* PODMÍNKA IREÁLNÁ V MINULOSTI] *v pátek musel nakupovat na marži nebo bych měl málo hotovosti v portfoliích,*>
 <Arg1: *nedělal bych žádné nákupy.*>

Posledním typem příčinného vztahu je *pragmatická podmínka*. Jde o vztahy, v nichž se užívá konektoru typického pro vztahy podmínkové (anglicky *if – jestliže, pokud*), avšak Arg1 a Arg2 tu nejsou spojeny příčinně a pravdivost Arg1 není závislá na Arg2. U podtypu *relevance* vyjadřuje Arg2 podmínky, za nichž je tvrzení obsažené v Arg1 relevantní:

- (7) <Arg2: [*If* PODMÍNKA PRAGMATICKÁ] *you are thirsty,*>
 <Arg1: *there's beer in the fridge.*> (Prasad et al. 2007)
 <Arg2: [*Pokud* PODMÍNKA PRAGMATICKÁ] *budeš mít žízeň,*>
 <Arg1: *v lednici je pivo.*>

Speciální užití mívají také původně podmínkové konektory u podtypu *implicitní tvrzení*. Podle autorů tu jde o podmínkovou konstrukci využitou pro implicitní tvrzení, jeho vnitřní struktura se blíže nerozebírá. (Jde však zřejmě o konstrukci zřetelovou.)

- (8) <Arg2: [*If* IMPLICITNÍ TVRZENÍ] *you want to keep the crime rates high*,>
 <Arg1: *O'Connor is your man*.> (Prasad et al. 2007)
 <Arg2: [*Pokud* IMPLICITNÍ TVRZENÍ] *chcete udržet vysokou zločinnost*,>
 <Arg1: *je O'Connor váš člověk*.>

Další širokou třídu vztahů tvoří **vztahy srovnávací** (comparison). Autoři je charakterizují tak, že u těchto vztahů konektor po významové stránce zdůrazňuje rozdíly mezi situacemi v obou argumentech. U prvního typu vztahů, tzv. *kontrastu*, mají Arg1 a Arg2 jistý společný predikát či jiný výraz a v každém z argumentů se této společné jednotce připisuje jiná hodnota. U druhého typu vztahů, *přípustky*, se klade důraz na rozpor mezi očekáváním, které vyplývá z jednoho argumentu, a jeho popřením v druhém argumentu (op. cit., s. 32).

Vztahy typu kontrast lze pak dále rozdělit na dva podtypy, *juxtapozice* a *opozice*. U *juxtapozice* se jednotce společně pro oba argumenty připisují hodnoty, které lze považovat za alternativy, srov. příklad:

- (9) <Arg1: *John paid \$5*>
 <Arg2: [*but* JUXTAPOZICE] *Mary paid \$10*.> (Prasad et al. 2007)
 <Arg1: *John zaplatil 5 dolarů*,>
 <Arg2: [*ale* JUXTAPOZICE] *Marie 10*.>

Naproti tomu u vztahu *opozičního* jsou připisované hodnoty mezními hodnotami jisté škály, antonymy, srov.:

- (10) <Arg1: *Mary is black*,>
 <Arg2: [*whereas* OPOZICE] *John is white*.> (Prasad et al. 2007)
 <Arg1: *Mary je černoška*,>
 <Arg2: [*zatímco* OPOZICE] *John je běloch*.>

Modifikací typu *kontrast* je *kontrast pragmatický*, u něhož konektor vyjadřuje kontrastivní vztah nikoli mezi vlastními ději argumentů, ale mezi dějem jednoho argumentu a inferencí založenou na druhém argumentu.

K *přípustkovému* typu se řadí takové vztahy, u nichž konektor vyjadřuje nesoulad mezi argumenty: v jednom z nich je popsána situace, která obvykle vede k jistému běžnému výsledku, v druhém argumentu je však takový výsledek popřen. Podle umístění konektoru se přípustkový typ dělí na dva podtypy. U podtypu *očekávání* vytváří Arg2 očekávání, které je v Arg1 popřeno (typickou spojkou je např. *although – ačkoli*); u podtypu *popření očekávání* je očekávání vyjádřeno v Arg1 a popřeno v Arg2 (např. spojka *but – ale*).

Zvláštní typ tvoří *pragmatická přípustka*, kterou autoři blíže nepopisují.

Třída **rozšiřujících vztahů** (expansion) zahrnuje vztahy, které posouvají text dále (op. cit., s. 34–37). Prvním z jejích typů je *konjunkce*. U tohoto vztahu konektor vyjadřuje, že Arg2 obsahuje doplňující informaci k situaci popsané v Arg1; zároveň tento vztah nelze vyjádřit žádným jiným ze vztahů v této třídě.

Exemplifikace je takovým typem vztahu, u něhož konektor vyjadřuje, že Arg1 zmiňuje jistou množinu a Arg2 ji popisuje detailněji (např. konektor *for example* – *například*). Dalším typem je *přeformulování* (restatement), které se dělí do několika podtypů. Při *přeformulování* je v Arg2 nějakým způsobem nahrazeno tvrzení, které bylo uvedeno v Arg1. U podtypu *specifikace* je toto přeformulování konkrétnější (např. konektor *specifically* – *konkrétně*), u *generalizace* naopak obecnější (např. konektor *overall* – *celkově*). Podtyp *ekvivalence* zahrnuje případy, v nichž oba argumenty popisují tutéž situaci z různých hledisek (např. konektor *in other words* – *jinými slovy*).

Čtvrtým typem rozšiřujících vztahů je *alternativa*. U tohoto typu vztahu konektor vyjadřuje, že obsahy obou argumentů jsou navzájem alternativami. Tyto alternativy mohou platit obě zároveň (podtyp *alternativa slučovací*), platí pouze jedna z nich (*alternativa vylučovací*), nebo platí jedna místo druhé (podtyp *alternativa vybraná*, s konektorem *instead* – *místo toho*).

Předposledním typem rozšiřujících vztahů je *výjimka*, u níž konektor vyjadřuje, že Arg2 obsahuje výjimku k nějakému obecnému tvrzení obsaženému v Arg1 (typický konektor je *except* – *vyjma to, že*).

Posledním významovým typem v PDTB 2.0 je *seznam*, totiž takový druh vztahu, u něhož Arg1 a Arg2 jsou položkami seznamu, který byl definován v dřívějším kontextu. Mezi těmito položkami nemusí být žádný přímý významový vztah.

K uvedené klasifikaci dodáváme, že rozdělení významových vztahů podle několika úrovní obecnosti mělo v PDTB 2.0 přímé využití při značkování: anotátoři měli možnost označit mezi argumenty takovou úroveň vztahu, u níž si byli jisti významem. Pokud tedy např. váhali při určení podtypu, stačilo označit typ či pouze nejobecnější třídu vztahů. Na jedné straně se tím zvýšila míra neshody mezi anotátory, na druhé straně se jasněji ukázalo, které vztahy potřebují lépe vymezit, což se odrazilo později ve zpřesněních v další verzi PDTB.

Způsob anotace v Penn Discourse Treebank 2.0 (Prasad et al. 2008b) inspiroval řadu dalších korpusů zaměřených na diskurzní vztahy. Přístupy těchto korpusů a různé modifikace se staly tématem vědeckých diskusí, během nichž různá pracoviště upravují svoje pojetí konektoru, diskurzního argumentu i klasifikaci sémantických typů. Členové týmu Penn Discourse Treebank na základě této zpětné vazby, která pochází také z pražského diskurzního týmu,⁵ upravili svoje anotační schéma v nové verzi

⁵ Srov. bilaterální spolupráci v projektech Kontakt *Towards a Computational Analysis of Text Structure* (ME10018, v letech 2010–2012) a LH – KONTAKT II *Multilingual Corpus Annotation as a Support for Language Technologies* (LH14011, v letech 2014–2016).

korpusu, Penn Discourse Treebank 3.0 (Prasad et al. 2019). Nicméně pro náš výzkum sloužil jako prvotní zdroj právě přístup PDTB 2.0, proto věnujeme hlavní pozornost jemu.

Změny, které proběhly v PDTB 3.0, se české anotace implicitních vztahů přímo netýkají, uvádíme je zde jen pro úplnost. Na rozdíl od PDTB 2.0 je verze 3.0 (Webber et al. 2019) podstatně více zaměřena na diskurzní vztahy vnitrovětné, proto v anotačním schématu došlo k mnoha změnám spojeným právě s vnitrovětnými jevy: bylo upřesněno pojetí konektorů (např. modifikace konektorů, nespojitě konektory skládající se z více částí, odlišení konektorů od větných příslovcí, sekundární konektory vyjádřené gramatickou formou, nikoli lexikálně) a navrženo řešení pro anotaci koordinačních konstrukcí s elipsami. V sémantické klasifikaci diskurzních vztahů byly zavedeny nové typy: *způsob*, *negativní podmínka*, *účel*, *negativní výsledek* a *podobnost*. Podrobněji byly rozlišeny vztahy pragmatické, a to na vztahy vyjadřující epistemické znalosti a vztahy vyjadřující řečové akty. Anotátoři měli nově možnost vyznačovat mezi dvěma stejnými argumenty více diskurzních vztahů, nikoli jen jeden (nově typicky jeden explicitní a navíc jeden implicitní), což umožňuje lepší zachycení významové stavby textu. Změnilo se také označování (číslování) argumentů v diskurzním vztahu: není určováno povrchovou pozicí diskurzního konektoru, nýbrž (syntaktickým) významem argumentů.

4.2 Funkční generativní popis a tektogramatická rovina

Při přípravách k popisu stavby textu v češtině jsme měli k dispozici jako zdroj nejen Penn Discourse Treebank zaměřený na diskurzní vztahy jako takové, ale také Pražský závislostní korpus (Prague Dependency Treebank, dále PDT; Bejček et al. 2011), který již obsahoval anotaci některých jevů, jež by bylo možné převzít pro anotaci diskurzních vztahů. Šlo zejména o anotaci větné sémantiky. Protože se teoretický přístup a technická stránka tohoto korpusu staly základem pozdější anotace diskurzu, popisujeme jej v této podkapitole podrobněji.

Teoretickým východiskem Pražského závislostního korpusu je *funkční generativní popis* (Sgall et al. 1969; Sgall et al. 1986), který popisuje jazykový systém jako soustavu rovin. Jednotky na jednotlivých rovinách mají své *formy a funkce*. Přitom formy nižších rovin slouží jako vyjádření funkcí rovin vyšších: např. nepřítomnost hlásek na rovině fonologické (forma) je v jistém kontextu vyjádřením nulové koncovky nominativu na rovině morfologické (funkce). Tentýž nominativ, tentokrát jako forma na rovině morfologické, je vyjádřením funkce subjektu na rovině povrchové syntaxe. A subjekt je formou, která – opět v konkrétním kontextu – vyjadřuje funkci aktora v rovině větné sémantiky.

Tato soustava rovin byla s určitými technickými úpravami převzata při vytváření Pražského závislostního korpusu⁶, který rozlišuje tzv. rovinu morfologickou, analy-

⁶ Mikulová et al. 2006

tickou (povrchová syntax) a tektogramatickou (větná sémantika). Na těchto rovinách jsou komplexně anotovány věty z českých textů.

Rovina morfologická. Jednotlivým slovům ve větě jsou na morfologické rovině přiřazeny morfologické značky v souladu se zásadami Českého národního korpusu (Hnátková et al. 2014).

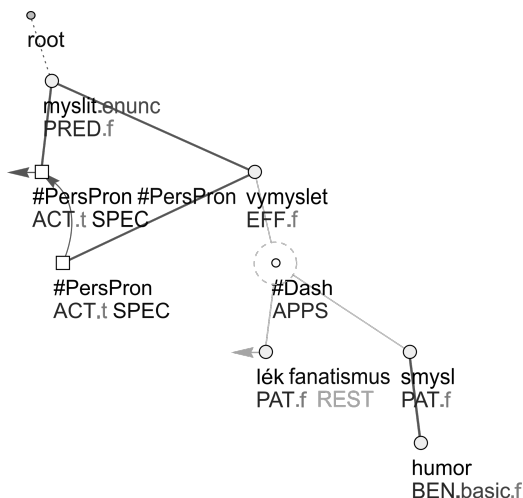
Rovina analytická. Na této rovině jsou slova ve větě uspořádána do grafu v podobě závislostního stromu. Jednotlivá slova zde tvoří uzly, které jsou pak spojeny závislostními hranami. Závislostní hrany jsou označovány druhem syntaktického vztahu mezi slovy, např. subjekt, objekt, pomocný slovesný tvar, předložka. Tato rovina je chápána jako technická, pomocná; jejím účelem je vyjádřit syntaktické vztahy mezi všemi vyjádřenými slovy ve větě, tj. mezi všemi vstupními jednotkami z morfologické roviny, nerekonstruuji se však na ní slova nevyjádřená (např. nevyjádřený podmět). Na rozdíl od tradičního větného rozboru (např. MČ 3, s. 13) se v ní nerozlišují slova s větněčlenskou platností a bez ní, ve větné struktuře se zachycují všechna.

Rovina tektogramatická. Tektogramatická rovina zachycuje významovou stavbu vět, opět v podobě závislostních stromů, stejně jako na rovině analytické. Uzly stromu tvoří lemmata plnovýznamových slov a některé uzly technické; vzájemně jsou spojeny závislostními hranami. Závislostní hrany jsou označeny druhem významového vztahu mezi uzly, např. predikát, aktor, patiens, směr kam, míra, doprovod, odkaz k předcházejícímu kontextu. Tektogramatická rovina obsahuje také uzly rekonstruované, např. nevyjádřené aktory a jiné valenční členy. U tektogramatických uzlů je značkováno rozsáhlé spektrum jejich jazykových vlastností, jako je např. číslo, rod, zdvořilostní forma, negace, slovesný čas, rezultativnost, původnost / rekonstrukce uzlu apod. Významy synsémantických slov, která jsou znázorněna v grafu na rovině analytické, bývají na tektogramatické rovině zachyceny jako význam závislostní hrany nebo vlastnost uzlu.

Součástí popisu významové stránky na tektogramatické rovině je také její aktuální členění. V anotaci aktuálního členění je u každého uzlu vyznačena hodnota jeho kontextové (ne)zapojenosti, uzly jsou ve stromě řazeny odleva doprava podle vzrůstající výpovědní dynamičnosti.

V Pražském závislostním korpusu byly na tektogramatické rovině vyznačeny také vztahy, které vystihují zapojení věty v textu, konkrétně vztahy koreference a asociační anafory. Anotace koreference spojuje v textu všechny uzly se společným referentem do jednoho koreferenčního řetězce, např. podstatné jméno a zájmeno či nevyjádřený uzel dále v textu, které k němu odkazují. Anotace asociační anafory podobným způsobem zachycuje specifické významové vztahy mezi entitami, jako je např. množina – podmnožina (*žáci – někteří žáci*), část – celek (*ouško – hrnek*).

Způsob anotace na tektogramatické rovině je představen na obr. 4.1, který znázorňuje závislostní strom jedné věty. Jednotlivé roviny Pražského závislostního korpusu jsou navzájem propojené, je tedy možné sledovat např. u jednoho slova jeho



Obrázek 4.1: Tektogramatický strom věty *Myslím, že jsem vymyslel lék – smysl pro humor.*

U každého uzlu je uvedeno lemma výrazu (např. *myslit*), funktor vztahu (např. *PRED* pro predikaci) a hodnota kontextové zapojenosti (např. *f* – uzel kontextově nezapojený). Apoziční dvojice je znázorněna zvláštním uzlem *APPS*, který spojuje oba členy apoziční dvojice (*lék – smysl*). Rekonstruované uzly jsou zakresleny jako čtverce (nevyjádřené aktory u sloves *myslím* a *vymyslel jsem*). Tyto uzly jsou spojeny vztahem koreference znázorněným šipkou, u každého z nich je vyznačen typ reference (*SPEC* – specifická). Z prvního nevyjádřeného uzlu vede další koreferenční šipka do předchozího kontextu, k uzlu s lemmatem *#PersPron*. Ke kontextu odkazuje také šipka asociační anafory spojující slovo *lék* se slovem *fanatismus* v předchozí větě *Bojíte se rostoucí síly fanatismu na Blízkém východě?*; zároveň je označen také významový podtyp asociační anafory (*REST* – ostatní).

morfologické, syntaktické a významové vlastnosti zároveň či vyhledávat syntaktické ztvárnění určitých struktur v aktuálním členění apod.⁷

4.2.1 Významová klasifikace tektogramatických vztahů v Pražském závislostním korpusu (funktory)

Ačkoliv tektogramatická rovina neobsahuje anotaci diskurzivních vztahů (což odpovídá konceptu tektogramatické roviny, diskurzivní vztahy jako celek přesahují úroveň větné sémantiky, pro niž je tektogramatická rovina vymezena), jsou v ní značeny některé významové typy závislostních a jiných vztahů, které v určitých syntaktických

⁷ Pro vyhledávání v Pražském závislostním korpusu slouží nástroje TrEd (Pajas - Štěpánek 2008) a PML-Tree Query (PML-TQ; Pajas - Štěpánek 2009).

strukturách odpovídají vztahům diskurzním. Jde zejména o různé druhy vět vedlejších a o vztahy souřadné, ale i o další jevy.

Obecně se významové typy značí u jednotlivých uzlů v závislostním stromě pomocí hodnoty *funktor*.

Funktory se podle sémantických a syntaktických vlastností uzlů dělí do několika rozsáhlejších skupin (viz Mikulová et al. 2006, s. 426):

- funktory jako sémantické ohodnocení syntaktického vztahu závislosti
- funktory pro efektivní kořeny nezávislých klauzí, které vyjadřují nezávislost a určují typ konstrukce
- funktory pro kořeny souřadných struktur, popisující významový druh souřadného spojení
- funktory pro závislé části víceslovných lexikálních jednotek
- funktor pro uzly reprezentující cizojazyčné výrazy
- funktory atomických uzlů

Značení některých funktorů lze využít buď přímo, či jako podkladu pro anotaci diskurzních vztahů. V první skupině, zachycující význam závislostního vztahu, odpovídají významu diskurzních vztahů funktory účelu, příčiny, přípustky, podmínky, konfrontace (vztah podřadný), záměru a vztahů časových (funktory AIM, CAUS, CNCS, COND, CONTRD, INTT, TPAR a další vztahy časové), viz tab. 4.2.⁸

Zkratka	Angl. označení	Význam	Příklad
ACMP	Accompaniment	Volné doplnění vyjadřující doprovod (v širokém smyslu)	tatínek s maminkou .ACMP
ACT	Actor	Aktant – aktor	Otec .ACT pracuje.
ADDR	Addressee	Aktant – adresát	Poslal dárek příteli .ADDR
AIM	Aim	Volné doplnění vyjadřující účel	Cvičí, aby zhubla .AIM
APP	Appurtenance	Volné doplnění substantiva vyjadřující přináležitost	můj .APP hrad
AUTH	Author	Volné doplnění substantiva označující autora	Nezvalovy .AUTH verše
BEN	Benefactor	Volné doplnění vyjadřující ne/prospěch	Pracuje pro firmu .BEN
CAUS	Cause	Volné doplnění vyjadřující příčinu	Z důvodu nemoci .CAUS zavřeno.

⁸ Tabulky s přehledem hodnot technického atributu *functor* v Pražském závislostním korpusu vycházejí ze souhrnné tabulky in Mikulová et al. 2006, s. 427–429.

Zkratka	Angl. označení	Význam	Příklad
CNCS	Concession	Volné doplnění vyjadřující přípustku	Navzdory studijním úspěchům. CNCS se v praxi neuplatnil.
COMPL	Complement	Volné doplnění – doplněk	Vrátila se unavená. COMPL
COND	Condition	Volné doplnění vyjadřující podmínku	Když spí. COND, nezlobí.
CONTRD	Contradiction	Volné doplnění vyjadřující konfrontaci	Zatímco mzdy klesají. CONTRD, ceny se zvyšují.
CPR	Comparison	Volné doplnění vyjadřující srovnání	víc než tisíc. CPR korun
CRIT	Criterion	Volné doplnění vyjadřující měřítko	Seřaď slova podle abecedy. CRIT
DIFF	Difference	Volné doplnění vyjadřující rozdíl	Je vyšší o dva centimetry. DIFF
DIR1	Directional: from	Volné doplnění místa odpovídající na otázku „odkud“	Přijel z Prahy. DIR1
DIR2	Directional: which way	Volné doplnění místa odpovídající na otázku „kudy“	Jdou lesem. DIR2
DIR3	Directional: to	Volné doplnění místa odpovídající na otázku „kam“	Přišel domů. DIR3
EFF	Effect	Aktant – efekt	Jmenovali ho předsedou. EFF
EXT	Extent	Volné doplnění vyjadřující míru	V nádobě je přesně. EXT litr vody.
HER	Heritage	Volné doplnění vyjadřující dědictví	šátek po matce. HER
ID	Identity	Nominativ jmenovací a genitiv explikativní	hrad Karlštejn. ID; trest smrti. ID
INTT	Intent	Volné doplnění vyjadřující záměr	Šel nakoupit. INTT

Zkratka	Angl. označení	Význam	Příklad
LOC	Locative	Volné doplnění místa odpovídající na otázku „kde“	Pracuje v Praze .LOC
MANN	Manner	Volné doplnění vyjadřující způsob	Mluví hlasitě .MANN
MAT	Material	Aktant substantiva vyjadřující obsah kontejneru	sklenice vody .MAT
MEANS	Means	Volné doplnění vyjadřující prostředek	Píše perem .MEANS
ORIG	Origo (pův. lat. označení)	Aktant – origo	Vyrábí nábytek ze dřeva .ORIG
PAT	Patient	Aktant – patiens	Vaří oběd .PAT
REG	Regard	Volné doplnění vyjadřující zřetel	Vzhledem k počasí .REG nelze nic plánovat.
RESL	Result	Volné doplnění vyjadřující účinek	Mluví tak potichu, že mu nerozumíme .RESL
RESTR	Restriction	Volné doplnění vyjadřující výjimku	Kromě tebe .RESTR tam byli všichni.
RSTR	Adnominal modifications (pův.: restrictive adnominal modifications)	Volné doplnění blíže specifikující řídící substantivum	velký .RSTR dům
SUBS	Substitution	Volné doplnění vyjadřující substituci	Za otce .SUBS jednal strýc.
TFHL	Temporal: for how long	Volné doplnění času odpovídající na otázku „na jak dlouho?“	Přijel na měsíc .TFHL
TFRWH	Temporal: from when	Volné doplnění času odpovídající na otázku „ze kdy?“	Přeložil jednání ze soboty .TFRWH na dnešek.
THL	Temporal: how long	Volné doplnění času odpovídající na otázku „jak dlouho?“ a „za jak dlouho?“	Stihnul to za týden .THL

Zkratka	Angl. označení	Význam	Příklad
THO	Temporal: how often	Volné doplnění času odpovídající na otázku „jak často?“ a „kolikrát?“	Pracuju na tom každý den .THO
TOWH	Temporal: to when	Volné doplnění času odpovídající na otázku „na kdy?“	Přeložil jednání ze soboty na dnešek .TOWH
TPAR	Temporal: parallel	Volné doplnění času odpovídající na otázku „současně s čím?“ a „během jaké doby?“	Během naší dovolené .TPAR ani jednou nepršelo.
TSIN	Temporal: since when	Volné doplnění času odpovídající na otázku „od kdy?“	Budu pracovat od zítra .TSIN
TTILL	Temporal: till	Volné doplnění času odpovídající na otázku „do kdy?“	Udělám to do pátku .TTILL
TWHEN	Temporal: when	Volné doplnění času odpovídající na otázku „kdy?“	Přijdu zítra .TWHEN

Tabulka 4.2: Funktory jako sémantické ohodnocení syntaktického vztahu závislosti

Ve skupině funktorů pro kořeny souřadných struktur se pak význam diskurzivních vztahů objevuje u funktorů pro vztah odporovací, konfrontační (souřadný), slučovací, důsledkový, vylučovací, stupňovací a důvodový (funktory ADVS, CONFR, CONJ, CSQ, DISJ, GRAD, REAS), viz tab. 4.3.

Zkratka	Angl. označení	Význam	Příklad
ADVS	Adversative	Kořen koordinační struktury vyjadřující odporovací vztah	Viděl, ale .ADVS neslyšel.
CONFR	Confrontation	Kořen koordinační struktury vyjadřující konfrontační vztah	Pavel se zlepšuje, kdežto .CONFR Jan dostává čtyřky.
CONJ	Conjunction	Kořen koordinační struktury vyjadřující slučovací vztah	Pavel a .CONJ Jan

CONTRA	Opposing entities	Kořen členské koordinační struktury vyjadřující vztah dvou vzájemně se střetávajících subjektů	otec versus .CONTRA syn
CSQ	Consequence	Kořen koordinační struktury vyjadřující důsledkový vztah	Pracoval nezodpovědně, a proto .CSQ dostal výpověď.
DISJ	Disjunction	Kořen koordinační struktury vyjadřující vylučovací vztah	Pojedu já, nebo .DISJ ty.
GRAD	Gradation	Kořen koordinační struktury vyjadřující gradační vztah	Běžel, ba .GRAD utíkal.
REAS	Reason	Kořen koordinační struktury vyjadřující důvodový vztah	Dostal výpověď, neboť .REAS pracoval nezodpovědně.
APPS	Apposition	Kořen apoziční struktury	substantivum neboli .APPS
OPER	Operand	Kořen souřadné struktury vyjadřující matematickou operaci nebo interval	podstatné jméno pět až .OPER deset hodin

Tabulka 4.3: Funktory pro kořeny souřadných struktur, popisující významový druh souřadného spojení

U obou těchto skupin, tj. u struktur souřadných a podřadných, je nutné sledovat, jaké jednotky jsou tímto vztahem spojeny, zda jde o skutečné diskurzní argumenty, tj. minimálně větně vyjádřené propozice. Pokud se jedná pouze o vztah mezi členy, nikoli mezi klauzemi či vyššími jednotkami, nehodnotíme ho jako vztah diskurzní. Proto také funktory jako např. protiklad (CONTRA), apozice (APPS) či operand (OPER) nezařazujeme jako možné ekvivalenty diskurzních vztahů.

Další skupiny funktorů mohou poukazovat nikoli na diskurzní vztah, ale buď na možné diskurzní argumenty (funktory označující efektivní kořeny závislostních stromů), nebo na diskurzní konektory. Možné diskurzní argumenty mohou mít jako vrchol jeden z efektivních kořenů závislostních stromů, viz tab. 4.4.

Zkratka	Angl. označení	Význam	Příklad
PRED	Predication	Efektivní kořen nezávislé slovesné klauze, která není vsuvkou	Pavel dal .PRED kytku Martině.
DENOM	Denom	Efektivní kořen nezávislé nominativní klauze, která není vsuvkou	Základní škola .DENOM
VOCAT	Vocative case	Efektivní kořen nezávislé vokativní klauze	Hanko .VOCAT, podej mi to.
PARTL	Interjectional clause (pův.: particle)	Efektivní kořen nezávislé citoslovečné klauze	Hurá .PARTL, vyhráli jsme!
PAR	Parenthesis	Efektivní kořen vsunuté slovesné nebo nominativní klauze	Přijedu 13. prosince (pátek .PAR)

Tabulka 4.4: Funktory pro efektivní kořeny nezávislých klauz, které vyjadřují nezávislost a určují typ konstrukce

Diskurzní konektory pak mohou odpovídat funktorům atomických uzlů (viz tab. 4.5), zejména funktoru návaznosti na předchozí kontext (PREC), ale okrajově i ostatním funktorům (rematizátory, výrazy vyjadřující modální charakteristiku či postoj mluvčího – RHEM, MOD, ATT).

Zbýlé funtory užívané v Pražském závislostním korpusu přímo ke stavbě diskurzu neodkazují (tab. 4.6).

Celkově lze tedy z anotace funktorů v Pražském závislostním korpusu získat informace o diskurzních vztazích odpovídajících syntaktickým vztahům souřadným a podřadným a dále o možných diskurzních konektorech. Tektogramatická anotace je zaměřena na zachycení větné sémantiky, nikoli diskurzních vztahů, proto také obsahuje mnoho informací a značek, které jsou z hlediska stavby diskurzu nadbytečné. Nicméně jevy, které patří do oblasti diskurzních vztahů, lze v Pražském závislostním korpusu dobře vyfiltrovat. Napomáhá tomu jednak značení funktorů, jednak propojení tektogramatické roviny s dalšími anotovanými rovinami, zejména s rovinou analytickou (syntaktickou). Díky tomuto propojení lze dobře najít možné diskurzní argumenty, tj. odlišit jednotky členské a větné.

Zkratka	Angl. označení	Význam	Příklad
ATT	Attitude	Atomický výraz vyjadřující postoj mluvčího	Je to samozřejmě .ATT pravda
MOD	Modality	Atomický výraz vyjadřující modální charakteristiku obsahu výpovědi	Pracuje asi .MOD na půl úvazku.
PREC	Reference to preceding clause	Atomický výraz, který vyjadřuje návaznost klauze na předchozí kontext	A .PREC pak odešel.
RHEM	Rhematizer	Atomický výraz – rematizátor	Jen .RHEM Karel odešel.
INTF	Intensification	Atomický výraz vyjadřující „falešný podmět“	Ono .INTF prší.

Tabulka 4.5: Funktory atomických uzlů

Zkratka	Angl. označení	Význam	Příklad
CPHR	Compound phraseme	Jmenná část složeného predikátu	mít plán .CPHR
DPHR	Dependent part of phraseme	Závislá část frazému	křížem krážem .DPHR
CM	Conjunction modifier	Modifikátor souřadícího spojovacího výrazu	otec a také .CM syn
FPHR	Foreign phrase	Součást cizojazyčného textu	Cash .FPHR flow .FPHR

Tabulka 4.6: Funktory pro závislé části víceslovných lexikálních jednotek a pro uzly reprezentující cizojazyčné výrazy

4.3 Pražský diskurzní korpus a navazující výzkum diskurzních vztahů

Výzkum diskurzních vztahů, který na Matematicko-fyzikální fakultě UK začal v roce 2007, zkombinoval oba uvedené přístupy, postup Pražského závislostního korpusu (PDT) i korpusu Penn Discourse Treebank (PDTB).

Jako základ pro vznikající Pražský diskurzní korpus 1.0 (dále PDiT 1.0, Poláková et al. 2012a) posloužila data a anotace Pražského závislostního korpusu. Toto východisko se ukázalo jako výhodné, protože jednak PDT obsahuje na tektogramatické rovině značné množství údajů, které lze přímo či nepřímo využít pro anotaci dis-

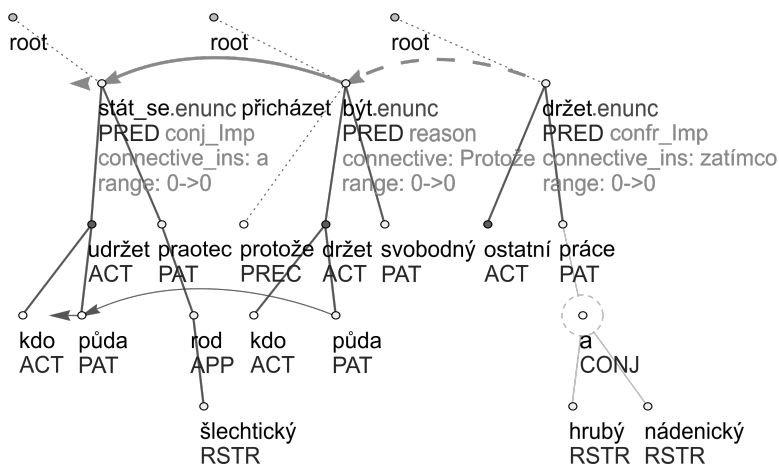
kurzu, jednak jsou jeho data již analyzována v jednotném syntaktickém rámci, který umožňuje přistupovat konzistentně ke složitějším syntaktickým jevům. K nim patří např. koordinační struktury, elipsy, vsuvky, zpracování přímé řeči apod.

Na PDTB pak Pražský diskurzní korpus 1.0 navazuje v pojetí diskurzního konektoru, diskurzního argumentu a diskurzního vztahu (viz kap. 2). Role diskurzního konektoru v systému diskurzních vztahů je však v Pražském diskurzním korpusu po teoretické stránce hodnocena trochu odlišně. Pro Lexicalized Tree-Anjoining Grammar je lexikální jednotka základem, od něž vychází při interpretaci věty. Proto také Penn Discourse Treebank u implicitních vztahů pracuje s představou konektoru, který je možné mezi diskurzní argumenty vsunout, a značuje pak tento konektor. Naproti tomu Pražský závislostní korpus a s ním i Pražský diskurzní korpus vycházejí z představy funkce, kterou lze v textu vyjadřovat různými formami. Forma, v tomto případě diskurzní konektor, je pouze projevem funkce; důležité je odhalit funkce a potom sledovat možnosti jejich vyjádření. Proto také Pražský diskurzní korpus pracuje s otevřenou sadou konektorů; u implicitních vztahů se zde anotuje přímo sémantika vztahu, konektor či parafrázi lze do vztahu vsunout, aby bylo zřejmé, jak anotátor vztah pochopil, vsunutí konektoru / parafráze však není podmínkou pro anotaci implicitního vztahu.

Diskurzní vztahy byly v PDiT anotovány na tektogramatické rovině, tj. na tektogramatických závislostních stromech. (Zachycení diskurzního vztahu na tektogramatické rovině je znázorněno na obrázku 4.2.) Tyto vztahy jsou v korpusu vyznačeny pomocí orientované diskurzní šipky, která spojuje přesně vymezené argumenty. Argument přitom může mít rozsah od jednoho uzlu ve stromě až po několik stromů. Diskurzní šipka je označována významovým typem vztahu (např. přípustka, gradace, rektifikace); o významové klasifikaci viz níže. Orientace šipky se pak řídí významem jednotlivých argumentů ve vztahu: šipka vede vždy od Arg2 k Arg1, přičemž číslo argumentu je přiděleno na základě jeho významu. Např. u vztahu nesoučasnosti vede šipka od argumentu, jehož děj probíhá dříve, k argumentu s dějem pozdějším, bez ohledu na povrchové pořadí těchto argumentů v textu. V tomto bodě se liší označování argumentů v Pražském diskurzním korpusu a v Penn Discourse Treebank: v PDTB 2.0 se argumenty číslují podle toho, ve kterém z nich se nachází konektor. Naproti tomu rozlišování argumentů podle významu umožňuje značkovat stejně vztahy, které mají stejný význam, avšak různou formu (např. podřadná souvětí s VV přípustkovou a souřadná odporovací spojení vět s přípustkovým významem či podřadná souvětí s VV příčinnou a souřadná důvodová spojení vět).⁹

Diskurzní šipce je vedle významového vztahu připsán také konektor, který tento vztah vyjadřuje. Slouží to pro kontrolu jednoty anotace. U implicitních vztahů bývá připsán konektor či vysvětlení, které ukazují, jak anotátoři pochopili význam daného vztahu.

⁹ Používáme tu vazbu *spojení vět* proto, že nejde pouze o souvětí souřadná, ale také o spojení více vět či souvětí graficky oddělených koncovým interpunkčním znaménkem.



Obrázek 4.2: Obrázek znázorňuje anotaci vět *Kdo půdu udržel, stal se praotcem šlechtického rodu. Protože kdo držel půdu, byl svobodný. Ostatní drželi hrubé a nádenické práce. Mezi první a druhou větou je pomocí plně silné šipky vyznačen explicitní diskurzní vztah příčiny, který je vyjádřen pomocí konektoru *protože*. U výchozího uzlu šipky (s lemmatem *být*) je vztah popsán: je vyznačen jeho sémantický typ (*reason*) a příslušný konektor (*protože*). Hodnota *range* ukazuje rozsah jednotlivých argumentů, *0->0*, tj. argumenty přesně odpovídají výchozímu a cílovému stromu. Orientace šipky naznačuje význam argumentů: u příčinného vztahu vede šipka od argumentu s významem příčiny (Arg2) k argumentu s významem důsledku (Arg1). Druhá a třetí věta jsou spojeny implicitním vztahem konfrontace, který je vyznačen pomocí přerušované šipky. U výchozího uzlu šipky (s lemmatem *držet*) je opět vyznačen sémantický typ vztahu a konektor (tj. u implicitního vztahu vsunutý konektor, u něhož anotátor předpokládal, že nejlépe vystihuje význam vztahu mezi diskurzními argumenty). Hodnota *range* zde opět označuje, že do výchozího i cílového argumentu patří vždy řídící uzel a všechny uzly závislé. Třetí diskurzní vztah je implicitní konjunkce, která spojuje první větu s předchozím kontextem (vede od slova *stát se* ke slovu *přicházet* v předcházejícím kontextu).*

4.3.1 Významová klasifikace diskurzních vztahů v Pražském diskurzním korpusu

Sémantická klasifikace vztahů v Pražském diskurzním korpusu vychází z přístupu Penn Discourse Treebank 2.0, který doplňuje a upravuje na základě rozdělení funktořů v Pražském závislostním korpusu. Na některých místech mají obě zdrojové klasifikace úplné ekvivalenty, jinde však funktory PZK neodpovídají vymezení sémantických typů v PDTB přímo. Obě klasifikace se liší na různých místech v míře podrobnosti (srov. např. velmi podrobné členění typů podmínky v PDTB, které nemá ekvivalent v PZK; naproti tomu zase PZK člení podrobněji než PDTB vztahy časové). Obě klasifikace také rozlišují některé vlastní vztahy, které se na druhé straně neobje-

vují. U PDTB jsou to např. vztahy generalizace, specifikace či ekvivalence, tj. vztahy neopírající se o syntax; u PZK pak jde např. o vztah účelu a gradace.¹⁰

Rozdíl je také v tom, že PDTB používá trojúrovňovou hierarchii vztahů, zatímco v PZK existuje pouze jedna úroveň funktorů. To znamená, že jev, který je na tektoqramatické rovině anotován jedním způsobem, může být podle schématu PDTB anotován třemi způsoby, v závislosti na jistotě anotátora. Jednoúrovňový typ anotace byl v PZK přijat na základě zkušeností z předchozích výzkumů v Ústavu formální a aplikované lingvistiky: možnost anotovat jak detailní, tak obecnější kategorie (či možnost přiřazovat jednomu jevu více značek v případech pochyb) vede ve svém důsledku k vyšší anotátorské neshodě, protože tak vzniká více možností, jak zkoumaný jev označovat.

Pro Pražský diskurzní korpus byla vytvořena klasifikace vztahů, kterou představuje tab. 4.7.

Ačkoli se diskurzní vztahy v Pražském diskurzním korpusu dělí do čtyř skupin jako v PDTB, je vlastní česká anotace jednoúrovňová: anotátoři neměli možnost vyznačit u vztahů, kde si nebyli jisti významem, pouze obecnější významovou skupinu. Naopak, jejich úkolem bylo vždy se pokusit o přesnější interpretaci. Rozdělení vztahů do skupin tedy slouží pro teoretický popis, příp. pro srovnání s PDTB a jinými korpusy.¹¹

Časové vztahy se dělí na vztah *současnosti* a *nesoučasnosti*. Vztah současnosti odpovídá vztahu současnosti v PDTB a vyjadřuje, že děje obou diskurzních argumentů probíhají zároveň, srov. příklady 11–12. Jako argument 1 je označen ten argument, který je v textu pozičně vyjádřen dříve (vlevo).^{12, 13}

- (11) <Arg1: *Město postihla krize a nezaměstnanost.*>
<Arg2: [**Zároveň** SOUČASNOST] *začala nová éra svobodných celních zón.*> (PDT 3.5)
- (12) „*Mezi Dánskem a Českou republikou neexistují žádné sporné body,*“
<Arg1: *prohlásil včera dánský ministr zahraničí Niels Helveg Petersen,*>

¹⁰ Účelový vztah byl zařazen do novější klasifikace významových vztahů v PDTB 3.0 (Webber et al. 2019).

¹¹ Vztah *kontinuace* byl do klasifikace zařazen až s anotací implicitních vztahů, tedy v korpusu PDiTEDA 1.0, viz níže.

¹² V příkladech z pražských korpusů značíme rozsah jednotlivých argumentů pomocí špičatých závorek, pro označení argumentů používáme zkratky Arg1 a Arg2. Pokud je ve vztahu vyjádřen diskurzní konektor, je vtištěn tučně. U implicitních diskurzních vztahů je doplněný konektor (příp. parafráze), který osvětluje význam diskurzního vztahu, vložen v hranatých závorkách na příslušné místo. Sémantický typ vztahu je označen buď u explicitního konektoru, nebo u implicitních vztahů na začátku druhého argumentu.

¹³ Příklady explicitních diskurzních vztahů v tomto přehledu jsou převzaty z článku Zikánová et al. 2015b a pocházejí z verze korpusu PDT 3.0 (dále cit. jako PDT 3.5, Hajič et al. 2018).

Skupina vztahů	Název diskurzního vztahu	Značka
Časové vztahy	současnost nesoučasnost	synchr preced
Kauzální vztahy	příčina-důsledek nepravá příčina-důsledek explikace podmínka nepravá podmínka účel	reason f_reason explicat cond f_cond purp
Kontrastivní vztahy	konfrontace opozice restriktivní opozice nepravý kontrast přípustka rektifikace gradace	confr opp restr f_opp conc corr grad
Navazovací vztahy	konjunkce kontinuace konjunktivní alternativa disjunktivní alternativa exemplifikace specifikace ekvivalence generalizace	conj empty_relation conjalt disjalt exempl spec equiv gener

Tabulka 4.7: Významová klasifikace diskurzních vztahů v Pražském diskurzním korpusu

<Arg2: [vsunuto: **kdýž** IMPLICITNÍ SOUČASNOST] který byl v Praze na krátké návštěvě.> (PDt-EDA 1.0)

Vztah *nesoučasnosti* vyjadřuje, že děje diskurzních argumentů probíhají po sobě. Posloupnost dějů je vyjádřena číslováním argumentů: děj probíhající dříve je označen jako argument 1, ať se v textu pozičně nachází kdekoli. (Viz příklady 13–15.)

- (13) <Arg1: Na žádost soudů musel poskytnout některé informace, a tudíž krabice s uloženými dokumenty musel otevřít.>
<Arg2: [**Poté** NESOUČASNOST] je opatřil novými pečeti.> (PDT 3.5)
- (14) <Arg2: Společnost vznikla transformací větší části bývalého státního podniku Jihlavan.>

<Arg1: [**Předtím** NESOUČASNOST] z něho byly vyčleněny a v malé privatizaci samostatně vydraženy výrobní hala, galvanika a kalírna.> (PDT 3.5)

- (15) <Arg1: Gyula Horn se během své návštěvy omluvil za křivdy spáchané na maďarských Slovácích, řekl slovenský prezident Michal Kováč zpravodaji ČTK.>
 <Arg2: [vsunuto: **nato** IMPLICITNÍ NESOUČASNOST] Hornova mluvčí zprávu de-mentovala a prohlásila, že tomu bylo právě naopak: Horn prý vyzval Slovensko, aby se omluvilo za odsun slovenských Maďarů a za konfiskaci jejich majetků na základě Benešových dekretů.> (PDiT-EDA 1.0)

Mezi **kauzální vztahy** patří dva vztahy příčinné (příčina-důsledek a nepravá příčina-důsledek), dále *explikace*, dva vztahy podmínkové (podmínka a nepravá podmínka) a vztah účelu. U vztahu *příčiny a důsledku* vyjadřuje argument 2 příčinu děje, který nastal jako důsledek děje v argumentu 1 (příklady 16–18). Nezáleží přitom na formálním syntaktickém ztvárnění vztahu, za synonymní zde považujeme podřadná souvětí s VV příčinnou i souřadná spojení důsledková.

- (16) <Arg2: Pivo je podle jeho slov vhodné zejména po tělesné námaze.>
 <Arg1: [**Proto** PŘÍČINA-DŮSLEDEK] bude ve sklenicích o obsahu 0,25 litru nabízeno například ve fitnesscentrech a na plovárnách.> (PDT 3.5)
- (17) Minulý týden došlo na stránce Automobilismus k politováníhodné chybě.
 <Arg2: U článku o Hondě Legend byla zaměněna fotografie s Roverem 214i.>
 <Arg1: Čtenářům i Hondě Česká republika, která nám vůz k testování zapůjčila, se [vsunuto: **proto** IMPLICITNÍ PŘÍČINA-DŮSLEDEK] omlouváme a auto ještě jednou připomínáme.> (PDiT-EDA 1.0)
- (18) <Arg1: Nejsem ještě připravena na to, mít děti.>
 <Arg2: [vsunuto: **protože** IMPLICITNÍ PŘÍČINA-DŮSLEDEK] Mám příliš mnoho práce.> (PDiT-EDA 1.0)

Vztah *nepravé příčiny a důsledku* vyjadřuje, že mezi oběma argumenty dochází k vysuzování, případně nejde přímo o spojení dvou vyjádřených argumentů, ale o inference, které se z těchto argumentů vyzumívají. Jako argument 2 se přitom označuje ta část vztahu, která vyjadřuje (domnělou) příčinu děje ve druhém argumentu. V příkladu 19 je vztah mezi argumenty interpretován takto: domníváme se, že veřejnost bere v úvahu na premiéra Klause v úvahu také jeho odpovědnost za celou vládu, protože víme, že obecně je důvěra v premiéra závislá na důvěře ve vládu. V příkladu 20 se pak objevuje komentář k průběhu komunikace: vyzývám vás, abyste se neptal, proč jsem přijel, protože pokládám vaši otázku za nepatřičnou, neadekvátní – nepatřičná je proto, že přijet sem je normální.

- (19) <Arg1: Při posuzování premiéra Klause bere veřejnost patrně v úvahu i jeho odpovědnost za činnost celého kabinetu.>
 <Arg2: Dlouhodobé výsledky STEM [**totiž** NEPRÁVÁ PŘÍČINA-DŮSLEDEK] ukazují, že vždy, když klesala důvěra ve vládu, klesala i důvěra v premiéra.> (PDT 3.5)
- (20) <Arg1: Neptejte se mě, proč jsem přijel do Prahy.>
 <Arg2: [vsunuto: **protože** IMPLICITNÍ NEPRÁVÁ PŘÍČINA-DŮSLEDEK] Je to přece normální sem přijet.> (PDiT-EDA 1.0)

Vztahem *explikace* jsou spojeny dva argumenty, z nichž jeden (argument 2) obsahuje vysvětlení obsahu druhého (argument 1), přičemž nejde o vztah příčinný ani specifikační. Obvykle toto vysvětlení představuje jiný pohled na skutečnost popsanou v argumentu 1 (příklady 21–22).

- (21) <Arg1: Včerejší porada ministrů o státním rozpočtu na rok 1995 dopadla víc než dobře.>
 <Arg2: Václav Klaus ani Ivan Kočárník [**totiž** EXPLIKACE] nenašli v Kramářově vile nikoho, kdo by se s nimi chtěl prát o ideu vyrovnaného rozpočtu.> (PDT 3.5)
- (22) <Arg1: Ve státech ES je navíc právem spotřebitele, aby měl na výrobku napsáno, z čeho je složen.>
 <Arg2: To mu [vsunuto: **totiž** IMPLICITNÍ EXPLIKACE] umožňuje vyhnout se konzumaci výrobků, které obsahují velké množství cizorodých látek.> (PDiT-EDA 1.0)

U vztahu *podmínky* vyjadřuje obsah argumentu 2 podmínku, za níž dochází k ději vyjádřenému v argumentu 1 (příklady 23–24).

- (23) <Arg1: Ceny centrálně vyráběného tepla by mohly být i nižší.>
 <Arg2: [**Ale** PODMÍNKA] výrobci tepla musí začít hodně investovat.> (PDT 3.5)
- (24) <Arg1: Unionističtí vůdci odmítli zúčastnit se jednání,>
 <Arg2: [vsunuto: **pokud** IMPLICITNÍ PODMÍNKA] na němž by byla zastoupena Sinn Fein.> (PDiT-EDA 1.0)

Vztah *nepravé podmínky* byl v Pražském diskurzním korpusu zaveden pro konstrukce s konektory typickými pro vztah podmínky, nicméně v odlišném užití. Jako argument 2 se označuje argument vyjadřující nepravou podmínku, argument 1 pak obsahuje výsledek působení nepravé podmínky (příklad 25). U implicitních vztahů se nepodařilo v korpusu PDiT-EDA 1.0 tento vztah identifikovat.

- (25) <Arg2: [**Jestliže** NEPRÁVÁ PODMÍNKA] se někteří levicoví politici snaží porovnávat úroveň dotací u nás a v EU,>
 <Arg1: je to velmi zavádějící.> (PDT 3.5)

Posledním ze vztahů kauzálních je *účel*. U něj argument 2 vyjadřuje účel, proč probíhá děj argumentu 1 (příklady 26–28).

- (26) <Arg2: [**Abychom** ÚČEL] si já a manžel mohli založit termínovaný vklad,>
<Arg1: vybrali jsme tenkrát všechny úspory ze spořitelny.> (PDT 3.5)
- (27) <Arg1: Volební urny byly dokonce průhledné.>
<Arg2: [**Tak, aby** ÚČEL] bylo vidět, že tentokrát neobsahují žádné hlasovací lístky vhozené ještě před příchodem prvního voliče.> (PDT 3.5)
- (28) <Arg1: Známo je, že v roce 1959 irská vláda nabídla zahraničním firmám nesmírně výhodné podmínky podnikání v postižené oblasti.>
<Arg2: [vsunuto: **účelem bylo, aby** IMPLICITNÍ ÚČEL] Cizí kapitál měl pomoci vytvořit nové pracovní příležitosti a zároveň přispět k oživení vzdušného provozu.>
(PDIT-EDA 1.0)

Další rozsáhlou skupinu tvoří vztahy **kontrastivní**, mezi něž se řadí vztah *konfrontace*, *opozice*, *restriktivní opozice*, *nepravý kontrast*, *připustka*, *rektifikace* a *gradace*. Těmto vztahům je společný význam určité protikladnosti či porovnávání obsahu obou argumentů.

V případě *konfrontace* jsou v obou argumentech postaveny do kontrastu určité entity, jimž se připisují rozdílné vlastnosti. Typickým diskurzním konektorem konfrontačního vztahu je např. spojka *kdežto* či *zatímco*. Jako argument 1 je označen ten argument, který stojí v textu vlevo (příklady 29–30).

- (29) <Arg1: Klasická teplárna nebo kondenzační elektrárna nemá účinnost vyšší než 33 %.>
<Arg2: Kogenerační jednotka má [**naopak** KONFRONTACE] účinnost až 87 %.>
(PDT 3.5)
- (30) *V ČR dnes operuje 57 bank.*
<Arg1: Pouze čtyři z nich však existovaly již před listopadem 1989.>
<Arg2: [vsunuto: **kdežto** IMPLICITNÍ KONFRONTACE] Ostatní začínaly na absolutní nule.> (PDIT-EDA 1.0)

Také ve vztahu *opozice* jsou postaveny obsahu diskurzních argumentů do kontrastu, tento vztah však na rozdíl od konfrontace neobsahuje srovnání dvou entit obsažených v argumentech. Typickým diskurzním konektorem tohoto vztahu je spojka *ale*; jako argument 1 se značí ten, který se v textu vyskytuje dříve (příklady 31–32).

- (31) <Arg1: Lidé chtějí platit jen to, co skutečně spotřebovali.>
<Arg2: Ještě dlouho tomu tak [**ale** OPOZICE] patrně nebude.> (PDT 3.5)
- (32) <Arg1: Popravdě řečeno, z Bohumína jsem měl strach,> přiznal po zápase trenér slováckého týmu Antonín Juran.
<Arg2: Obavy [vsunuto: **však** IMPLICITNÍ OPOZICE] asi nebyly namístě,> vždyť Uherské Hradiště přesvědčivě zvítězilo 4:0. (PDIT-EDA 1.0)

Ve vztahu *restriktivní opozice* je přítomen významový rys omezení platnosti obsahu jednoho z argumentů. Argument vyjadřující omezení či výjimku je značen jako argument 2 (příklady 33–34).

- (33) <Arg1: *Každá krajina má svou krásu.*>
 <Arg2: [**Jenom** RESTRIKTIVNÍ OPOZICE] *ji musíte umět vidět.*> (PDT 3.5)
- (34) <Arg1: *To by dovolilo vlastníkům domů, aby se „zbavili“ nepohodlných nájemníků.*>
 <Arg2: [vsunuto: **ale** IMPLICITNÍ RESTRIKTIVNÍ OPOZICE] *Samozřejmě při splnění dalších právem požadovaných skutečností.*> (PDIT-EDA 1.0)

U vztahu *nepravého kontrastu* nestojí vůči sobě v opozici obsahy vyjádřených vět, nýbrž nevyjádřených inferencí. V příkladu 35 by fakt, že má firma rozpracováno mnoho zakázek, naznačoval, že firma bude mít velký zisk. Na druhou stranu však zákazníci vyvíjejí tlak na snižování cen, a to může zisk naopak snížit. Ve vlastní opozici pak stojí tyto dva možné vývoje výše zisku. Jako argument 1 je u *nepravého kontrastu* značen první argument v toku textu. (Implicitní forma *nepravého kontrastu* není v našich datech doložena.)

- (35) <Arg1: *Podle vedoucího výroby Miloše Přiklopila má Seba rozpracovanou celou řadu zakázek.*>
 <Arg2: *Zákazníci [však NEPRAVÝ KONTRAST] vyvíjejí velký tlak na snižování cen tkanin.*> (PDT 3.5)

V *přípustkovém vztahu* je v argumentu 1 vyjádřen obsah, který probíhá navzdory očekávání vyvolanému obsahem argumentu 2 (příklady 36–37). Typickým diskurzním konektorem s významem *přípustky* je např. spojka *ačkoliv* či *přesto*.

- (36) <Arg2: *Když přistál letoun v Montrealu, ležela kolem letiště asi metrová vrstva sněhu a foukal silný vítr.*>
 <Arg1: [**Přesto** PŘÍPUSTKA] *fungoval letecký i silniční provoz normálně.*> (PDT 3.5)
- (37) *Loňské léto bylo pro trenéra Jurana obtížné, přestože dovedl Drnovice do ligy,* <Arg2: *byl na hodinu vyhozen> a na jeho místo tehdy nastoupil trenér Dejmal.*
 <Arg1: [vsunuto: **přesto** IMPLICITNÍ PŘÍPUSTKA] *Za pár dnů už vedl přípravu v Uherském Hradišti:*
Po nechutné záležitosti v Drnovicích, kdy mi bylo den před zahájením letní přípravy oznámeno, že končím, jsem si chtěl chvíli od fotbalu odpočinout. (PDIT-EDA 1.0)

Také u vztahu *rektifikace* jsou obsahy obou argumentů postaveny proti sobě: obsah argumentu 1 je negován, vyvrácen, v argumentu 2 se pak uvádí náležitá informace či uvedení na pravou míru (viz příklady 38–39). Charakteristickým diskurzním konektorem s významem *rektifikace* je spojka *nýbrž* či výraz *naopak*.

- (38) <Arg1: *Sestupná tendence porodnosti přitom není jevem poslední doby,*>
 <Arg2: [**ale** REKTIKACE] *výrazně se prohlubuje od poloviny 60. let.*> (PDT 3.5)
- (39) <Arg1: *Můj otec byl lékař a ve společnosti, ve které jsem vyrůstala, nebylo zvykem užívat ve vztahu k člověku slova zadek. Lapidárnější výraz na pět písmen nepřicházel vůbec v úvahu.*>
 <Arg2: [vsunuto: **místo toho** IMPLICITNÍ REKTIKACE] *Mluvílo se poněkud nepřesně o bocích, a pokud se chtěla ona tělesná partie, kde se záda přestávala slušně jmenovat, označit výstižněji, přišlo ke cti ono nelibozvučné slovo hýždě, které pana Měšťana tak zaujalo.*> (PDfT-EDA 1.0)

Poslední ve skupině vztahů kontrastivních je *gradace*. U tohoto vztahu se srovnávají obsahy dvou argumentů, z nichž jeden vyjadřuje vyšší míru či intenzitu jisté vlastnosti popisované v druhém z nich (viz příklady 40–41). Jako argument 1 se označuje ta část, jež vyjadřuje nižší stupeň. Typickými gradačními diskurzními konektory jsou např. *navíc* či *dokonce*.

- (40) <Arg1: *Porušení zákona může zprostředkovatele zahraniční pojišťovny přijít až na 10 milionů korun pokuty.*>
 <Arg2: *Ti, kteří se v této činnosti angažují, se [**navíc** GRADACE] mohou dostat do rozporu s devizovým zákonem.*> (PDT 3.5)
- (41) *Nosíme s sebou [matce do domova důchodců] zubní pastu s kartáčkem, protože tohle k normě ranní toalety nepatří a s jednou ochrnutou rukou ani sama nevytlačí pastu.*
 <Arg1: *Prostředí je ubíjející.*>
 <Arg2: [vsunuto: **navíc** IMPLICITNÍ GRADACE] *Předpisy organizace jsou nepružné – není možné ani připlatit za nějakou péči navíc – např. to vyčištění zubů, uvaření dobrého čaje nebo kávy, více rehabilitačních procházek, aby na ni někdo občas promluvil.* (PDfT-EDA 1.0)

Poslední skupinu vztahů tvoří **vztahy navazovací**, mezi něž řadíme *konjunkci* a *kontinuaci*, *konjunktivní* a *disjunktivní alternativu*, *exemplifikaci* a dále tři vztahy, které vyjadřují tutéž skutečnost na různých rovinách obecnosti: *specifikaci*, *ekvivalenci* a *generalizaci*.

Vztah *konjunkce* vyjadřuje prosté spojení obsahu dvou argumentů, které jsou na sobě významově navzájem nezávislé, mezi nimiž však existuje zjevná koherenční vazba, platí oba současně. Typickým diskurzním konektorem s konjunktivním významem je spojka *a*. V anotaci se jako argument 1 označuje ta část, která se jako první objevuje v textu (příklady 42–43).

- (42) <Arg1: *Dánsko plně podporuje budoucí zapojení ČR do Unie.*>
 <Arg2: *Obě strany jsou [**také** KONJUNKCE] spokojeny s růstem vzájemného obchodu.*> (PDT 3.5)

- (43) *V Plzni je stánkařům k dispozici tržnice pronajatá soukromé firmě Bratři Flaxové.*
 <Arg1: Prodává se také na náměstí,>
 <Arg2: [vsunuto: **a** IMPLICITNÍ KONJUNKCE] *od prodejců vybírá poplatek každé ráno správce tržště.*> (PDiT-EDA 1.0)

Také u vztahu *kontinuace* platí obsahy obou argumentů současně, nicméně koherenční vazba mezi nimi je slabší. Argument, který se v textu vyskytuje jako první, je označován jako argument 1; druhý argument pak na něj prostě navazuje, uvádí další informaci k témuž tématu, která může být jiného druhu či souviset s obsahem předchozího argumentu jen volněji (příklady 44–45).

- (44) *Do zástavy může přitom být vždy dána jen taková pohledávka, podle níž má být věřiteli poskytnuta nějaká věc, právo nebo jiná majetková hodnota –*
 <Arg1: zpravidla bývá zastavována pohledávka peněžní.>
 <Arg2: [Jiným KONTINUACE, SEKUNDÁRNÍ KONEKTOR] *zvláštním druhem zástavního práva je zastavení cenných papírů.*> (PDT 3.5)
- (45) *Se zájmem jsem si přečetl pojednání Arno Paříka o Karlu Poláčkovi.*
 <Arg1: Děkuji za sérii povídání o Osvětami.>
 <Arg2: [vsunuto: **k tomu dále uvádím, že** IMPLICITNÍ KONTINUACE] *Jsem rád, že se konečně píše o tom, jak to skutečně bylo.*> (PDiT-EDA 1.0)

Dalšími dvěma vztahy jsou vztahy alternativní. U *konjunktivní alternativy* jsou obsahy obou argumentů předloženy jako možnosti, u nichž může dojít k výběru, ale mohou také platit zároveň (příklady 46–47). Argument 1 je v anotaci ten z argumentů, který se v textu vyskytuje dříve.

- (46) <Arg1: Obdobně lze této možnosti [zastupování účastníka řízení notářem] využít například u daňového přiznání souvisejícího s převodem vlastnictví k nemovitému majetku.>
 <Arg2: [Případně KONJUNKTIVNÍ ALTERNATIVA] *notář může klienta zastupovat v daňovém řízení u příslušného finančního úřadu.*> (PDT 3.5)
- (47) *Trh s těmito právy existuje v té nejhorší poloegální a nelegální podobě.*
 <Arg1: Nájemníci byty vyměňují,>
 <Arg2: [vsunuto: **nebo** IMPLICITNÍ KONJUNKTIVNÍ ALTERNATIVA] *nelegálně pronajímají.*> (PDiT-EDA 1.0)

U *disjunktivní alternativy* se pak možnosti výběru navzájem vylučují: buď platí jedna, nebo druhá (příklady 48–49). Argument 1 opět stojí v textu vlevo.

- (48) <Arg1: Vlády na krátko se termín blokáce zákona o bankrotu zřejmě jednoduše zapomněly.>
 <Arg2: [Nebo DISJUNKTIVNÍ ALTERNATIVA] *mu nevěnovaly dostatečnou pozornost.*> (PDT 3.5)

- (49) <Arg1: *Na obchodní místo RM-S je nejlépe dostavit se osobně s platným občanským průkazem.*>
 <Arg2: [vsunuto: **nebo** IMPLICITNÍ DISJUNKTIVNÍ ALTERNATIVA] *Podáním obchodního pokynu můžeme někoho pověřit,> ale je nutné ho vybavit notářsky ověřenou plnou mocí k tomuto úkonu.>* (PDIT-EDA 1.0)

U vztahu *exemplifikace* obsahuje jeden z argumentů (označovaný jako argument 2) příklad toho, o čem se obecněji mluví v argumentu 1 (příklady 50–51). Typickým diskurzním konektorem *exemplifikace* je výraz *například*.

- (50) <Arg1: *Jde o ceny průměrné, které se začnou v rámci státu výrazně lišit podle druhu paliva, používaného k výrobě tepla.*>
 <Arg2: [**Například** EXEMPLIFIKACE] *v Brně lze očekávat letos průměrnou cenu až 180 Kč.>* (PDT 3.5)
- (51) *Vláda nyní zapracovává do svého návrhu novely zákona o dluhopisech možnost krytí hypotečních zástavních úvěrů také zástavním právem k budované nemovitosti. ... Při realizaci zákona bude záležet na tom, zda a* <Arg1: *v jaké formě se bude na podpoře hypotečních úvěrů podílet stát,> uvedl náměstek ministra financí Vladimír Rudlovčák, který novelu v parlamentu zdůvodňoval.*
 <Arg2: [vsunuto: **například** IMPLICITNÍ EXEMPLIFIKACE] *Novela zákona o daních z příjmů osvobodila úrokové výnosy z hypotečních zástavních listů od této daně,> připomněl.* (PDIT-EDA 1.0)

U následujících tří vztahů představuje vždy jeden z argumentů jiné vyjádření obsahu druhého argumentu. V případě *specifikace* (příklady 52–53), pro niž je charakteristickým konektorem příslovce *konkrétně*, obsahuje argument 1 obecnější vyjádření, zatímco v argumentu 2 uvádí nějaký jeho detail (nikoli však jako příklad, na rozdíl od *exemplifikace*).

- (52) <Arg1: *V souladu se západními vzory je možná i omezená preference soukromého pojištění před sociálním pojištěním.*>
 <Arg2: [**Konkrétně** SPECIFIKACE], *pokud si výdělečně činná osoba zaplatí dostatečně vysoké soukromé pojištění, bude se moci ze sociálního pojištění „vyvoléknout“.>* (PDT 3.5)
- (53) <Arg1: *Teprve po uvolnění embarga na dovoz špičkové technologie si uvědomujeme míru našeho zaostávání.*>
 <Arg2: [vsunuto: **konkrétně** IMPLICITNÍ SPECIFIKACE] *V plné míře se to týká telekomunikačních zařízení.>* (PDIT-EDA 1.0)

U vztahu *ekvivalence* je obsah jednoho argumentu jiným vyjádřením obsahu druhého argumentu, na stejné úrovni obecnosti či specifičnosti (příklady 54–55). Význam ekvivalence vyjadřují např. diskurzní konektory *jinými slovy* či *lépe řečeno*; argumentům jsou v anotaci přiřazována čísla podle pořadí, v jakém se argumenty objevují v textu.

- (54) <Arg1: Od 1. srpna jsou nové akcie Ekoagrobanky prodávány s třicetiprocentním emisním áziem.>
 <Arg2: Investor [**tak** EKVIVALENCE] tisícikorunové akcie může nakoupit nejméně za 1300 korun.> (PDT 3.5)
- (55) <Arg1: Teplíctí si dobře uvědomují, že těžiště jejich činnosti musí být v léčení tuzemských pacientů.>
 Těm je dáována přednost před zahraničními hosty, a to i přesto, že cizinci zaplatí za léčebný pobyt až čtyřikrát více, než dostanou lázně od zdravotní pojišťovny za pobyt tuzemského pacienta.
 <Arg2: [vsunuto: **jinými slovy** IMPLICITNÍ EKVIVALENCE] Vedení lázní si dobře uvědomuje, že zahraniční klientela může tvořit jen doplňkovou, nikoli rozhodující část zákazníků.> (PDiT-EDA 1.0)

Posledním vztahem ve významové klasifikaci je vztah *generalizace*. U něj je významový poměr mezi argumenty opačný než u specifikace: argument 1 vyjadřuje určitý obsah konkrétněji, argument 2 ho pak zobecňuje (příklady 56–57). Význam generalizace má např. diskurzní konektor *zkrátka*, *celkově* či *tedy*.

- (56) <Arg1: [Slovo klika] znamená jednak „držadlo k otvírání dveří, oken apod.“, nebo také „rameno s rukojetí k otáčení různých strojů“, jednak „skupinu vzájemně se podporujících jednotlivců se sobeckými cíli“ a konečně „štěstí“. Je zřejmé, že se zde spojily výrazové prostředky pocházející z různých zdrojů.>
 <Arg2: Jde [**tedy** GENERALIZACE] o trojnásobné homonymum.> (PDT 3.5)
- (57) Další důvody pro zahájení výroby byly:
 <Arg1: * Zámková dlažba, která zaplavila trh před čtyřmi lety, je příliš jednotvárná a nedává architektovi dostatek možností pro estetické ztvárnění ploch.
 * Nový typ dlažby má velkou pevnost oproti jiným typům (vyrábí se tlakem 600 tun) i při malé tloušťce a rozměrech 40x40, resp. 40x60 cm.
 * Cena za m² může být srovnatelná s cenou za zámkovou dlažbu.
 * Nový typ dlažby je možné vhodně prostorově kombinovat se zámkovou dlažbou okolo stromů, na okrajích chodníků a jinde.>
 <Arg2: Vše [vsunuto: **tedy** IMPLICITNÍ GENERALIZACE] hovořilo pro.> (PDiT-EDA 1.0)

4.3.2 Od Pražského závislostního korpusu po Rozšířenou anotaci diskurzních vztahů (PDiT-EDA 1.0)

Data, která jsme vytvořili pro výzkum implicitních vztahů v češtině, navazují na korpusy, které vznikly v Ústavu formální a aplikované lingvistiky na MFF UK v předchozích letech. Na základě Pražského závislostního korpusu vznikl Pražský diskurzní korpus 1.0, oba korpusy se pak vzájemně ovlivňovaly v dalších verzích. Pražský diskurzní korpus 2.0 posloužil jako východisko pro naše data, tzv. Rozšířenou anotaci diskurzních vztahů v Pražském diskurzním korpusu 1.0 (Enriched Discourse

Annotation of PDiT Subset 1.0, Zikánová et al. 2018). Aby byly zřejmé souvislosti mezi těmito korpusy a jejich verzemi, podáváme níže stručný přehled, jak na sebe jednotlivé verze navazují a co obsahují.¹⁴

Pražský závislostní korpus 2.5 (Prague Dependency Treebank 2.5)

Data Pražského závislostního korpusu 2.5 (PDT 2.5; Bejček et al. 2011) se stala východiskem pro anotaci diskurzních vztahů. Texty PDT pocházejí z českých novin 90. let 20. století (Mladá fronta Dnes, Lidové noviny, Českomoravský profit). Celkový rozsah PDT je přibližně 50.000 vět v 3.165 textech. PDT obsahuje vícerovinnou anotaci textu (viz podkap. 4.2), která zahrnuje morfologii, analytickou syntax, syntaktickou sémantiku, aktuální členění, koreferenci a asociační anaforu.

Pražský diskurzní korpus 1.0 (Prague Discourse Treebank 1.0)

Materiál Pražského závislostního korpusu jako celek posloužil jako vstupní data pro Pražský diskurzní korpus 1.0 (PDiT 1.0; Poláková et al. 2012a), v němž byla provedena anotace explicitních diskurzních vztahů. Tato anotace je začleněna do analýz vět na tektogramatické rovině a vedle koreference a asociační anafory je dalším typem analýzy, která přesahuje hranice věty.

Pražský závislostní korpus 3.5 (Prague Dependency Treebank 3.5)

Anotace diskurzních vztahů v PDiT 1.0 byly zahrnuty do nového vydání Pražského závislostního korpusu (PDT 3.0/PDT 3.5, Hajič et al. 2018), které obsahuje řadu oprav na různých rovinách, úpravu anotace explicitních diskurzních vztahů, doplněnou anotaci koreference a také nově vytvořenou anotaci žánrů jednotlivých textů.

Pražský diskurzní korpus 2.0 (Prague Discourse Treebank 2.0)

Nová verze Pražského diskurzního korpusu (PDiT 2.0; Rysová, M. et al. 2016) vychází z PDT 3.0/PDT 3.5 a přinesla vedle oprav dosavadních dat významné doplnění anotace: obsahuje totiž ucelenou koncepci a anotaci diskurzních vztahů vyjádřených sekundárními, tj. víceslovnými diskurzními konektory.

Rozšířená anotace diskurzních vztahů na části Pražského diskurzního korpusu 1.0 (Enriched Discourse Annotation of PDiT Subset 1.0)

V Rozšířené anotaci diskurzních vztahů (PDiT-EDA 1.0; Zikánová et al. 2018) byla použita část anotovaných textů z PDiT 2.0 (viz tab. 4.8).¹⁵ Na nich pak byla provedena

¹⁴ Srov. kol. autorů, <https://ufal.mff.cuni.cz/pdit2.0>, cit. 20. 4. 2021.

¹⁵ U počtu tokenů v tabulce odkazujeme na počet uzlů na analytické rovině, protože jde o údaj, která nejlépe vystihuje počet slov. Analytické uzly totiž reprezentují autosémantika i synsémantika v textu; pro každou větu jsou uspořádány v tzv. analytických stromech, viz podkap. 4.2.

	PDiT-EDA 1.0	PDiT 2.0
Počet textů	100	3 165
Počet vět	2 592	49 431
Počet tokenů (analytických uzlů)	41 877	833 193

Tabulka 4.8: Rozsah PDiT-EDA 1.0 a PDiT 2.0

anotace implicitních diskurzních vztahů a dále vztahů založených na entitách, otázek a odpovědí, přerывů v koherenci, přiřazení argumentů mluvčímu a dalších jevů souvisejících se stavbou textu.

4.4 Korpus PDiT-EDA 1.0 (Rozšířená anotace diskurzních vztahů)

Korpus PDiT-EDA 1.0 jsme vytvořili jako souhrn dat pro výzkum představený v této práci, proto zasluhuje na tomto místě podrobnější popis. Korpus PDiT-EDA 1.0 převzal přibližně 5 % textů z korpusu PDiT 2.0 (viz tab. 4.8), obsahuje celkově 100 dokumentů o 2 592 větách.

Při volbě textů pro PDiT-EDA 1.0 se stal klíčovým vodítkem textový žánr. U analýzy založené na textech Pražského závislostního korpusu a korpusů, které z něj vycházejí, je totiž třeba mít na paměti, že tyto korpusy obsahují pouze psané texty žurnalistické, jejichž jazyk může být v jistém směru specifický. Zároveň ale jde o poměrně pestrout sbírku textových žánrů, která zahrnuje např. zprávy, recenze, interview, předpovědi počasí či televizní programy. Pro PDiT-EDA 1.0 jsme proto volili takové žánrové zastoupení textů, aby byly v korpusu představeny texty s různými hlavními stylistovnými faktory, u nichž lze také očekávat variabilitu v typických jazykových jevech.

Tak např. pro některé žánry zastoupené v korpusu je příznačná určitá komunikativní funkce, může jít o funkci informativní (u žánru *zpráva*, *popis*, *sbírka krátkých zpráv*, *přehled*), funkci argumentativní (u *komentáře*, *recenze*, *eseje* či *dopisů čtenářů*) či funkci persvazivní (typickou pro *radu čtenářům* nebo *pozdávku*). Pro další žánry je charakteristické téma a jeho vývoj v textu, to platí pro *sportovní zprávy* a *předpověď počasí*. Stavba textu může být ovlivněna také počtem mluvčích – liší se od sebe texty monologické a dialogické. Texty v novinách jsou převážně monologické, nicméně některé žánry jsou založeny na dialogičnosti. K nim patří např. *interview s osobností*, *interview k tématu* či *čtenářská anketa*. Je třeba podotknout, že jde nikoli o klasické dialogy se dvěma či více autory replik, nýbrž o dialogy konceptuální, v nichž jeden z autorů původního dialogu (žurnalista) posléze upravuje repliky ostatních do tisku. V tab. 4.9 je představeno 15 textových žánrů, které byly vybrány pro korpus PDiT-EDA 1.0.

Textový žánr	Počet textů	Počet vět
Čtenářská anketa	8	163
Dopis čtenáře	6	176
Esej	4	170
Interview k tématu	7	291
Interview s osobností	2	104
Komentář	6	171
Popis	4	170
Pozvánka	7	170
Předpověď počasí	5	91
Přehled	6	171
Rada čtenářům	10	200
Recenze	7	167
Sbírka krátkých zpráv	4	163
Sportovní zprávy	9	196
Zpráva	15	189
Celkem	100	2 592

Tabulka 4.9: Zastoupení textových žánrů v korpusu PDiT-EDA 1.0

Jednotlivé žánry jsou podrobně charakterizovány v článku Poláková et al. (2014, s. 1322, tab. 2) a Mikulová et al. (2013, s. 33–36), a to na základě distinktivních formálních, obsahových a pragmatických rysů. V našem vysvětlení se omezíme na popis pouze vybraných žánrů, jejichž rozlišení by mohlo být nejasné. *Zprávou* rozumíme krátký text o aktuálních událostech, které autor popisuje z objektivní perspektivy. Jiným žánrem s objektivní perspektivou je *popis*: ten zachycuje objekt či událost, která neprobíhá v aktuálním čase (např. popis zámku či historické události). *Sbírka krátkých zpráv* představuje zřetězení objektivních zpráv, které spolu navzájem nesouvisejí – typický název sbírky by byl např. *Minulý týden v zahraničí*. Někdy bývají objektivně podané zprávy také navzájem tematicky propojeny a řazeny v charakteristických textových strukturách (jako např. seznam) – takový textový útvar nazýváme *přehled*.

U další skupiny žánrů žurnalistického stylu podávají autoři informace ze své subjektivní perspektivy. K těmto žánrům patří např. *krátký komentář* či rozsáhlejší *esej*. Zvláštním žánrem je pak *recenze*, která mívá specifické téma (zprav. z oblasti kultury) a hodnotí nějakou aktuální událost.

4.4.1 Implicitní diskurzní vztahy

Na textech zařazených do korpusu PDiT-EDA 1.0 pak byla provedena anotace implicitních diskurzních vztahů. Anotátoři měli za úkol vyznačit implicitní diskurzní

vztah, pokud mezi argumenty identifikovali významový vztah, který odpovídá jednomu ze vztahů v sémantické klasifikaci, a zároveň pokud povaha samotných potenciálních argumentů odpovídala vymezení diskurzního argumentu (tj. typicky nešlo o člen, nýbrž o klauzi). Jako pomůcka přitom anotátorům sloužilo ověření pomocí doplnění konektoru, který jednak poukazuje na existenci diskurzního vztahu vůbec, jednak napomáhal přesnějšímu určení významu diskurzního vztahu. Zvolený konektor, případně delší vysvětlení, zapisovali anotátoři u jednotlivých šipek implicitních vztahů. U každého anotovaného diskurzního vztahu byl vyznačen také rozsah obou diskurzních argumentů. Rozsah nebyl předem nijak omezován; pokud podle anotátorů existoval vztah mezi určitým argumentem a stavbou textu jako celku, zachycovali tento jev pomocí značky *makrostruktura* (*macrostructure*).

Anotovány byly vztahy vnitrovětné a mezivětné. U *vnitrovětného vztahu* se hranice mezi argumenty nachází uvnitř jednoho tektogramatického stromu. *Mezivětným vztahem* pak rozumíme takový vztah, jehož výchozí a cílový uzel se nacházejí v různých tektogramatických stromech, tj. hranici mezi argumenty tvoří grafické znaménko označující konec věty (případně více takových znamének, pokud argumenty neleží v textu vedle sebe). Argumenty nemusejí v textu bezprostředně sousedit a mohou se nacházet i v různých odstavcích.¹⁶ Toto řešení – omezovat anotátory ve vyznačování různých struktur co nejméně – se nám osvědčilo již při předchozích fázích anotace u explicitních diskurzních vztahů, PDiT 1.0, Poláková et al. 2012a). Ukázalo se totiž, že v anotaci nedochází k excesivnímu propojování všeho se vším, nýbrž naopak, stavba samotného jazykového materiálu dosahuje jen jisté úrovně složitosti a ta nebývá překračována. Zároveň nám tato volnost umožňuje odhalit a konzistentně popsat jevy, které jsme nepředpokládali. K nim patří např. větosled přímé a autorské řeči v delších sekvencích, kde propojování vzdálených argumentů umožňuje anotovat diskurzní vztahy mezi jednotlivými částmi přímé řeči oddělenými řečí autorskou.

U vnitrovětných vztahů sloužila jako podklad pro určování diskurzních vztahů anotace tektogramatické roviny, na níž jsou klasifikovány významové typy vztahů koordinačních. Anotátoři měli za úkol v těchto případech tektogramatickou anotaci ověřit a buď přenést do anotace diskurzu, nebo ji v anotaci diskurzu změnit. (K tomu může dojít např. na místech, kde je možné dvojí čtení textu.) Jako implicitní diskurzní vztah hodnotíme také spojení vedlejší věty vztahné s větou řídící, pokud je význam věty vedlejší samostatný a jedná se o tzv. nepravou větu vedlejší. Proto u vedlejších vět vztahných ověřovali anotátoři jejich významový vztah k větě řídící a v náležitých případech mezi oběma větami značili implicitní diskurzní vztah.

¹⁶ Toto rozlišení uvádíme proto, že by mohlo dojít k nedorozumění u rozsáhlých argumentů, skládajících se z více vět. Teoreticky může existovat např. diskurzní vztah příčiny uvnitř podřadného souvětí s VV příčinnou, kde však příčina je vyjádřena nejen ve větě vedlejší, ale ještě v následující větě, která na větu vedlejší významově úzce navazuje (tj. jeden z argumentů se skládá z více vět či jejich částí). V souladu s vymezením uvedeným výše považujeme tento vztah za vnitrovětný, protože hranice mezi argumenty probíhá uvnitř souvětí.

4.4.2 Další druhy vztahů

Při rozpoznávání implicitních diskurzních vztahů je nutné také odlišit tyto vztahy od dalších typů vztahů, které se mohou vyskytovat mezi jednotlivými částmi textu. Aby bylo zřejmé, že anotátoři nepřehlédli potenciální implicitní vztah, anotovali také tyto jiné způsoby začlenění částí do textu; úkolem anotátorů proto bylo pro kontrolu nějakým způsobem označit všechny hranice mezi větami (mezivětné vztahy).

Součástí tohoto úkolu bylo např. doplnit anotaci vztahů explicitních: pokud se ukázalo, že argumenty jsou spojeny explicitním konektorem, který z různých důvodů nebyl zahrnut do dosavadní anotace explicitních vztahů, byl v textu vyznačen explicitní diskurzní vztah. To se týkalo zejména dosud neidentifikovaných sekundárních konektorů, které nebylo možné zachytit automatickou kontrolou textu v předchozích vydáních různých diskurzních anotací.

Dalším druhem vztahu vyznačovaným při doplňkové anotaci jsou vztahy založené na entitách (entity-based relation, EntRel).¹⁷ Tyto vztahy v zásadě vytvářejí podmnožinu dříve anotovaných vztahů koreference a asociační anafory. Typicky se vyskytují mezi jmennými skupinami, ale také mezi jmennou skupinou a větou či mezi rozsáhlejším úryvkem textu a jmennou skupinou. Do anotace vztahů mezi skupinami byly převzaty zdaleka ne všechny vyznačené koreferenční a asociační vztahy, ale pouze takové, které vytvářejí silný koherenční mezivětný vztah (srov. příklady 58–62). Pro dvojice takových vět bývá také příznačný jistý pevně daný průběh aktuálního členění, např. druhá věta nějak specifikuje jmennou skupinu, která byla uvedena v rématu věty první (tematizace rématu).

- (58) <Arg1: *Firmy jdou za levnou, leč zkušenou či učenlivou pracovní silou a za výhodami* ENTITA, které by doma nedostaly.>
<Arg2: Patří *mezi ně* ENTITA různé úlevy, jako například bezcelní dovoz anebo lacinější služby, avšak nejvíce jim jde o daňové prázdniny, které mohou trvat až patnáct let.> (PDIT-EDA 1.0)
- (59) <Arg1: *Asrael – Rozměrná pětivěťá symfonie* ENTITA o smrti, inspirovaná úmrtím skladatelových blízkých, byla dokončena roku 1906.>
<Arg2: *Skladba* ENTITA bohatá na melodiku, harmonii i instrumentaci, patřící k vrcholům Sukova díla, došla ve světě ocenění.> (PDIT-EDA 1.0)
- (60) *Přesto má katedrála od loňského roku nového varhaníka, Martina Porubu.*
<Arg1: *Ten* ENTITA také v neděli sérii varhanních koncertů otevře.>
<Arg2: *Poruba* ENTITA (1951) byl varhaníkem u sv. Mikuláše v Mnichově a od roku 1988 působí jako docent na mnichovské Akademii múzických umění.> (PDIT-EDA 1.0)

¹⁷ Převzato z terminologie PDTB.

- (61) <Arg1: „Lukáš škrtl pádlem o kámen a **ztratil rovnováhu** _{DĚJ},“> řekl trenér Miroslov Nedvěd.
<Arg2: „Stává se **to** _{ENTITA}.“> (PDiT-EDA 1.0)
- (62) <Arg1: Gyula Horn se vyslovil pro možné **zavedení majetkové daně** _{DĚJ}.>
<Arg2: Zdůvodnil **to** _{ENTITA} tím, že utahování opasků se nemůže vztahovat pouze na lidi žijící ze mzdy.> (PDiT-EDA 1.0)

Pro srovnání uvádíme také příklad koreferenčního vztahu, který nevytváří silný koherenční mezivětný vztah, jenž bychom vyznačovali zvlášť jako vztah mezi entitami, a který jsme proto nepřevzali z anotace koreference (příklad 63). Jde často např. o opakování téhož slova v různých, zejména nerematických pozicích v obou větách nebo o koreferenci podmětů či vypuštěných zájmen v různých pozicích.

- (63) *Takže jak jsme na tom – jsme kvalifikovaní pro Evropu?*
My jsme z ní nikdy neodešli, a přesto se musíme hodně učit.
[My] Můžeme Evropě hodně dát a ona by měla hodně dát **nám**. (PDiT-EDA 1.0)

Další typy vztahů byly v PDiT-EDA 1.0 anotovány úplně nově, nelze je získat z předchozích anotací. Jak již bylo řečeno, hlavním cílem anotace těchto vztahů bylo jejich odlišení od vztahů implicitních, přesto je zřejmé, že vytvářejí vlastní subsystémy v jazykové koherenci a zaslouží si podrobnější rozbor. Z tohoto pohledu může být naše anotace chápána také jako základní jazykový materiál pro budoucí výzkum těchto jevů.

Prvním z nich je vztah mezi otázkou a odpovědí. Tento vztah se odlišuje od běžných diskurzních vztahů, protože situace popsaná v otázce nemá stejnou faktičnost jako situace v ostatních větách. Stejně tak není běžným diskurzním vztahem relace mezi otázkou a kontextovou větou, která jí předchází. Toto téma s sebou přináší další podtémata, jako je střídání mluvčích, otázky bez odpovědí a jejich zapojení do textu, řečnické otázky, nepřímé otázky ad. V korpusu PDiT-EDA 1.0 se otázky vyskytují poměrně často (65 výskytů celkem); jejich anotace může sloužit jako podklad pro další výzkum.

Dalším typem anotovaných vztahů je skupina jevů souvisejících s makrostrukturou textu. Jak se ukázalo během anotace, některé diskurzní vztahy spojují nikoli sousední věty, ale rozsáhlejší úseky textu, a některé odkazují k textu jako celku (vyznačují např. začátek závěrečného shrnutí). Tyto výskyty jsou anotovány jako běžné diskurzní vztahy, ale navíc byly označeny poznámkou *makrostruktura*. Ostatně na jevy makrostruktury textu reagovala už anotace Penn Discourse Treebank 2.0, která používala značku *list* (seznam) jako jeden z typů diskurzních vztahů. Do naší anotace byla značka seznam převzata již v Pražském diskurzním korpusu a uvádíme ji také v tabulce jako zvláštní položku, a to proto, aby bylo možné srovnávat tato data s daty PDTB. Nicméně na rozdíl od PDTB nepovažujeme seznam za jeden z významů diskurzních vztahů (položky seznamu mohou být spojeny vztahy, které lze běžně zařadit podle

významové klasifikace bez zavádění zvláštní kategorie pro seznam, např. konjunkce či časová následnost).¹⁸

Myšlenka, že dříve anotovaný seznam (již v PDiT 1.0, převzato do PDiT-EDA 1.0) je vlastně jedním z podtypů šířeji pojaté makrostruktury, je v našem výzkumu úplně nová; anotace seznamů a nově zavedené makrostruktury nebyla v PDiT-EDA 1.0 dosud sjednocena. V počtech výskytů uvedených v tabulce je pouze v jednom případě ze 13 začátek seznamu označen jako makrostruktura, místy jsou tak označeny také vnitřní položky seznamů. Soustavný výzkum textové makrostruktury bude předmětem dalších projektů.

K jevům makrostruktury se řadí také *mezery v koherenci*, tj. přerušení koherence v textu, jichž se v datech PDiT-EDA 1.0 vyskytlo 25. Typicky se objevují u žánru sbírka zpráv, na předělu dvou zpráv, které spolu nesouvisejí a jsou jen volně řazeny za sebe, viz následující příklady. Vnitrovětné předěly v koherenci jsme v datech nenalezli, což ukazuje, že větný celek je vnímán jako koherentní jednotka.¹⁹

- (64) *Odborníci jsou toho názoru, že virus se dostal do vodovodní sítě z řeky Latorice.*
 <Arg1: *Nelze proto vyloučit, že se brzy objeví i na území Maďarska.*>
 <Arg2: **MEZERA V KOHERENCI** *Jednou z největších atrakcí 45. sněhového festivalu v japonském Sapporu bude zřejmě budova maďarského parlamentu vymodelovaná ze sněhu.*> (PDiT-EDA 1.0)
- (65) *Předcházející národně konzervativní koalice učinila před několika lety podobný pokus o zavedení majetkové daně.*
 <Arg1: *Pod tlakem Ústavního soudu však byla nucena své rozhodnutí odvolat.*>
 <Arg2: **MEZERA V KOHERENCI** *Za zoufalý krok osamělého člověka označilo britské ministerstvo zahraničí chování maďarského diplomata, jenž se podle britského tisku pokusil se zbraní v ruce vynutit zastavení hlučných stavebních prací poblíž svého obydlí v Londýně.*>
Záležitost byla hladce vyřešena. Pracovníci maďarského velvyslanectví posadili diplomata na první letadlo do Budapešti. (PDiT-EDA 1.0)

K jiným typům začlenění jednotek do textu, jež anotátoři označovali, patří také vztahy vyplývající ze zvláštní povahy argumentů (*otázka, titulek, autor, popis* – např. u fotografie, *metatext* či uvedení *místa*, odkud se podává zpráva). V těchto případech nebyl vyznačován vztah, nýbrž přímo u argumentu samotného se vyznačovalo, že jeho povaha je specifická. Tento postup se nám zdál jako přesnější než vyznačovat všechny jevy textové stavby pomocí vztahů (různých druhů šipek). Některé rysy jsou totiž vlastní spíše samostatným jednotkám než vztahům mezi těmito jednotkami.

¹⁸ Podrobněji se otázkám makrostruktury v PDT věnuje článek Poláková et al. 2021.

¹⁹ V anotaci značíme u mezer v koherenci argument 1 a argument 2 pouze z technických důvodů: potřebujeme naznačit, mezi kterými jednotkami se přerýv v koherenci nachází. Nejde samozřejmě o druh diskurzního vztahu, nýbrž naopak o nepřítomnost jakéhokoliv diskurzního vztahu.

	Explic. (prim.)	Explic. (sek.)	Implic.	Entrel	Seznam	Otázka – odp.	Makro	Norel
Explic. (prim.)	ano							
Explic. (sek.)	ano	ano						
Implicit	ne	ne	ne					
Entrel	ano	ano	ano	ano				
Seznam	ano	ano	ano	ano	–			
Otázka – odp.	ano	ano	ne	ano	ano	–		
Makro	ano	ano	ano	ano	ano	ano	ano	
Norel	ne	ne	ne	ne	ne	ne	ne	–

Tabulka 4.10: Povolené překryvy vztahů v anotaci PDiT-EDA 1.0

Některé jevy v textu, které ovlivňují jeho koherenci, se v datech PDiT-EDA 1.0 vyskytovaly jen zřídka, eventuálně jsme se na ně nesoustředili soustavně, ale pouze v nápadných případech. Komentář k nim jsme zachytili ve značce *diskurzní poznámka*. Tyto značky jsou pouze pracovní a zachycují jádra jevů, k nimž se můžeme vrátit v dalším výzkumu. Konkrétně jde o hodnotící výrazy, neustálené sekundární diskurzní konektory, řečnické otázky, vyznačení zdroje atribuce²⁰ a obsahu, který někdo tvrdí (značeno jako *atribuce*), dále o bibliografické údaje (*quotation:reference*) a o diskurzní argumenty s explicitním konektorem, k nimž se nepodařilo najít druhý argument ve vztahu (*one_arg*). Značka *diskurzní poznámka* a její hodnoty nebyly zařazeny do vydaných dat.

K anotačním zásadám je třeba doplnit, že zkusmá anotace ukázala, že některé druhy anotovaných vztahů se mohou překrývat, např. mezi dvěma argumenty může být vztah konjunkce, zároveň ale oba mohou být položkami seznamu. V instrukcích pro anotátory jsme proto vymezili předem, které kombinace vztahů lze anotovat jako vztahy paralelní (viz tab. 4.10). Může jít i o stejný druh vztahů: např. dva argumenty mohou být zároveň spojeny vztahem explicitním vztahem exemplifikace a příčiny (s kombinací diskurzních konektorů *například protože*).

Při vlastním anotačním postupu se ukázalo jako praktické anotovat text nejprve na jeho běžné lineární formě (obvykle v tištěné formě), bez opory v tektogramatické rovině. Tento postup umožňoval anotátorům dobře se soustředit na smysl textu a na

²⁰ Zdrojem atribuce se rozumí člověk, činitel, případně dokument, který uvádí nějaké tvrzení, srov. následující příklady:

- (66) *Zdá se, že se postoj americké diplomacie ZDROJ ATRIBUCE zjednodušil na jedinou otázku: „Kam až hodláš zajít, Slobodane?“* (PDiT-EDA 1.0)
- (67) *Odpověď srbského prezidenta Slobodana Miloševiče ZDROJ ATRIBUCE je neměnná: „Ještě dál.“* (PDiT-EDA 1.0)

pochopení návazností v rozsahu celého textu, aniž by byli rozptylováni různými otázkami syntaktické stavby a také např. ovlivňování interpretací koordinačních struktur ad. v tektogramatické anotaci.

V další fázi pak byla anotace přenášena na tektogramatickou rovinu, přičemž se zohledňoval způsob reprezentace vět na této rovině, zejména v otázce rekonstruovaných částí vět.

Před vlastním vydáním korpusu PDiT-EDA 1.0 pak byly anotace zkontrolovány a sjednoceny.

Po technické stránce byly v korpusu pro anotaci diskurzních vztahů používány značky, které lze rozdělit do tří skupin: rysy textových šipek, zachycující koherenční vztahy obecně, rysy diskurzních šipek, které popisují charakteristiky diskurzních vztahů, a rysy jednotlivých uzlů, jež vypovídají o povaze argumentu, který pak vstupuje do případných vztahů.

Rysy textových šipek mohou mít čtyři hodnoty:

- *discourse* (diskurzní vztah)
- *entrel* (vztah založený na entitách)
- *list* (seznamová struktura)
- *question – answer* (vztah otázky a odpovědi)

Textové šipky s hodnotou *discourse*, tj. diskurzní vztahy, jsou pak podrobněji charakterizovány následujícími značkami:

- *is_secondary*, s hodnotami 0 a 1 (vyjadřuje, zda je vztah vyjádřen sekundárním konektorem)
- *is_implicit*, s hodnotami 0 a 1 (označuje, zda jde o vztah implicitní)
- *connective_inserted* (používá se u implicitních vztahů, obsahuje konektor nebo vysvětlení, které na dané místo vkládají anotátoři při interpretaci)
- *connective* (používá se u explicitních vztahů s primárními a sekundárními konektory, zachycuje konektor přiřazený danému diskurznímu vztahu)
- *comment* (obsahuje jakýkoliv komentář k anotaci)
- *is_negated*, s hodnotami 0 a 1 (používá se u sekundárních konektorů, vyznačuje, zda jde o afirmativní, či negativní tvar konektoru, např. u spojení z *toho nevyplývá*, že je vyznačeno, že je v negativním tvaru)
- *discourse_type* (jeden z 23 sémantických typů diskurzních vztahů, např. *přípustka*)
- *start_range* (rozsah výchozího argumentu (Arg2), např. 1, což znamená aktuální strom či podstrom kromě vyjmutých uzlů a jeden strom navíc kromě vyjmutých uzlů)
- *target_range* (rozsah cílového argumentu, tj. Arg1)

Poslední skupinou anotovaných diskurzních rysů jsou rysy zachycované na tekto-gramatických uzlech. Tyto rysy se připisují kořeni argumentu, nejčastěji jde o uzel s funktorem PRED (tj. predicate). Jedná se o následující značky:

- *discourse_feature* (charakteristika zvláštních jednotek v textu, s hodnotami *heading, location, author, caption, metatext, question*)
- *discourse_macrostructure*, s hodnotami *macrostructure* a *coherence_gap*
- *discourse_comment*, s hodnotami *evaluation, possible altlex, rhetorical_question, attribution_source, attribution, quotation:reference, one_arg*.

Pro vyhledávání v korpusu PDiT-EDA 1.0 slouží stejné nástroje jako pro data z rodiny Pražských korpusů, tj. TrEd (Pajas - Štěpánek 2008) a PML-Tree Query (PML-TQ; Pajas - Štěpánek 2009).

4.4.3 Mezipřetřesová shoda

Anotaci korpusu PDiT-EDA 1.0 prováděli dva anotátoři, kteří měli zkušenost s anotováním explicitních diskurzních vztahů a dalších jevů v předchozím výzkumu; absolvovali také zácvik pro anotování vztahů implicitních. Během anotace probíhala kontrola, nakolik se anotátoři shodují v hodnocení jazykových jevů jako implicitních diskurzních vztahů. Pro účely kontroly jsme nechali dva anotátory označkovat některé soubory korpusu paralelně (12 textů obsahujících 233 vět). Mezipřetřesovou shodu jsme pak měřili na základě těchto dvou kritérií:

1. Shoda na existenci diskurzního vztahu mezi stejnými argumenty (tj. oba anotátoři shodně označí tutéž dvojici výchozího a cílového uzlu diskurzní šipky, nemusejí se ale shodovat ve významovém popisu diskurzního vztahu). Tuto shodu udáváme pomocí míry F1,²¹ což je harmonický průměr přesnosti (precision)²² a úplnosti (recall).²³

2. Shoda na významovém popisu diskurzního vztahu (pouze v případech, kde se oba anotátoři shodnou na existenci diskurzního vztahu, tj. kde dosáhnou shody podle bodu 1); tento typ shody jsme měřili jednak jako prostou procentuální shodu (správnost, accuracy), jednak pomocí Cohenovy kappy, která udává, o kolik procent

²¹ $F1 = 2 \frac{pr}{p+r}$

²² Přesnost (*p*) vyjadřuje podíl relevantních nalezených výskytů v rámci všech nalezených výskytů, popisuje tedy kromě jiného, jaký podíl nálezů je nevyžádaných.

²³ Úplnost (*r*) udává podíl relevantních nalezených výskytů v rámci všech relevantních výskytů, vyjadřuje tedy mimo jiné, jaký je podíl výskytů, které měly být nalezeny, avšak nebyly. V našich anotacích není stanovena ideální interpretace, podle níž by se měla posuzovat anotace obou anotátorů. Poměříme anotátory mezi sebou. Pro stanovení přesnosti tedy bereme jednu anotaci jako ideální a hodnotíme, nakolik jí druhá anotace odpovídá a co zachycuje navíc („nevyžádané nálezy“). Pro stanovení úplnosti pak bereme jako ideální druhou anotaci a poměříme, nakolik jí odpovídá první anotace a co zachycuje navíc.

Data pro paralelní anotaci	12 textů, 233 vět
Počet označených implicitních vztahů	Anotátor 1: 131 Anotátor 2: 130
F1 – shoda na existenci vztahu	0,54
Správnost – shoda na významovém zařazení v případě shody na jeho existenci	58 %
Cohenova κ u shody na významovém zařazení v případě shody na jeho existenci	0,47

Tabulka 4.11: Meziprotátní shoda u anotace implicitních vztahů v PDIT-EDA 1.0

častěji / méně často se dva anotátoři shodli při daném počtu možných kategorií, než kdyby anotace probíhala náhodně.²⁴

Ve sledovaných datech vyhledali anotátoři přibližně stejný počet implicitních vztahů: anotátor A jich označil 131, anotátor B 130. (Z toho nevyplývá, že se skupiny vyhledaných vztahů dokonale překrývají, anotátoři mohli vyhledat 130 různých vztahů a ostatní označit např. jako vztahy mezi entitami aj.)

Míra F1, udávající, nakolik se u vyhledaných vztahů anotátoři shodují na prosté existenci implicitního vztahu na daném místě v textu, je 0,54. V rámci těch implicitních vztahů, na jejichž existenci se anotátoři shodují, se pak v 58 % případů shodují i na jejich významovém zařazení, což odpovídá hodnotě Cohenovy kappy 0,47 (viz tab. 4.11).

Meziprotátní shoda v anotaci implicitních vztahů je poměrně nízká. Jako příčinu tohoto jevu vidíme jednak nemožnost opřít se o lexikální význam konektoru, jednak poměrně velké množství sémantických typů, do nichž se implicitní vztahy třídí (23 typů). Podrobněji se příčinám meziprotátní shody a srovnání s anotacemi jiných druhů vztahů věnujeme v částech 4.4.4, 4.4.5 a 4.4.6.

Při měření meziprotátní shody jsme sledovali také rozdíly mezi anotátory ve značkování implicitních vztahů mezivětných a vnitrovětných. Tyto výsledky jsou zachyceny v tabulce 4.12. Shoda pro mezivětné vztahy je podstatně vyšší než u vztahů

²⁴ V ideálním případě dosahuje Cohenova κ 1,00, tj. 100 % (anotátoři dokázali jednotně vyřešit všechny případy, které se vyřeší při náhodné anotaci, a navíc i 100 % případů, které by zůstaly při náhodné anotaci nevyřešeny). V nejhorším případě se Cohenova κ rovná -1 (anotátoři se nedokázali shodnout na ničem, a nevyřešili ani ty případy, které by vyřešila pouhá náhodná anotace). Pokud dosahují anotátoři stejného výsledku jako náhodná anotace, je Cohenova κ rovna 0.

Hodnota Cohenovy κ se počítá následovně:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = \frac{1 - p_0}{1 - p_e} \quad (4.1)$$

kde p_e je pozorovaná vzájemná shoda mezi anotátory (správnost) a p_0 je hypotetická pravděpodobnost náhodné shody (srov. Cohen, 1960; Artstein – Poesio, 2008; Hoek – Scholman 2017; wikipedia, heslo Cohen's kappa, online https://en.wikipedia.org/wiki/Cohen%27s_kappa, cit. 28. 8. 2020)

	Vnitrovětné vztahy	Mezivětné vztahy
Počet označených implicitních vztahů	Anotátor 1: 20 Anotátor 2: 34	Anotátor 1: 111 Anotátor 2: 96
F1 – shoda na existenci vztahu	0,59	0,53
Správnost – shoda na význ. zařazení v případě shody na jeho existenci	38 %	64 %
Cohenova κ u shody na význ. zařazení v případě shody na jeho existenci	0,9	0,56

Tabulka 4.12: Mezipřímá shoda v anotaci vnitrovětných a mezivětných implicitních vztahů v PDiT-EDA 1.0

vnitrovětných. Domníváme se, že nízká míra shody u vnitrovětných vztahů je dána faktem, že anotátoři označili ve sledovaných datech relativně málo vnitrovětných vztahů, přičemž toto množství zařazovali do jedné ze 23 tříd sémantických typů (tj. málo výskytů v mnoha třídách). Každý jednotlivý výskyt neshody má pak poměrně velkou váhu. Číselné údaje pro vnitrovětné vztahy proto slouží spíše jako orientační vodítko.²⁵

4.4.4 Typické příčiny neshody na existenci implicitního diskurzního vztahu

Neshoda na rozsahu diskurzních argumentů

Neshoda na existenci diskurzního vztahu má v analyzovaném materiálu několik příčin. Jedna z nich spočívá v tom, že anotátoři se shodují na tom, že ve sledované části textu existuje implicitní diskurzní vztah, odlišně ale vnímají rozsah jeho argumentů. Např. jeden anotátor pojímá jako argument celé souvětí, druhý pouze jeho část. Diskurzními šipkami pak anotátoři při analýze spojují odlišné úseky stromů, a tento způsob paralelní anotace je vyhodnocen jako neshoda anotátorů. K tomuto typu neshody dochází jak u implicitních vztahů vnitrovětných, tak i mezivětných, srov. příklad 68.

(68) Anotace 1

<Arg1: Loňské léto bylo v Uherském Hradišti dost bouřlivé,>

<Arg2: fotbalový klub [vsunuto: **totiž** IMPLICITNÍ EXPLIKACE] zůstal naprosto bez prostředků> a teprve příchod nových sponzorů přinesl obrat k lepšímu.

²⁵ Poměr vnitrovětných a mezivětných vztahů v paralelní anotaci přibližně odpovídá zastoupení těchto vztahů v celém korpusu PDiT-EDA 1.0, kde bylo označeno celkově 1 427 implicitních vztahů, z toho 163 vnitrovětných a 1 264 mezivětných, viz tab. 5.2.

Anotace 2

<Arg1: Loňské léto bylo v Uherském Hradišti dost bouřlivé,>

<Arg2: fotbalový klub [vsunuto: **totiž** IMPLICITNÍ EXPLIKACE] zůstal naprosto bez prostředků a teprve příchod nových sponzorů přinesl obrat k lepšímu.> (PDIT-EDA 1.0, paralelní anotace)

V uvedeném úryvku následuje po prvním argumentu vysvětlení, proč lze označit loňské léto v Uherském Hradišti za bouřlivé. Zatímco první anotátor zahrnuje do vysvětlení pouze fakt, že klub zůstal bez prostředků, druhý k němu připojuje i údaj o obratu k lepšímu.

K podobnému typu neshody na rozsahu diskurzních argumentů dochází také u mezivětných implicitních vztahů, srov. následující příklad. (V příkladu se liší také významové zařazení identifikovaného vztahu, o tento rozdíl však na tomto místě nejde.)

(69) Anotace 1

<Arg1: Stýkal se [Asger Jorn] s mnohými z nejuznávanějších osobností a tvůrčím způsobem rozvíjel jejich odkaz.>

<Arg2: [vsunuto: **k tomu uvádíme, že** IMPLICITNÍ KONTINUACE] Jornovo rané období ovlivňoval surrealismus.>

Anotace 2

Stýkal se [Asger Jorn] s mnohými z nejuznávanějších osobností a <Arg1: tvůrčím způsobem rozvíjel jejich odkaz.>

<Arg2: [vsunuto: **konkrétně** IMPLICITNÍ SPECIFIKACE] Jornovo rané období ovlivňoval surrealismus.> (PDIT-EDA 1.0, paralelní anotace)

V prvním případě je věta o Jornově raném období chápána jako volné navázání na celý předchozí větný celek. V druhé anotaci je pojímána jako upřesnění obecného tvrzení, že Asger Jorn tvůrčím způsobem rozvíjel odkaz uznávaných osobností. Liší se tedy rozsah prvního argumentu.

Jindy, zejména v textech s volnější stavbou a různými odbočkami, bývá obtížné určit, ke které části v toku textu přesně se vztahuje jistý diskurzní argument. Vymezení argumentu se pak může lišit nejen v syntaktickém rozsahu jako v předchozích případech (typicky celé souvětí vs. pouze jeho jistá část), ale celkově v umístění: anotátoři např. rozdílně interpretují, co je součástí hlavního toku textu a co patří spíše do tematické odbočky (viz příklad 70).

- (70) *Je smutným dědictvím, že logický požadavek zajištění volného přístupu k domovním schránkám byl u nás, na rozdíl od zahraničí, dlouhá léta ignorován, a že tudíž mnoho domů této podmínce nevyhovuje. Za to obyvatelé domu nemohou, nicméně jsou to především oni, kdo se musí na řešení vzniklých problémů podílet. Nedobrá tradice odmítat spolupráci s poštou bohužel pokračuje i nadále.*

Anotace 1

Předplatí-li si ovšem někdo dodávku novin do domu, který je v době roznášky zamčen, stěží může očekávat, že noviny budou řádně dodány.

<Arg1: V případě pisatelky navíc nejde o poštovního doručovatele, ale o doručovatele První novinové společnosti.>

<Arg2: [vsunuto: **k tomu lze uvést, že** IMPLICITNÍ KONTINUACE] *V zahraničí se pošta v takové situaci zachová nekompromisně: není-li přístup ke schránkám zajištěn, zásilky se přestanou doručovat.*>

Anotace 2

<Arg1: *Předplatí-li si ovšem někdo dodávku novin do domu, který je v době roznášky zamčen, stěží může očekávat, že noviny budou řádně dodány.*>

V případě pisatelky navíc nejde o poštovního doručovatele, ale o doručovatele První novinové společnosti.

<Arg2: [vsunuto: **k tomu lze uvést, že** IMPLICITNÍ KONTINUACE] *V zahraničí se pošta v takové situaci zachová nekompromisně: není-li přístup ke schránkám zajištěn, zásilky se přestanou doručovat.*> (PDiT-EDA 1.0, paralelní anotace)

U tohoto úryvku pojmají anotátoři odlišně začlenění poslední věty do textu. První anotátor hodnotí stavbu textu jako lineární, kde se nepřilíhově navazující informace řadí volně za sebe. Naproti tomu druhý anotátor považuje větu o doručovateli První novinové společnosti za vsuvku, a větu o zahraničí proto spojuje se vzdálenějším předchozím textem, před vsuvkou.

Význam vedlejších vět vztahných

Druhou skupinu příčin, proč se anotátoři neshodují na vymezení implicitního diskurzního vztahu v daném místě, je třeba spatřovat nikoli v rozsahu argumentů, ale v konkurenci implicitních vztahů a jiných typů vztahů. Na vnitrovětné úrovni jde často o rozdílnou interpretaci vedlejších vět přívlastkových uvedených vztahným zájmenem. Vztah mezi řídicí klauzí a závislou větou vztahnou byl anotován jako diskurzní vztah pouze tehdy, pokud děj obou klauzí je nezávislý (tzv. nepravá věta vztahná) a lze rozeznat sémantiku diskurzního vztahu. K rozporu v anotaci tu typicky dochází tak, že jeden anotátor chápe vztahnou větu jako výhradně větu přívlastkovou a vztah mezi řídicí a závislou klauzí hodnotí jako vztah čistě syntaktický, a proto ho neanotuje, kdežto druhý mezi nimi spatřuje implicitní diskurzní vztah, viz následující příklady:

- (71) <Arg1: *Vláda by podle svého návrhu dokonce měla vliv i na ústavní soud,*>
 <Arg2: [vsunuto: **protože** IMPLICITNÍ PŘÍČINA-DŮSLEDEK] *jehož třetinu by jmenovala.*> (PDiT-EDA 1.0, paralelní anotace)

- (72) <Arg1: Předplatí-li si ovšem někdo dodávku novin do domu,>
 <Arg2: který je [vsunuto: **však** IMPLICITNÍ OPOZICE] v době roznášky zamčen,>
 stěží může očekávat, že noviny budou řádně dodány. (PDiT-EDA 1.0, paralelní anotace)
- (73) <Arg1: Další úspěch zaznamenal v Drnovicích,>
 <Arg2: [vsunuto: **a to ten, že** IMPLICITNÍ SPECIFIKACE] s nimiž se před více než
 rokem dostal do ligy.> (PDiT-EDA 1.0, paralelní anotace)

(Druhá anotace v uvedených dokladech implicitní vztah nezaznamenává.)

Pro úplnost dodejme, že v dokladech se vyskytuje velké množství vztažných vět, v nichž se diskurzni vztah nevyskytuje a oba anotátoři se na tom shodují, viz příklady 74 a 75:

- (74) Přes výborný start do druhé ligy se v Uherském Hradišti drží **cíle, který** si vytkli už před sezonou – umístění v první polovině tabulky. (PDiT-EDA 1.0, paralelní anotace)
- (75) Ve státech ES je navíc právem spotřebitele, aby měl na výrobku napsáno, z čeho je složen. To mu umožňuje vyhnout se konzumaci **výrobků, které** obsahují velké množství cizorodých látek. (PDiT-EDA 1.0, paralelní anotace)

Neshoda na implicitnosti diskurzního vztahu

V další skupině dokladů anotátoři shodně spatřují diskurzni vztah, neshodují se však na tom, zda je implicitní, nebo je vyjádřen konektorem. Narážíme zde na neshodu, která vyplývá z rozdílného pojetí konektorů. V případech (možných) primárních konektorů jde zejména o to, zda chápat určitý výraz jako diskurzni konektor, nebo jako rematizátor či postojovou částici (srov. Hajičová et al. 2020). Připomeňme, že rozlišovacím znakem konektoru je schopnost spojovat dva argumenty, tj. otevřít pozice pro dva argumenty. Naproti tomu rematizátor či postojová částice se vztahují pouze k jedné větě či části ve větě. K tomuto typu neshody došlo při anotaci následujícího úryvku:

- (76) Anotace 1
 <Arg1: Neptejte se mě, proč jsem přijel do Prahy.>
 <Arg2: [vsunuto: **protože** IMPLICITNÍ NEPRÁVÁ PŘÍČINA-DŮSLEDEK] Je to přece normální sem přijet.>
- Anotace 2
 <Arg1: Neptejte se mě, proč jsem přijel do Prahy.>
 <Arg2: Je to [**přece** EXPLIKACE] normální sem přijet.> (PDiT-EDA 1.0, paralelní anotace)

V prvním případě anotátor rozvádí svoje porozumění textu vysvětlením v poznámce „protože je to přece normální sem přijet“. Slovo *přece* chápe jako postojovou částici, oba diskurzni argumenty pak podle něj nemají žádný vyjádřený spojovací prvek. Druhý anotátor hodnotí slovo *přece* jako diskurzni konektor, uvedené argumenty jsou

tedy podle něj spojeny explicitním diskurzním vztahem. (Anotátoři se zde liší také v sémantickém zařazení pozorovaného diskurzního vztahu, o tento rozdíl ale na tomto místě výkladu nejde.)

K podobnému jevu dochází u další skupiny dokladů. Anotátoři se opět shodují na tom, že mezi jistými dvěma argumenty existuje diskurzní vztah, neshodují se však na tom, zda jde o vztah implicitní, či explicitní. Jeden z anotátorů totiž interpretuje nějaký výraz či sousloví ve větě nikoli jako primární konektor (jak tomu bylo v případě slova *přece* v předchozím příkladě), ale jako konektor sekundární. Množina sekundárních konektorů je otevřená a produktivní, proto k tomuto typu neshod může docházet častěji. Tento jev ilustrují doklady 77 a 78.

- (77) *Chystaná vyhláška ministerstev financí, která umožní zvyšování nájemného nad míru inflace, snad část bytů z regulovaného trhu do trhu bez státní regulace dostane.*
 <Arg1: Jakmile někteří majitelé určí nájemné pod vládou určeným stropem, lze ho-
 vořit o tržní ceně, dané vyrovnáním nabídky s poptávkou.>

Anotace 1

<Arg2: [vsunuto: **a** IMPLICITNÍ KONJUNKCE] *V takovém okamžiku problém dědičného nároku na nájem, který pravice s levicí tak intenzivně diskutovala, ztratí význam.*>

Anotace 2

<Arg2: [**V takovém okamžiku** PŘÍČINA-DŮSLEDEK] *problém dědičného nároku na nájem, který pravice s levicí tak intenzivně diskutovala, ztratí význam.*> (PDiT-EDA 1.0, paralelní anotace)

V první anotaci je spojení vět považováno za implicitní, bez spojovacího výrazu. Anotátor zde vyznačil vztah konjunkce, protože vidí posloupnost vět jako řadu na sebe prostě navazujících dějů bez příčinné vazby.

Ve druhé anotaci je sousloví *v takovém okamžiku* považováno za synonymní s *proto*, a je tedy označeno jako sekundární diskurzní konektor. Vztah obou argumentů pak není implicitní. (Při této interpretaci významu konektoru je tu diskurzní vztah vyhodnocen jako příčinný.)

Konkurenční výklady charakteru diskurzního vztahu (vztah implicitní vs. se sekundárním diskurzním konektorem) jsme pozorovali také v následujícím příkladu:

- (78) <Arg1: *Z celkového počtu 457 319 voličů, kteří se voleb zúčastnili, hlasovalo pro blok Vlast asi 20,5 procenta, což představuje 28 poslaneckých mandátů v novém parlamentu.*>

Anotace 1

<Arg2: [vsunuto: **a** IMPLICITNÍ KONJUNKCE] *Na druhém místě s 13,9 procenty hlasů (18 mandátů) je zatím blok Spolehlivý dům dosavadního nejvyššího představitele republiky Arnolda Rüütelä.*>

Anotace 2

<Arg2: [Na *druhém místě* KONFRONTACE] s 13,9 procenty hlasů (18 mandátů) je zatím blok Spolehlivý dům dosavadního nejvyššího představitele republiky Arnolda Rüütela.> (PDIT-EDA 1.0, paralelní anotace)

První anotátor opět hodnotí spojení vět jako implicitní, po významové stránce pak vnímá výčet údajů o umístění ve volbách jako konjunktivní. Naproti tomu druhý anotátor označil sousloví *na druhém místě* za sekundární konektor s významem konfrontace.

Poslední, okrajovou příčinou neshody anotátorů na existenci implicitního diskurzního vztahu bývá anotátorská chyba – například jeden z anotátorů přehlédne diskurzní konektor, a označí vztah chybně za implicitní.

Shrňme tedy hlavní příčiny anotátorské neshody na samotné existenci implicitního diskurzního vztahu. Za prvé jak u mezivětných, tak u vnitrovětných vztahů může jít o neshodu na rozsahu diskurzních argumentů. Oba anotátoři vnímají, že na jistém místě v textu existuje implicitní diskurzní vztah, umísťují ho ale ve struktuře textu trochu jinak. Tento typ neshody může být doprovázen také neshodou na významovém popisu diskurzního vztahu, viz podkap. 4.4.5. Za druhé jde o konkurenci implicitních diskurzních vztahů a vztahů jiných. Ve vnitrovětné stavbě se to týká syntaktických vztahů, kterým může anebo nemusí být připsán také diskurzní význam, jde tedy o zhodnocení vztahu jako čistě syntaktického (tj. neanotovaného), nebo diskurzního. K tomu typicky dochází u přívlaskových vět vedlejších. V mezivětné stavbě pak stojí proti sobě diskurzní vztahy implicitní a explicitní. Neshoda tu vyplývá z toho, zda anotátoři jisté výrazy vyhodnotí jako konektory, nebo ne (a to konektory primární i sekundární). Pokud označí ve větě diskurzní konektor, neanotují potom sledovaný diskurzní vztah jako implicitní. K tomuto druhu neshody může docházet proto, že zejména třída sekundárních konektorů je otevřená a dovoluje různé interpretace.

4.4.5 Typické příčiny neshody na významovém popisu diskurzního vztahu

Nyní se zaměříme na ty výskyty, kde se anotátoři přesně shodli na samotné existenci implicitního diskurzního vztahu a na rozsahu jeho argumentů, avšak rozdílně je interpretovali z významového hlediska, viz tab. 4.13 a 4.14. Při analýze vztahů měli anotátoři možnost přiřadit nalezeným implicitním diskurzním vztahům vždy právě jedno z 23 významových označení. Protože v případě implicitních diskurzních vztahů nemají anotátoři lexikální vodítko, o něž by se mohli v interpretaci vztahu opřít, a zároveň je k dispozici mnoho tříd na výběr, lze zde v interpretaci významového vztahu očekávat rozpory.

Celkově se anotátoři shodli na samotné existenci implicitního diskurzního vztahu v 71 případech, z toho v 55 dokladech jde o vztahy mezivětné, v 16 o vnitrovětné. Vzhledem k tomu, že vnitrovětné vztahy jsou zde zastoupeny nízkým počtem a bylo

An1/ An2	konf	konj	kont	ekv	expl	gen	opo	nes	příč	rest	spec	Celk
konf	1	0	0	0	0	0	0	0	0	0	0	1
konj	0	12	7	0	0	0	0	0	0	0	1	20
kont	0	0	3	0	0	0	0	0	0	0	0	3
ekv	0	0	0	0	0	0	0	0	0	0	0	0
expl	0	0	0	0	3	0	0	0	3	0	1	7
gen	0	0	0	0	0	1	0	0	0	0	0	1
opo	0	0	0	0	0	0	2	0	0	1	0	3
nes	0	0	0	0	0	0	1	0	0	0	0	1
příč	0	0	0	0	1	0	1	0	5	0	0	7
rest	0	0	0	0	0	0	0	0	0	0	0	0
spec	0	3	0	1	0	0	0	0	0	0	8	12
Celk	1	15	10	1	4	1	4	0	8	1	10	55

Tabulka 4.13: Přehled anotátorské shody a neshody ve významovém zařazení mezivětných implicitních vztahů v paralelních anotacích PDiT-EDA 1.0

by nepřiměřeně vyvozovat závěry o shodách a neshodách při členění 16 výskytů do 23 možných významových skupin, budeme se zabývat především vztahy mezivětnými. U vnitrovětných pak rámcově hodnotíme, zda jejich typy anotátorské neshody odpovídají jevům u vztahů mezivětných.

Jak vyplývá z obou tabulek, neshodám v přiřazení významového typu vztahu podléhají zejména tři vztahy, konjunkce, specifikace a explikace. Ke společným rysům těchto vztahů patří skutečnost, že jde o vztahy vyjadřující navazování a že neobsahují rys protikladnosti či negace. Vztah konjunkce spojuje jednotky podobného druhu; ve vztahu specifikace je druhý argument upřesněním argumentu prvního; explikace pak představuje vysvětlení dění v prvním argumentu pomocí obsahu druhého argumentu.

Záměna vztahu konjunkce s jiným významovým vztahem

V našich datech je vztah konjunkce typicky zaměňován jednak se vztahem kontinuity, jednak s konjunktivní alternativou a konfrontací. V jistém smyslu se tyto vztahy pohybují na jedné škále: konjunkce vyjadřuje spojení, návaznost jednotek téhož typu. Kontinuita je významově volnější: vyjadřuje jistý, nepříliš přesný druh návaznosti. Naproti tomu konjunktivní alternativa a konfrontace jsou významově vymezenější, užší než konjunkce. Konjunktivní alternativa vyjadřuje možnosti, které mohou platit jak každá zvlášť, tak i obě naráz. Konfrontace pak vedle prostého vztahu konjunktivního obsahuje také rys protikladnosti, kontrastu (typicky se vyjadřuje spojovacím výrazem *kdežto* či *naproti tomu*). Při neshodách v mezích této skupiny tedy můžeme

An1/ An2	konfr	konj	konj. alt	ekviv	explik	příčina	spec	Celkem
konfr	1	3	0	0	0	0	0	4
konj	0	5	2	0	0	0	1	8
konj. alt	0	0	0	0	0	0	0	0
ekviv	0	0	0	0	1	0	0	1
explik	0	0	0	0	0	1	1	2
příčina	0	0	0	0	0	0	0	0
spec	0	1	0	0	0	0	0	1
Celkem	1	9	2	0	1	1	2	16

Tabulka 4.14: Přehled anotátorské shody a neshody ve významovém zařazení vnitrovětných implicitních vztahů v paralelních anotacích PDiT-EDA 1.0

říci, že anotátoři rozdílně vnímají těsnost návaznosti argumentů, případně další významové rysy vztahu.

V následujícím textu hodnotí anotátoři návaznost vět za sebou jako konjunkci (první anotátor) či kontinuuaci (druhý anotátor), a to dokonce dvakrát za sebou (příklady 79–80). U druhého z těchto výskytů uvádíme pro přehlednost znovu předchozí větu, aby bylo zřejmé, co je pro příklad 80 prvním argumentem. Nicméně jde o přímé pokračování téhož textu.

- (79) *Jornovo rané období ovlivňoval surrealismus. Neustálým hledáním tvarů a znaků, smyslem pro úsporné řazení čar a barevných tónů i proměnlivostí výrazu ovlivnil jeho tvorbu Paul Klee. Citem pro vyjádření přírodních dějů v jejich mnohoznačnosti a zároveň chápáním přesných zákonitostí vesmírného řádu se stal jeho předchůdcem Vasilij Kandinskij.*

<Arg1: Asger Jorn se zajímal také o historii, filozofii a literaturu, organizoval různá setkání, která vedla ke vzájemnému poznávání umělců.>

<Arg2:

Anotace 1 [vsunuto: **a** IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: **k tomu dále uvádíme** IMPLICITNÍ KONTINUACE]

Z rozmanitých podnětů se zvolna utvářel jeho nezaměnitelný názor a rukopis, určený především osobní intuicí, dojmy vycházejícími ze současnosti a ostrým viděním kulturního vývoje i společenských vztahů.> (PDiT-EDA 1.0, paralelní anotace)

- (80) <Arg1: *Z rozmanitých podnětů se zvolna utvářel jeho nezaměnitelný názor a rukopis, určený především osobní intuicí, dojmy vycházejícími ze současnosti a ostrým viděním kulturního vývoje i společenských vztahů.>*

<Arg2:

Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: *k tomu dále uvádíme* IMPLICITNÍ KONTINUACE]

Umělcovo dílo se od předválečné doby přes čtyřicátá léta postupně vyhraňovalo.>

Odkaz národní kultury se prolínal s rozsáhlým poznáním, zprostředkovaným i spoluprací s různými časopisy a autorovým podílem na zakládání rozličných uměleckých seskupení. (PDiT-EDA 1.0, paralelní anotace)

Na tomto textu se ukazuje zajímavý jev, řetězení neshody, resp. důslednosti v hodnocení u jednotlivých anotátorů. Pokud např. jeden anotátor vyhodnotí styl jednoho autora spíše jako „kontinuační“, vyznačuje kontinuu i na jiných místech v textu, kde druhý anotátor – taktéž opakovaně – vidí např. konjunkci. V uvedeném úryvku chápe první anotátor jednotlivá fakta o Asgeru Jornovi jako údaje podobného druhu, v konjunktivním vztahu, vytvářející celkový obrázek o představované osobnosti. Naproti tomu druhý anotátor vidí tyto údaje jako různorodé, spojené pouze osobností Asgera Jorna, ale velice volně. Proto označuje toto spojení jako kontinuu.

Stejně tak byl dvojím způsobem vyhodnocen vztah argumentů v následujícím úryvku o předpovědi počasí:

(81) <Arg1: *Nejnižší ranní teploty ve středu 10 až 6 st. C, ve čtvrtek 12 až 8 st. C, nejvyšší odpolední teploty 18 až 22 st. C.>*

<Arg2:

Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: *k tomu dále uvádíme* IMPLICITNÍ KONTINUACE]

OZÓN: Koncentrace ozónu v ozónové vrstvě nad naším územím je 3,6 % nad dlouhodobým průměrem.> (PDiT-EDA 1.0, paralelní anotace)

Při konjunktivním chápání tohoto spojení je zdůrazněn fakt, že obojí, teplota a stav ozónu, jsou parametry počasí a společně o něm postupně, strukturovaně vypovídají. Pokud vztah interpretujeme jako kontinuu, znamená to, že text považujeme za méně strukturovaný: podle tohoto chápání se v textu uvádějí různé informace, které se dají o počasí uvést, aniž by mezi těmito informacemi byl nějaký jasný vztah.

Konkurence konjunkce a kontinuity, tj. volnější chápání konjunkce, je v našich paralelních datech příznačná pro mezivětné implicitní vztahy. U vnitrovětných vztahů si konjunkce konkuruje spíše se vztahy újeji vymezenými, s konjunktivní alternativou a s konfrontací. V příkladech 82–83 je patrná záměna konjunkce a konjunktivní alternativy hned dvakrát, mezi první a druhou a potom druhou a třetí větou v souvětí. Pro přehlednost opět přetiskujeme souvětí ještě jednou, aby byla v obou případech patrná struktura argumentů diskurzních vztahů.

(82) *Trh s těmito právy existuje v té nejhorší pololegální a nelegální podobě.*

<Arg1: *Nájemníci byty vyměňují,>*

<Arg2:

Anotace 1 [vsunuto: **a** IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: **anebo** IMPLICITNÍ KONJUNKTIVNÍ ALTERNATIVA]

používají pro výdělečnou činnost,> nelegálně pronajímají. (PDiT-EDA 1.0, paralelní anotace)

- (83) *Nájemníci byty vyměňují,*

<Arg1: *používají pro výdělečnou činnost,>*

<Arg2:

Anotace 1 [vsunuto: **a** IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: **anebo** IMPLICITNÍ KONJUNKTIVNÍ ALTERNATIVA]

nelegálně pronajímají.> (PDiT-EDA 1.0, paralelní anotace)

U tohoto textu vidí jeden anotátor nelegální činnosti, které mohou nájemníci provozovat s pronajatým bytem, jako řadu prvků ve slučovacím vztahu. Je možné, že chápe aktéry těchto činností jako různé nájemníky – jedni byty vyměňují, druzí používají pro výdělečnou činnost atd.; tyto děje jsou pak ve vztahu konjunkce. V druhé anotaci jsou tyto činnosti vnímány více jako alternativy: nepředpokládá se, že by nájemník prováděl s bytem uvedené činnosti zároveň, ale výslovně vyloučené to není.

Poměrně často se lze u implicitních vztahů setkat s možným dvojím čtením také u vztahů konjunkce a konfrontace. Konjunkce spojuje dva prvky, které mají v jistém smyslu shodné rysy, patří do stejné množiny. Konfrontace spojuje stejné prvky, avšak zdůrazňuje jejich různost, protikladnost. Vnímání protikladnosti se může lišit, a to právě u implicitních vztahů, kde není vyjádřena konektorem, viz příklad 84:

- (84) *Pro trh s právy nakládání s byty spadajícími pod státní regulaci platí, že*

<Arg1: *poptávka je obrovská,>*

<Arg2:

Anotace 1 [vsunuto: **a** IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: **kdežto** IMPLICITNÍ KONFRONTACE]

nabídka nízká>

a právní regulace téměř nulová. (PDiT-EDA 1.0, paralelní anotace)

Anotátoři do takovýchto spojení doplňují v anotaci konektory, které podle nich nejlépe vyjadřují vztah mezi argumenty; tyto konektory se liší, bývá to spojka *a* a *kdežto/zatímco*. (Tj. *poptávka je obrovská a nabídka nízká* vs. *Poptávka je obrovská, kdežto nabídka nízká.*)

Typický je tento rozpor pro předpovědi počasí a vůbec pro texty, kde se řadě parametrů za sebou připisuje řada hodnot, srov. příklady 85–86:

- (85) <Arg1: Nejvyšší teplota 33,0 st. C byla v roce 1834,>
 <Arg2:
 Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]
 Anotace 2 [vsunuto: *kdežto* IMPLICITNÍ KONFRONTACE]
nejnižší 9,0 st. C v roce 1907.> (PDiT-EDA 1.0, paralelní anotace)
- (86) <Arg1: Ve středu vychází Slunce v 05:06 a zapadá ve 21:05,>
 <Arg2:
 Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]
 Anotace 2 [vsunuto: *kdežto* IMPLICITNÍ KONFRONTACE]
Měsíc vychází ve středu v 00:43 a zapadá v 16:27 hodin letního času.> (PDiT-EDA 1.0, paralelní anotace)

Opět tu záleží na tom, zda je řada údajů chápána jako prostý výčet, nebo jako řazení konfrontační.

V některých případech se konjunkce zaměřuje také se specifikací, a to jak u mezivětných vztahů, tak i u vnitrovětných. Znamená to, že anotátoři si odlišně vysvětlují význam druhého argumentu: jeden ho vidí jako další část textu, která je na stejné úrovni jako předchozí argument; druhý v něm spatřuje upřesnění informace z prvního argumentu.

- (87) <Arg1: Spěšný vlak do Děčína odjíždí z pražského Masarykova nádraží v 02:20 hod.>
 <Arg2:
 Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]
 Anotace 2 [vsunuto: *konkrétně* IMPLICITNÍ SPECIFIKACE]
Zastavuje v Praze-Bubenči, Roztokách u Prahy, Libčicích nad Vltavou, Kralupech nad Vltavou, Nelahozevsi, Hněvicích, Roudnici nad Labem, Bohušovicích nad Ohří, Lovosicích, Prackovicích nad Labem a hlavním nádraží v Ústí nad Labem.>
 (PDiT-EDA 1.0, paralelní anotace)

V příkladu 87 lze chápat informace o vlaku jako strukturu dvojího typu. U konjunkčního výkladu to znamená, že o vlaku se uvádějí dva údaje stejné důležitosti: děčínský vlak odjíždí v tolik a tolik hodin, staví v těch a těch stanicích. Při specifikačním výkladu je informace o zastávkách upřesněním, které se vztahuje k děčínskému vlaku.

Totéž kolísání mezi konjunkcí a specifikací vidíme také ve vnitrovětné stavbě:

- (88) *Je uvažovaný pětapadesátikorunový poplatek za denní pobyt na nemocničním lůžku přiměřený?*
 ...
Milena Veselovská, jednasemdesátiletá důchodkyně: Je to dost peněz, ale
 <Arg1: *doma stejně musím jíst,>*

<Arg2:

Anotace 1 [vsunuto: *a* IMPLICITNÍ KONJUNKCE]

Anotace 2 [vsunuto: *konkrétně* IMPLICITNÍ SPECIFIKACE]

denně s manželem tu stovku za jídlo utratíme.> (PDIT-EDA 1.0, paralelní anotace)

Tady jeden anotátor označuje vztah za konjunkci („musíme jíst a utratíme stovku“), zatímco druhý cítí, že už v prvním argumentu se nepřímo mluví o penězích a ve druhém se pak upřesňuje konkrétní částka („musíme jíst, a tedy platit za jídlo, a konkrétně utratíme stovku“).

Záměna vztahu specifikace s jiným významovým vztahem

Specifikace je druhým vztahem, u něhož dochází k charakteristické záměně. Kromě konkurence s konjunkcí, která byla popsána výše, se v datech objevuje také konkurence se vztahem explikace. U vztahu explikace obsahuje druhý argument vysvětlení pro obsah prvního argumentu; obsah propozic bývá podobný či přímo totožný, obě propozice jsou si blízké (Poláková et al. 2012b, s. 38–39), zároveň má vztah explikace i rys zdůvodnění, příčinnosti. U explicitních vztahů se typicky vyjadřuje např. slovem *totiž*, *vždyť*. Ve vztahu specifikace je v druhém argumentu vyjádřen detail či specifičtější informace k obsahu prvního argumentu (Poláková et al. 2012b, s. 52–53), nejde však o zdůvodňování. Jeho charakteristickými konektory při explicitním vyjádření jsou např. *přesněji řečeno*, *konkrétně*.

Příklad 89 ukazuje rozdílné hodnocení diskurzního vztahu mezi oběma argumenty:

- (89) *V utkání plném karet (celkem jsme napočítali 13 žlutých a jednu červenou) měli domácí s výjimkou okamžiků po první brance trvalou převahu.*

<Arg1: *Gólově ji však výrazněji vyčíslili až ve druhé půli.*>

<Arg2:

Anotace 1 [vsunuto: *konkrétně* IMPLICITNÍ SPECIFIKACE]

Anotace 2 [vsunuto: *totiž* IMPLICITNÍ EXPLIKACE]

kdy kromě tří branek ještě nastřelili tyč a břevno.> (PDIT-EDA 1.0, paralelní anotace)

První anotátor zde vidí uvedení čísel v druhém argumentu jako specifikaci, jak přesně vypadala převaha vítězů v druhé polovině utkání. Naproti tomu druhý anotátor chápe vztah mezi oběma argumenty jako explikační, doplňuje při anotaci pomocný konektor *tím, že*. V druhém argumentu pro něj není tolik podstatné, že se zpřesňuje informace z prvního argumentu, ale že se z jiného úhlu pohledu mluví o téže skutečnosti, o jednom faktu.

Častá je konkurence specifikace a explikace v takových dvojicích argumentů, kde v prvním z nich se nachází nějaký hodnotící výraz, viz příklad 90.

- (90) *Oproti předchozím je [album Síla v nás] méně explozivní. Dechy jsou sice stále výrazné, ale latinské rytmy i perkuse s výjimkou úvodní skladby Lidi se shromáždí ustoupily do pozadí. Větší roli naopak hraje kytara. Album je rockovější, což je patrné v písni To se mi snad zdá.*

<Arg1: Na desce je jiná Laura.>

<Arg2:

Anotace 1 [vsunuto: **konkrétně** IMPLICITNÍ SPECIFIKACE]

Anotace 2 [vsunuto: **totiž** IMPLICITNÍ EXPLIKACE]

Napětí ve swingujícím blues Já a Bůh vystupuje na povrch jen místy a ani rock'n'roll Zůstaňte se mnou není tak žhavý.>

Laura si teď více libuje v klidu. (PDiT-EDA 1.0, paralelní anotace)

Tento příklad dobře reprezentuje řadu dalších podobného typu. Specifikační interpretace zde vidí upřesnění, v čem přesně lze hudbu Laury z prvního argumentu hodnotit jako jinou – napětí se v blues objevuje jen místy, rokenrol není tak žhavý. U explikace je podstatný rys zdůvodnění: hudbu Laury můžeme hodnotit jako jinou, protože v blues je méně napětí a rokenrol není tak žhavý. U explikačního pojetí lze druhý argument vidět jako obhajobu volby slov v prvním argumentu (můžeme si dovolit nazvat něco nějak, protože...).

Záměna vztahu explikace s jiným významovým vztahem

Kromě záměny explikace se specifikací, již jsme se zabývali výše, se vztah explikace někdy zaměňuje se vztahem příčinným a s ekvivalencí. Na tuto skutečnost upozorňuje již příručka pro anotaci diskurzních vztahů (Poláková et al. 2012b, s. 38–39, 53), která se věnuje vztahům explicitním. Uvádí, že konkrétní význam vztahu při nejasnostech bývá určen použitým konektorem. U implicitních vztahů se o vodítko konektoru opřít nemůžeme, je tedy zřejmé, že kolísání, které se projevovalo i u vztahů explicitních, se zde bude objevovat tím spíše.

U konkurence vztahu explikace a příčiny jde o to, zda je druhý argument chápán jako vysvětlení téhož jevu z jiného hlediska, např. jako obhajoba volby jistého výrazu, nebo zda je vyjádřením jiného, samostatného obsahu, který je příčinou obsahu prvního argumentu. Srov. příklad 91:

- (91) *Práce s mužstvem ve špičce tabulky je velmi vyčerpávající.*

<Arg1: Ve Slušovicích nad námi neustále visel Damoklův meč,>

<Arg2:

Anotace 1 [vsunuto: **to znamená, že** IMPLICITNÍ EXPLIKACE]

Anotace 2 [vsunuto: **protože** IMPLICITNÍ PŘÍČINA-DŮSLEDEK]

postupovat jsme museli,>

podobně to bylo i v Drnovicích, tady je situace trochu odlišnější. (PDiT-EDA 1.0, paralelní anotace)

V tomto případě lze spojení vět rozumět dvěma způsoby. V prvním z nich je Damoklův meč přeneseným pojmenováním pro nutnost postupovat, do souvětí je možné doplnit konektor *totiž* (visel nad námi Damoklův meč, museli jsme totiž postupovat). Druhý argument je tedy explikací, vysvětlením faktu, který je obsažen v prvním argumentu; týž fakt se popisuje jinými slovy. Pokud pojmáme vztah mezi argumenty jako příčinný, je jeho interpretace jiná: *visel nad námi Damoklův meč* tu znamená „cítili jsme se pod tlakem“ (v tomto smyslu tato věta také dobře navazuje na předcházející kontext o vyčerpávající práci ve špičce tabulky). Druhá věta pak vyjadřuje jiný obsah, tj. „cítili jsme se pod tlakem, protože jsme museli postupovat.“

Explicace se může zaměřovat také s ekvivalencí. U obou těchto vztahů se popisuje jeden fakt dvěma způsoby, jinými slovy. V explikaci je však na rozdíl od ekvivalence obsažen navíc význam vysvětlení, zdůvodnění. Tento rys nemusejí oba anotátoři vidět v textu shodně, srov. příklad 92:

- (92) <Arg1: *První trh s byty funguje*,>
 <Arg2:
 Anotace 1 [vsunuto: *můžeme to tvrdit, protože* IMPLICITNÍ EXPLIKACE]
 Anotace 2 [vsunuto: *neboli* IMPLICITNÍ EKVIVALENCE]
cena bytu, byť závrtná dána je nabídkou a poptávkou.> (PDIT-EDA 1.0, paralelní anotace)

První anotátor se zde zaměřuje na zdůvodnění tvrzení, že trh s byty funguje (tedy odůvodňuje své hodnocení, volbu slov) – toto chápání textu je explikační. Druhý anotátor pak vidí oba argumenty jako vyjádření dějů s totožným obsahem: trh s byty funguje, neboli cena bytu je dána tržně.

Povaha záměn u vztahů mezivětných a vnitrovětných

Celkově se záměny typů diskurzních vztahů liší také podle toho, zda jde o vztahy vnitrovětné, nebo mezivětné. U vnitrovětných vztahů dochází ve vnímání významu k typické konkurenci u vztahu konjunkce a konfrontace / konjunktivní alternativy. Naproti tomu v mezivětných vztazích se konjunkce zaměřuje spíše s kontinuíací či se specifikací. K typickým mezivětným různocením patří také konkurence vztahu explikace a příčiny.

Je třeba však vzít v úvahu, že implicitní vnitrovětné vztahy jsou v našich paralelních anotacích zastoupeny poměrně málo. Mají také chudší rejstřík než vztahy mezivětné: vnitrovětné vztahy celkově obsahují 7 významových druhů, zatímco vztahy mezivětné jich pokrývají 11.

Zároveň je třeba také podotknout, že vedle vztahů, které se typicky zaměňují, existují i vztahy bez záměn, např. generalizace ve vztazích mezivětných.

	Diskurzní vztahy – implicitní (PDiT-EDA 1.0)	Diskurzní vztahy – explicitní (PDiT)	Textová koreference (PDiT)	Asociační anafora (PDiT)
F1 – přítomnost vztahu	0,54	0,43	0,72	0,46
Shoda na druhu vztahu	0,58	0,8	0,9	0,92
Cohenova κ u druhu vztahu	0,47	0,74	0,73	0,89

Tabulka 4.15: Mezipřímá shoda u různých typů textových vztahů

4.4.6 Mezipřímá shoda u implicitních diskurzních vztahů ve srovnání s jinými druhy textových vztahů

Srovnáme nyní míru mezipřímé shody v naší analýze implicitních vztahů s mezipřímou shodou v jiných projektech, které se zabývaly jinými typy koherenčních vztahů. Pokud jde o češtinu, máme k dispozici údaje o shodě u explicitních diskurzních vztahů, vztahů textové koreference²⁶ a u asociační anafory (vztahy typu *rodina – matka* apod.), viz tab. 4.15.²⁷

U srovnání v tabulce je třeba sledovat, na čem jsou míry shody založeny. Pro určení mezipřímé shody u implicitních vztahů platí jistá omezení, která je nutné mít na paměti. Tento výpočet se liší od nejčastěji používaného výpočtu mezipřímé shody u explicitních diskurzních vztahů. Jak ukazuje práce Mírovský et al. (2010), shodu u explicitních vztahů dobře vystihuje měření shody na konektorech (connective-based measuring). Při něm se přihlíží k tomu, zda oba anotátoři zaznamenali konektor. Nepřihlíží se pak k tomu, zda tomuto konektoru připsali strukturně naprosto shodný vztah, tj. takový, který by měl shodné výchozí a cílové uzly (viz naše měření 1), ani k významové klasifikaci vztahu (naše měření 2), srov. podkap. 4.4.3.

Protože implicitní vztahy nejsou formálně vyjádřeny konektorem, nemůžeme měření založené na konektorech využít. Na druhou stranu ale měření, která jsou v naší analýze představena u implicitních vztahů, je možné provést stejným způsobem i u vztahů explicitních (pro PDT 2.0 také byla provedena, srov. Mírovský et al. 2010), a srovnání těchto výsledků v tabulce 4.15 je tedy průkazné.

²⁶ Jde především o mezivětnou koreferenci v těch případech, kde koreferenční vztah není založen přímo na gramatických pravidlech. Příkladem textové koreference může být sekvence *chlapec – on – nevyjádřený podmět*; příkladem gramatické koreference je vazba mezi členy *chlapec – který*.

²⁷ Údaje o explicitních diskurzních vztazích byly přejaty z práce Mírovský et al. 2010, s. 777. Údaje o textové koreferenci a asociační anafoře byly publikovány ve článku Poláková et al. (2013, s. 96).

Ze srovnání vyplývá, že anotátoři implicitních diskurzních vztahů rozeznali samotnou existenci vztahu lépe než u vztahů explicitních, což je překvapivé. (Očekávali bychom, že explicitní konektor slouží jako jasné vodítko pro rozpoznání vztahu.) Snad tu mohla sehrát roli skutečnost, že na rozdíl od anotace explicitních vztahů, při níž šlo pouze o označování tohoto druhu vztahů, měli anotátoři u implicitních vztahů také označit všechny jiné druhy vztahů (otázka – odpověď, titulky apod.), aby bylo zjevné, že nějaký implicitní vztah neminuli či nepochopili jinak. Možná je právě tento postup navedl na pečlivější čtení textu. Výrazně vyšší je shoda na existenci vztahu u textové koreference, kde jde o poměrně snadný úkol – vyhledávání jednotlivých entit se shodnou referencí. Naopak anotace asociační anafory představuje náročný úkol v tom, že není a nemůže být jasně vymezena – kde jeden anotátor vidí nějaký vztah dvěma entitami (které však nejsou koreferenční), tam druhý takový vztah vnímat nemusí.

Shoda na druhu vztahu je vysoká u asociační anafory a textové koreference, kde se pracuje s relativně málo třídami. V rámci diskurzních vztahů tu lze pozorovat velký rozdíl mezi vztahy explicitními a implicitními: u explicitních se anotátoři shodují častěji (80 % případů proti 58 % u vztahů implicitních). Zde se bezesporu projevuje vliv explicitního diskurzního konektoru jako významového vodítka, o něž se mohou anotátoři v interpretaci vztahu opřít.

Srovnání hodnot Cohenovy kappy ukazuje, že anotace implicitních vztahů byla ze všech čtyř anotací nejnáročnější:²⁸ bylo třeba identifikovat vztah bez opory ve formálním ztvárnění (tento rys mají implicitní diskurzní vztahy společný s asociační anaforou) a dále zařadit do jedné z poměrně mnoha tříd (23 sémantických typů), a to opět bez opory např. ve významu vyjádřeného diskurzního konektoru.

Výsledky srovnání anotací odpovídají dřívějšímu pozorování: čím je anotace méně založena na formě, resp. čím se postupuje dále od nižších jazykových rovin (např. morfologie) k vyšším (např. koherenční vztahy), tím menší je mezianotátorská shoda. Nicméně tento rys je pro jazyk přirozený, vzrůstá zde jazyková vágnost a příliš vysoká shoda by tu odrážela spíše jakousi neadekvátnost v anotačních zásadách, totiž příliš striktní vyžadování jednoznačných řešení tam, kde nejsou z povahy věci možná (srov. Poláková 2014, s. 251–255).

²⁸ Toto číslo by teoreticky mohlo odrážet také např. nedbalost anotátorů či jejich neproškolenost; na základě zkušenosti s nimi ale víme, že tato příčina neshody bude spíše okrajová.



Analýza

V analytické části práce se zabýváme nejprve rozbořem dat korpusu PDiT-EDA 1.0 (kap. 5), v němž sledujeme distribuci explicitních a implicitních diskurzních vztahů (5.1) a způsoby vyjádření jednotlivých sémantických typů diskurzních vztahů (5.3–5.7). V kap. 6 potom představujeme psycholingvistické experimenty, v nichž jsme ověřovali, nakolik implicitnost znesnadňuje porozumění diskurzním vztahům.

5

Výsledky anotace v korpusu PDiT-EDA 1.0

5.1 Implicitní a explicitní diskurzní vztahy

Výsledky anotace implicitních vztahů v korpusu PDiT-EDA 1.0¹ jsou představeny v tabulkách 5.1–5.4, které odrážejí zachycení diskurzních jevů v korpusu podle technických parametrů, jež v něm byly nastaveny.

V tab. 5.1 je představeno zastoupení různých koherenčních vztahů,² mezi něž se řadí také vztahy diskurzní. Důležitou položkou této tabulky jsou právě vztahy diskurzní; vztahy mezi entitami a vztahy mezi otázkou a odpovědí byly anotovány pouze doplňkově, pro kontrolu, že nebyl vypuštěn žádný implicitní diskurzní vztah. Seznamy pak představují základ pro budoucí anotaci makrostruktury. Z tabulky vyplývá, že diskurzní vztahy se v korpusu PDiT-EDA 1.0 realizují převážně mezivětně (65 % vztahů mezivětných, 35 % vztahů vnitrovětných).

Tab. 5.2 uvádí vnitřní strukturu skupiny diskurzních vztahů. Jak je patrné, implicitní vztahy lehce převažují nad vztahy explicitními (53 % vztahů implicitních, 43 % vztahů explicitních s primárními konektory a 4 % explicitních vztahů se sekundárními konektory). Zastoupení implicitních a explicitních diskurzních vztahů se přitom liší v závislosti na syntaktické realizaci: pro vnitrovětná spojení jsou příznačné explicitní vztahy s primárním konektorem, zatímco v mezivětných spojeních se užívá především vztahů implicitních a okrajově explicitních vztahů se sekundárním konektorem.³

Distribuce implicitních a explicitních vztahů je znázorněna také v grafu na obrázku 5.1. Z grafu vyplývá, že vnitrovětné diskurzní vztahy bývají především explicitní (89 % dokladů), zatímco mezivětné vztahy jsou převážně implicitní (62 % dokladů). Test χ^2 pak potvrzuje, že mezi explicitností / implicitností diskurzního vztahu a jeho syntaktickou realizací existuje statisticky signifikantní spojitost ($X^2(1, N = 2715) = 756.0026, p < 0.00001$).

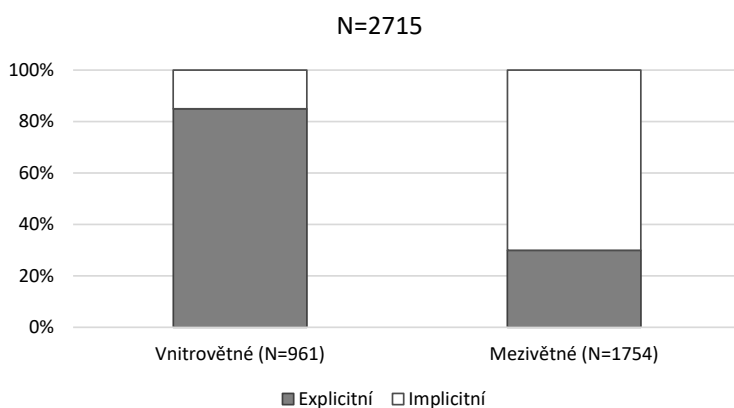
¹ Některé údaje o anotaci uvedené v této kapitole byly zveřejněny v článku Zikánová et al. 2019.

² Typy koherenčních vztahů a další pojmy jsou vymezeny v částech 4.4.1 a 4.4.2.

³ Je zajímavé, že explicitní vztahy se sekundárním konektorem se v tomto ohledu chovají stejně jako implicitní vztahy (jsou spíše mezivětné), a že tedy jistá dělicí čára nevede mezi explicitními vztahy jako celkem a vztahy implicitními, ale mezi vztahy s primárním konektorem, tj. jádrem explicitních vztahů a ostatními dvěma typy. Společný rys těchto dvou typů můžeme spatřovat v tom, že nejsou ustálené v jednoduché jasné explicitní formě, a možná proto je vysuzování těchto diskurzních vztahů obtížnější. Nabízí se pak interpretovat tento rozdíl tak, že pro jazyk je důležité jasně a jednoznačně vyjádřit koherenci uvnitř věty, zatímco mezivětná koherence může být volnější, a nemusí se proto vyjadřovat tak zřetelně.

	Vnitrovětné	Mezivětné	Celkem
Diskurzní vztahy	961	1 754	2 715
Vztahy mezi entitami	0	264	264
Vztahy mezi otázkou a odpovědí	2	63	65
Seznamy	0	92	92
Celkem	963	2 173	2 715

Tabulka 5.1: Zastoupení koherenčních vztahů v korpusu PDiT-EDA 1.0



Obrázek 5.1: Distribuce explicitních a implicitních vztahů ve vnitrovětné a mezivětné struktuře

V tab. 5.3 jsou pak zachyceny výsledky anotace diskurzní makrostruktury. Anotace mezer v koherenci představuje opět doplňkovou anotaci, jejímž cílem je výslovně označit další jevy, zaměnitelné s implicitními vztahy, a ověřit tak, že anotátor nepřehlédl implicitní diskurzní vztah. Anotace makrostruktury pak představuje základ pro budoucí výzkum. Je poněkud překvapivé, jak nízce jsou v korpusu zastoupeny mezery v koherenci; předpokládali jsme, že budou příznakovým rysem některých žánrů, např. sbírka zpráv, a že jistá útržkovitost je typická pro žurnalistický styl jako takový.

Poslední z tabulek (tab. 5.4) pak představuje výsledky doplňkové anotace zachycující specifické rysy přímo argumentů (tj. nikoliv vztahů), které ovlivňují jejich zapo-

	Vnitrovětné	Mezivětné	Celkem
Explicitní (primární konektory)	782	409	1 191
Explicitní (sekundární konektory)	16	81	97
Implicitní	163	1 264	1 427
Celkem	961	1 754	2 715

Tabulka 5.2: Zastoupení diskurzních vztahů v korpusu PDiT-EDA 1.0

	Vnitrovětné	Mezivětné	Celkem
Mezera v koherenci	0	25	25
Makrostruktury	0	126	126

Tabulka 5.3: Zastoupení vztahů diskurzni makrostruktury v korpusu PDiT-EDA 1.0

Titulek	189
Místo	30
Autor	28
Popisek	11
Metatext	1
Celkem	259

Tabulka 5.4: Zastoupení diskurzních rysů v korpusu PDiT-EDA 1.0

jení do textu. Poměrně vysoce jsou v datech našeho korpusu zastoupeny titulky, což odpovídá charakteru jeho textů.⁴

5.2 Implicitnost a sémantické typy diskurzních vztahů

Doposud jsme sledovali obecné charakteristiky anotace diskurzních vztahů, nyní se zaměříme na jejich významovou stránku (podkap. 5.2–5.6). Budeme se přitom zabývat následujícími otázkami:

- Které sémantické typy diskurzních vztahů patří ve sledovaných datech k frekventovaným a které k okrajovým?

⁴ Tento fakt má vliv na ztvárňování koherence v žurnalistických textech: titulky ve výrazné míře utvářejí kontext a vzbuzují očekávání, jaké informace se v takovém kontextu mohou přibližně vyskytovat. Vztah titulků a dalších kohezních prostředků může být námětem pro další výzkum.

- Liší se nějak distribuce sémantických typů ve vztazích explicitních a implicitních?
- Jak souvisí explicitnost a implicitnost diskurzních vztahů s jejich strukturní realizací (vztahy mezivětné a vnitrovětné)? Může se distribuce sémantických typů lišit také v závislosti na strukturní realizaci?
- Mají některé sémantické typy diskurzních vztahů tendenci k charakteristické realizaci, ať už explicitní / implicitní, či mezivětné / vnitrovětné?

Před vlastním rozbořem je třeba připomenout, že odpovědi, které získáme, se týkají právě anotovaných dat, nikoli např. češtiny jako celku. Je možné, že v jiných datech bychom získali jiné distribuce diskurzních vztahů. Zároveň musíme brát v úvahu, že shoda na anotaci těchto dat dosahuje jen určitého stupně, není – a ani nemůže být – stoprocentní (viz část 4.4.6). Přesto takto získáme výstupy, které je možno na dalších datech ověřovat.

Výsledky anotace sémantických typů diskurzních vztahů v korpusu PDiT-EDA 1.0 jsou shrnuty v tab. 5.5. Zastoupení jednotlivých sémantických typů v anotovaných datech se značně liší. Prvních pět nejfrekventovanějších sémantických typů tvoří 71,6 % výskytů všech vztahů. Nejčastějším sémantickým typem je konjunkce (33,4 % všech výskytů), následuje ji specifikace, příčina–důsledek, opozice a kontinua.

Z celkem dvaceti tří sémantických typů jich má třináct menší než 2% zastoupení. Konkrétně jde o gradaci, exemplifikaci, generalizaci, restriktivní opozici, účel, rektifikaci, ekvivalenci, konjunktivní alternativu, současnost, disjunktivní alternativu, nepravou příčinu, nepravou opozici a nepravou podmínku.

Všechny nepravé vztahy (nepravá příčina, nepravá opozice a nepravá podmínka) jsou v datech PDiT-EDA 1.0 zastoupeny naprosto okrajově (mají každý méně než 10 výskytů, tj. 0,2 % všech výskytů a méně).

5.3 Způsoby vyjádření různých sémantických typů diskurzních vztahů

Analýzovali jsme také spojitost sémantických typů diskurzních vztahů s explicitní či implicitní formou jejich vyjádření. Je totiž možné, že určitý sémantický typ (např. přípusťka) se vyjadřuje jistým konkrétním způsobem a jiný způsob se u něj neobjevuje. Ukázalo se však, že není přesné hodnotit sémantické vztahy pouze podle tohoto jednoho kritéria, neboť současně se značně liší distribuce sémantických typů v každé skupině podle toho, zda jde o spojení vnitrovětné, nebo mezivětné. Mezi oběma těmito parametry, tj. formou vyjádření na jedné straně a strukturní realizací na druhé straně, existuje prokazatelná spojitost (viz podkap. 5.1); vzniká pak otázka, v jakém jsou vztahu k sémantickým typům diskurzních vztahů.

Kombinací obou hledisek můžeme doklady v korpusu PDiT-EDA 1.0 rozdělit do čtyř skupin, na vztahy mezivětné implicitní (dále MI), mezivětné explicitní (ME), vnitrovětné implicitní (VI) a vnitrovětné explicitní (VE). Nejvýrazněji je v datech zastou-

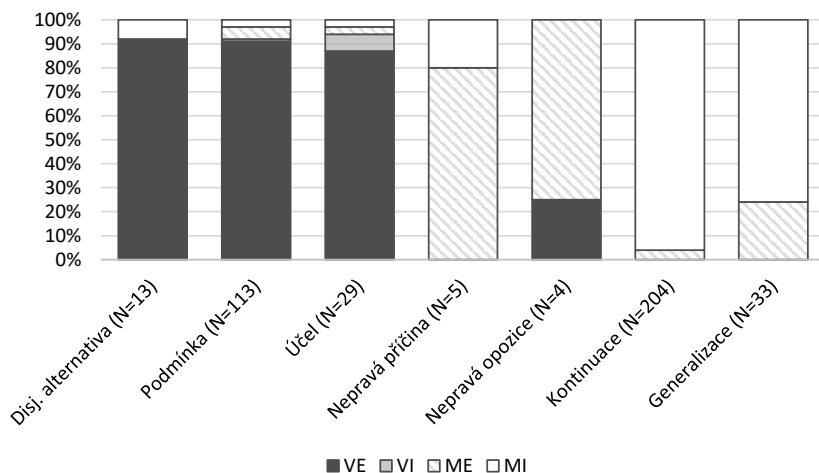
5.3 ZPŮSOBY VYJÁDRĚNÍ RŮZNÝCH SÉMANTICKÝCH TYPŮ DISKURZNÍCH VZTAHŮ

	Vnitrovětné Explicitní	Implicitní	Mezivětné Explicitní	Implicitní	CELKEM
Konjunkce	340	68	122	378	908
Specifikace	33	20	27	216	296
Příčina-důsledek	89	16	80	109	294
Opozice	75	10	103	57	245
Kontinuace	0	0	8	196	204
Explicke	15	14	14	90	133
Podmínka	103	1	6	3	113
Nesoučasnost	18	3	29	52	102
Konfrontace	12	20	19	46	97
Přípustka	33	1	11	10	55
Gradace	11	1	11	11	34
Exemplifikace	1	0	9	23	33
Generalizace	0	0	8	25	33
Restriktivní opozice	4	0	20	6	30
Účel	25	2	1	1	29
Rektifikace	9	1	6	9	25
Ekvivalence	2	2	2	17	23
Konj. alternativa	6	3	2	9	20
Současnost	8	1	4	4	17
Disj. alternativa	12	0	0	1	13
Nepravá přič.-důsl.	0	0	4	1	5
Nepravá opozice	1	0	3	0	4
Nepravá podmínka	1	0	1	0	2
CELKEM	798	163	490	1 264	2 715

Tabulka 5.5: Distribuce sémantických typů explicitních a implicitních diskurzních vztahů ve vnitrovětné a mezivětné strukturní realizaci (počty výskytů)

pena skupina vztahů MI (47 % ze všech výskytů) a dále VE (29 %). Další dvě skupiny jsou menší, vztahů ME je 18 % a podíl vztahů VI je nejmenší (6 %).

V následující analýze sledujeme, které sémantické typy se vyjadřují kterým z uvedených čtyř způsobů. Zajímá nás přitom, zda jistý sémantický typ má nějaký charakteristický způsob vyjádření, zda se typicky vyjadřuje jedním či více způsoby či zda je z tohoto hlediska nevymezený. Klademe si také otázku, které způsoby vyjádření jsou pro které sémantické typy výjimečné či nedoložené. Následně ještě sledujeme distribuci jednotlivých sémantických typů z obou hledisek zvlášť (explicitnost/implicitnost, vnitrovětnost/mezivětnost).



Obrázek 5.2: Sémantické typy s výrazně vymezeným způsobem vyjádření

V další fázi pak analyzujeme způsoby vyjádření sémantických typů nejen jednotlivě, ale také po skupinách. Zabýváme se otázkou, zda není některý ze způsobů vyjádření příznačný pro skupinu sémantických typů frekventovaných, zatímco ostatní by byly okrajové. Zjišťujeme také, zda mají sémantické typy s podobným významem (např. současnost a nesoučasnost jako vztahy časové) stejné způsoby vyjadřování.

Podle toho, nakolik silná je tendence sémantického typu k určitému způsobu vyjádření (ME, MI, VE, VI), dělíme sémantické typy do tří uskupení:

- sémantické typy s výrazně vymezeným způsobem vyjádření: typy, u nichž je jeden druh vyjádření zastoupen v 75 a více procentech výskytů
- sémantické typy s mírně vymezeným způsobem vyjádření: typy s jedním vyjádřením zastoupeným 50–74 % výskytů
- sémantické typy s nevymezeným způsobem vyjádření: ostatní typy

Sémantické typy s výrazně vymezeným způsobem vyjádření patří buď k jednomu, nebo k druhému hlavnímu způsobu vyjádření, VE a MI, viz graf na obr. 5.2. V datech korpusu PDIT-EDA 1.0 jsou doloženy také sémantické typy s charakteristickým vyjádřením ME, tedy z poměrně málo zastoupené skupiny. Nicméně jde o vztahy s velice nízkým počtem výskytů (nepravá příčina 4, nepravá opozice 3), jejich způsob vyjádření tedy nehodnotíme. V našich datech nejsou doloženy vztahy, pro něž by bylo výrazně typické vyjádření VI.

Ke vztahům, které se v naprosté převaze vyjadřují jako VE, patří disjunktivní alternativa (v 92 % dokladů VE; viz příklad 93), podmínka (91 %) a účel (86 %).

- (93) *Do té doby se aparát konfederace vystěhuje?*
„Nikdo z nás si nedovolí zbytečně zabírat nějaké prostory. Nejdříve se ovšem musí rozdělit majetek konfederace. A jestli v této budově tělovýchova zůstane i nadále, bude záležet na tom,
<Arg1: zda české svazy budou mít ve svých dosavadních kancelářích na Strahově dost prostoru,>
<Arg2: [či DISJUNKTIVNÍ ALTERNATIVA] vzhledem ke svým novým rozšířeným kompetencím budou mít zájem i o kanceláře tady Na Poříčí.“> (PDIT-EDA 1.0)

Ve skupině MI pak k výrazně vymezeným vztahům patří kontinua (96 %) a generalizace. Generalizace je v 76 % výskytů vyjadřována jako MI, viz příklad 94, ve zbylých 24 % pak jako ME; v našich datech je to tedy vztah výrazně mezivětný.

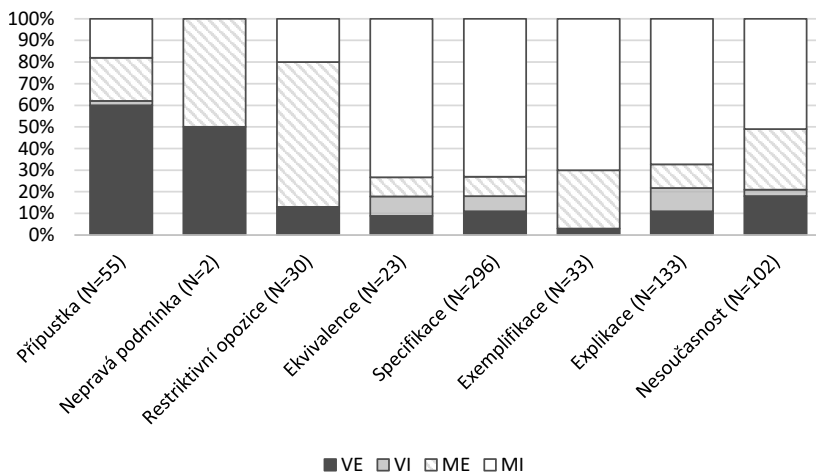
- (94) *Další důvody pro zahájení výroby byly:*
*<Arg1: * Zámková dlažba, která zaplavila trh před čtyřmi lety, je příliš jednotvárná a nedává architektovi dostatek možností pro estetické ztvárnění ploch.*
** Nový typ dlažby má velkou pevnost oproti jiným typům (vyrabí se tlakem 600 tun) i při malé tloušťce a rozměrech 40x40, resp. 40x60 cm.*
** Cena za m² může být srovnatelná s cenou za zámkovou dlažbu.*
** Nový typ dlažby je možné vhodně prostorově kombinovat se zámkovou dlažbou okolo stromů, na okrajích chodníků a jinde.>*
<Arg2: Vše [vsunuto: tedy IMPLICITNÍ GENERALIZACE] hovořilo pro.> (PDIT-EDA 1.0)

Sémantické typy s mírně vymezeným způsobem vyjádření se formálně realizují nejen jako dvě hlavní skupiny, VE a MI, ale také jako ME (viz obr. 5.3). Skupina VI není, stejně jako u typů s výrazně vymezeným způsobem vyjádření, u žádného sémantického typu doložena jako hlavní převažující.

K sémantickým typům s mírně vymezeným způsobem vyjádření VE se řadí přípustka (60 % výskytů), srov. příklad 95. (Nepravá podmínka je zastoupena pouze okrajově, její způsob vyjádření proto uvádíme pouze pro úplnost.)

- (95) *<Arg1: Ani Opava po propadáku předminulý týden na Xaverově tentokrát nezaváhala> (4:0 nad Turnovem),*
<Arg2: [ačkoliv PŘÍPUSTKA] měla před výkopem starosti se sestavením obrany – Orel zůstal po dlouhodobém zranění oka na lavičce a Hadašček s Onuferem byli v trestu za vyloučení v Horních Počernicích.> (PDIT-EDA 1.0)

Způsob MI pak mírně převažuje u sémantických typů ekvivalence (74 %), specifikace (73 %; viz příklad 96), exemplifikace (70 %), explikace (68 %) a nesoučasnosti. Vztah nesoučasnosti bývá přibližně v polovině případů vyjádřen jako MI (51 %), v další čtvrtině případů pak jako ME (28 %); jde tedy – ve zkoumaných datech – o vztah výrazně mezivětný.



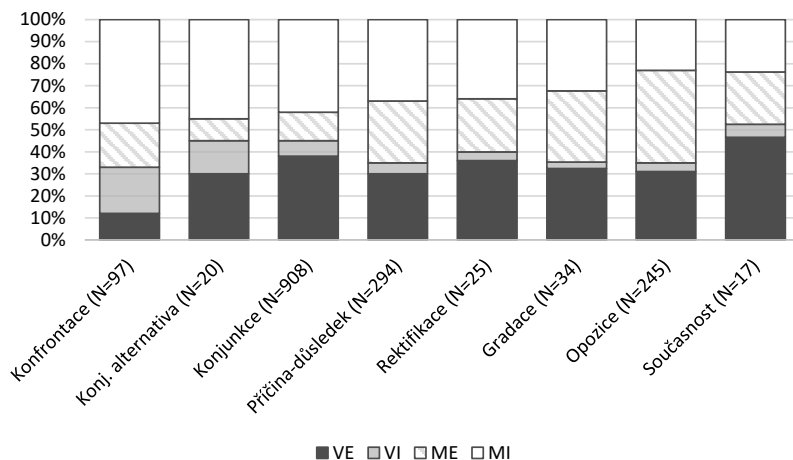
Obrázek 5.3: Sémantické typy s mírně vymezeným způsobem vyjádření

- (96) <Arg1: Minulý týden došlo na stránce Automobilismus k politováníhodné chybě.>
 <Arg2: [vsunuto: **konkrétně** IMPLICITNÍ SPECIFIKACE] U článku o Hondě Legend byla zaměněna fotografie s Roverem 214i.> Čtenářům i Hondě Česká republika, která nám vůz k testování zapůjčila, se omlouváme a auto ještě jednou připomínáme.>
 (PDIT-EDA 1.0)

Jistou zvláštností je vztah restriktivní opozice (srov. příklad 97), který se typicky vyjadřuje jako ME (67 %). Za zvláštnost ho považujeme proto, že skupina ME není celkově příliš frekventovaná, v rámci vztahů s výrazně vymezenou formou jsme ani nenašli sémantický typ, pro nějž by toto vyjádření bylo typické, odhlédneme-li od vztahů nepravých.

- (97) Česká spořitelna minulý týden zveřejnila svůj zájem získat šestnáctiprocentní podíl v České bance. Týden před tím převzala téměř beze zbytku AB Banku Mladá Boleslav. Investiční a poštovní banka, která vznikla v lednu sloučením dvou peněžních ústavů, zase před dvěma měsíci koupila více než třičtvrtinový podíl v Regiobance Hradec Králové. Zdá se tak, že období masového vzniku nových bank v České republice brzy vystřídají procesy jdoucí spíše opačným směrem. Restrukturalizace bankovního kapitálu totiž musí být logickým důsledkem dosavadního nerovnoměrného vývoje v českém bankovním sektoru.
 <Arg1: V ČR dnes operuje 57 bank.>
 <Arg2: Pouze čtyři z nich [**však** RESTRIKTIVNÍ OPOZICE] existovaly již před listopadem 1989.>
 Ostatní začínaly na absolutní nule. Bez depozit, bez pobočkové sítě, bez kvalifika-

5.3 ZPŮSOBY VYJÁDRĚNÍ RŮZNÝCH SÉMANTICKÝCH TYPŮ DISKURZNÍCH VZTAHŮ



Obrázek 5.4: Sémantické typy s nevymezeným způsobem vyjádření

něho personálu a také bez klientů, kteří by jim přinesli tolik potřebné primární zdroje. Aby nové úvěrové instituce vůbec mohly začít, byly nuceny nakupovat relativně drahé sekundární zdroje od velkobank a bojovat o přízeň vkladatelů o poznání vyššími úroky. Poměr sil mezi velkými a malými bankami se však přesto nezměnil. (PDIT-EDA 1.0)

Sémantické typy s nevymezeným způsobem vyjádření nemají tak vyhraněný způsob realizace jako předchozí vztahy, nicméně i u nich lze vypožorovat jisté tendence, nerealizují se nahodile (viz graf na obr. 5.4).

Tyto vztahy můžeme v zásadě rozdělit do tří podskupin. Sémantické typy jednovrcholové mají více než 25 % dokladů realizováno jedním způsobem. K nim patří konfrontace s hlavním vyjádřením MI (47 %) a současnost s hlavním vyjádřením VE (47 %).

Sémantické typy dvojevrcholové mají dva hlavní způsoby vyjádření, každým z nich je realizováno více než 25 % dokladů. Typickou dvojicí způsobů vyjádření je VE a MI, jde tedy o hlavní dva způsoby, které byly nejčastěji zastoupeny také v předchozích skupinách. Zatímco tam však stály v izolaci a mohly vést k dojmu, že vyjadřují nějaký protikladný význam, tady jsou společně využívány pro vyjádření téhož významu, byť jde o vyjádření ze dvou hledisek protikladné (explicitnost/implicitnost, strukturální realizace). Tato dvojice způsobů je typická pro vyjádření významu konjunktivní alternativy (45 % MI, 30 % VE), konjunkce (42 % MI, 38 % VE; viz příklady 98 a 99), příčiny-důsledku (37 % MI, 30 % VE) a rektifikace (36 % MI, 36 % VE).

- (98) *Občanský zákoník mu z titulu vlastnického práva umožňuje v § 711 odst. 1 vypovědět nájem bytu s přivolením soudu.*
Přitom musí jít o některé z těchto důvodů:
a) vlastník potřebuje byt pro sebe, manžela, své děti, vnuky, zete nebo snachu, své rodiče nebo sourozence,
b) jestliže <Arg1: nájemce přestal vykonávat práci, ke které je nájem služebního bytu vázán,>
<Arg2: [a KONJUNKCE] pronajimatel potřebuje byt pro jiného nájemce, který bude tuto práci vykonávat,>... (PDIT-EDA 1.0)
- (99) *<Arg1: Rozbití Varšavské smlouvy bylo jako odseknutí údů od těla.>*
<Arg2: [vsunuto: a IMPLICITNÍ KONJUNKCE] Od té doby se toho mnoho neudělalo.>
Ted' by se tyto (středoevropské) armády nedokázaly probíjet ani z mokrého papírového pytlíku.
Nejmenovaný západní vojenský odborník (PDIT-EDA 1.0)

Dvojvrcholová realizace je charakteristická také pro vztah opozice, a to jako VE a ME (31 % VE, 42 % ME; viz příklady 100 a 101). Opozice tak patří ke vztahům, pro něž je příznačná explicitnost, ať už je jejich strukturní realizace jakákoliv.

- (100) *Ve státech ES je navíc právem spotřebitele, aby měl na výrobku napsáno, z čeho je složen. To mu umožňuje vyhnout se konzumaci výrobků, které obsahují velké množství cizorodých látek. Tato práva občanů (a povinnosti výrobců) však zřejmě musí stanovit zákon.*
<Arg1: V našem případě by se mělo jednat o zákon o zdraví lidu,>
<Arg2: který se [však OPOZICE] nedá očekávat dříve než za dva roky.> (PDIT-EDA 1.0)
- (101) *<Arg1: Hodně se nyní v lázních hovoří o preventivních programech pro manažery.>*
<Arg2: Hezký nápad se [však OPOZICE] obtížně uskutečňuje.>
Málokterý vedoucí pracovník se může uvolnit na 21 dní, nebo alespoň na dva týdny. Proto jim vedení lázní Teplice vychází vstříc. Po dohodě s lázeňským lékařem si mohou rozložit dobu svého léčebného pobytu na několik částí. (PDIT-EDA 1.0)

V našich datech je doložen také trojvrcholový vztah s frekvencí vyšší než 25 % ve třech skupinách, VE, MI a ME. Jde o vztah gradace, srov. příklady 102–104. Složení této trojice skupin (32 % ME, 32 % MI, 32 % ME) odpovídá tendencím, které jsme pozorovali v předchozích případech: diskurzivní vztahy se obvykle vyjadřují jako VE či MI, případně ME, ale skupina VI je naprosto okrajová či nezastoupená, nedostává přednost před jiným způsobem vyjádření.

- (102) *Také já jsem se s Karlem Poláčkem za války setkal. Do Terezína jsem přišel koncem roku 1943. Bylo mi tehdy dvacet dva let. V tomto koncentračním táboře vládla zvláštní atmosféra, ve které se dařilo kultuře. My mladí jsme byli hladoví,*

<Arg1: nejen pokud jde o jídlo,>

<Arg2: [**ale** GRADACE] snad ještě více, pokud jde o kulturu.> (PDiT-EDA 1.0)

- (103) *Všechno se k lepšímu obrátilo s blížícími se volbami do kocourkovského magistrátu. <Arg1: V autobuse, pendlujícím mezi nádražím a kocourkovskou památkou, řidič každé tři minuty vykřikoval upozornění: Věnujte pozornost upozornění magistrátu!> Projíždíme zónou zvýšeného výskytu lupičů. Cestující jsou povinni dávat si na své prkenice pozor! Každého zadrženého lupiče nechť cestující předají radnímu pro bezpečnost!*

<Arg2: [vsunuto: **navíc** IMPLICITNÍ GRADACE] Pro zvýšení efektu řidič vždy zastavil a ládoval bambitku.>

To byl ovšem jen začátek. S blížícími se volbami přituhlo. Radní pro bezpečnost spolu s primátorem rozhodli, že provětrají noční metro. (PDiT-EDA 1.0)

- (104) <Arg1: Lidé, kteří uzavřou smlouvu se zahraniční pojišťovnou, která nemá povolení u nás působit, se vystavují nebezpečí, že pojišťovna nedodrží podmínky smlouvy a že v případě jejího vypršení či takzvané pojistné události, kterou může být například úraz, nedostanou žádnou pojistnou náhradu.>

<Arg2: [**Navíc** GRADACE] podmínky, které jsou často psány v cizím jazyce, takže potenciální klient má malou šanci jim do detailu porozumět, nemusí být mnohdy tak výhodné, jak se na první pohled zdá.> (PDiT-EDA 1.0)

Je třeba zdůraznit, že v přehledu typických způsobů vyjádření, jemuž jsme se doposud věnovali, jsou zachyceny hlavní tendence ztvárnění sémantických typů, to ale neznamená, že u nich nejsou možné i jiné způsoby vyjádření. Na druhou stranu některé sémantické typy se určitými způsoby v našich datech skutečně nevyjadřují. Jejich přehled uvádíme v tab. 5.6; zahrnujeme do něj sémantické typy s počtem výskytů 5 a méně ve sledované kategorii. Nepravé vztahy necháváme stranou pro jejich celkově nízké zastoupení v korpusu.

Klasifikace vztahů podle dvou parametrů zároveň, na něž jsme se nyní zaměřili (forma vyjádření a strukturní realizace) je sice přesná a může výborně vyhovovat zejména pro formálně jasně vymezené sémantické typy, může však být příliš detailní pro ostatní vztahy (mírně nevymezené, nevymezené). Pozorovali jsme již, že některé vztahy jsou např. typicky explicitní, ať už je jejich strukturní realizace jakákoliv. Proto se nyní zaměříme ještě na souhrnnou charakteristiku realizace jednotlivých sémantických typů z hlediska formy vyjádření a strukturní realizace zvlášť.

V následujícím přehledu zařazujeme do skupin s typickou formální realizací sémantické typy, které jsou sledovaným způsobem vyjádřeny v 70 a více procentech dokladů:

	Vztahy nedoložené	Vztahy níže zastoupené	Počet výskytů	Počet výskytů celkově
VE	Kontinuace	Restriktivní opozice	4	30
	Generalizace	Ekvivalence	2	23
		Exemplifikace	1	33
VI	Disjunktivní alternativa	Podmínka	1	113
		Účel	2	29
	Kontinuace	Přípustka	1	55
	Generalizace	Ekvivalence	2	23
	Restriktivní opozice	Nesoučasnost	3	102
		Konjunktivní alternativa	3	20
	Exemplifikace	Rektifikace	1	25
		Gradace	1	34
		Současnost	1	17
ME	Disjunktivní alternativa	Účel	1	29
		Ekvivalence	2	23
		Konjunktivní alternativa	2	20
		Současnost	4	17
MI	–	Disjunktivní alternativa	1	13
		Podmínka	3	113
		Účel	1	29
		Současnost	4	17

Tabulka 5.6: Řídké či nedoložené způsoby vyjádření jednotlivých sémantických typů v korpusu PDiT-EDA 1.0

- typicky explicitní: disjunktivní alternativa, účel, přípustka, podmínka, nepravá podmínka,⁵ restriktivní opozice, opozice, současnost
- typicky implicitní: kontinuace, generalizace, ekvivalence, specifikace, explikace, ekvivalence, exemplifikace, nepravá příčina-důsledek, nepravá opozice
- typicky mezivětné: restriktivní opozice, ekvivalence, specifikace, exemplifikace, explikace, nesoučasnost, nepravá příčina-důsledek, nepravá opozice
- typicky vnitrovětné: disjunktivní alternativa, podmínka, účel

⁵ Nepravé vztahy uvádíme v tomto přehledu pouze pro úplnost, mají velmi nízký počet výskytů, z něhož proto nelze vysuzovat žádné závěry (nepravá podmínka 2 výskyty, nepravá příčina 5, nepravá opozice 2 celkem).

Vztahy konjunktivní alternativy a konjunkce se typicky realizují ve dvou nejčastějších, avšak protikladných způsobech vyjádření, VE a MI, a nelze je proto zahrnout mezi typy uvedené výše.

Ostatní sémantické typy jsou z hlediska realizace nevyomezené. Jde o konfrontaci, gradaci, příčinu-důsledek a rektifikaci.

Stejně jako u kombinace obou hledisek, formy vyjádření a strukturní realizace, lze i u každého hlediska zvlášť sledovat, které možnosti se pro uvedené sémantické typy nevyužívají nebo se využívají okrajově. Do přehledu jsou opět zařazeny sémantické typy s počtem dokladů 5 a méně ve sledované kategorii, vypuštěny jsou vztahy nepravé:

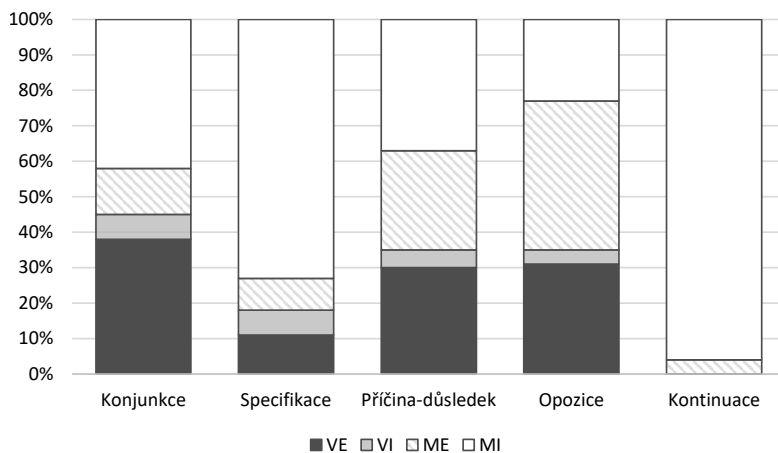
- explicitní: jsou zastoupeny všechny vztahy. Nízce je zastoupena ekvivalence (4 z celkových 23).
- implicitní: jsou zastoupeny všechny vztahy. Nízce je zastoupena disjunktivní alternativa (1 ze 13) a současnost (5 ze 17).
- mezivětné: jsou zastoupeny všechny vztahy. Nízce je zastoupena disjunktivní alternativa (1 ze 13) a účel (2 z 29).
- vnitrovětné: chybí generalizace a kontinua. Nízce je zastoupena exemplifikace (1 ze 33), restriktivní opozice (4 ze 30) a ekvivalence (4 z 23).

Dosud jsme při popisu distribuce sémantických typů podle způsobu vyjádření nepřihlíželi k frekvenci těchto typů ve zkoumaných datech. Vzniká otázka, zda se nejfrekventovanější sémantické typy neztvární nějakým charakteristickým způsobem, který by převažoval nad ostatními. Přehled způsobů vyjádření pěti nejčastějších sémantických typů, které tvoří dohromady 72 % všech vztahů, je shrnut v grafu na obr. 5.5.

Z grafu vyplývá, že nejfrekventovanější vztahy nemají jeden převažující způsob vyjádření: konjunkce a příčina-důsledek patří ke vztahům dvojvrcholovým (MI a VE; konjunkce 38 % VE, 42 % MI, příčina-důsledek 30 % VE, 37 % MI), specifikace a kontinua jsou typicky MI (specifikace 73 %, kontinua 96 %), opozice je vztahem dvojvrcholovým explicitním (31 % VE, 42 % ME). Pro nejčastější vztahy není charakteristický způsob VI; tento způsob ostatně není typický pro žádnou skupinu vztahů, které jsme dosud sledovali.

5.4 Souvislost mezi významem diskurzích vztahů a jejich vyjádřením

Prozatím jsme se zabývali sémantickými typy diskurzích vztahů a jejich způsoby vyjádření jednotlivě. Nyní se zaměříme jak u sémantické stránky, tak u stránky formální na celé skupiny a budeme sledovat, zda pro jistý způsob vyjádření jako celek jsou charakteristické významy, které mají nějaké společné rysy, a naopak, zda pro nějakou významovou skupinu diskurzích vztahů jako pro celek je příznačný jistý způsob vyjádření.



Obrázek 5.5: Způsoby vyjádření pěti nejčastějších sémantických typů diskurzních vztahů v PDIT-EDA 1.0

5.5 Distribuce z hlediska formy

Z výzkumu v této kapitole dosud vyplynulo, že z možných čtyř způsobů vyjádření (VE, VI, ME, MI) jsou nejběžnější dva, VE a MI, které jsou pro řadu vztahů typické jak izolovaně (typ VE a typ MI zvlášť), tak také ve dvojvrcholové distribuci (VE a MI společně jako typické způsoby vyjádření jednoho sémantického typu).

V izolované formě VE se typicky (výrazné a mírné vymezení, srov. výše) realizují vztahy podmínky, účelu, přípustky a disjunktivní alternativy. V případě prvních tří jde o vztahy s významem okolnosti. Zároveň ale ne všechny vztahy s významem okolnosti se přednostně ztvárňují právě tímto způsobem, srov. např. příčina-důsledek.

Izolovaný způsob vyjádření MI je pak charakteristický pro vztahy kontinuity, generalizace, ekvivalence, specifikace, exemplifikace, explikace a nesoučasnosti. (Jde opět o výrazné a mírné vymezení, viz výše). Kromě explikace a nesoučasnosti patří tyto vztahy do významové skupiny vztahů navazovacích.

Dvojvrcholová distribuce VE a MI je příznačná pro vztahy konjunktivní alternativy, konjunkce, příčiny-důsledku a rektifikace. Dva z nich, konjunktivní alternativa a konjunkce, patří do skupiny vztahů navazovacích, celkově je však sémantika vztahů s touto distribucí způsobů vyjádření různorodá.

5.6 Distribuce z hlediska významu

V korpusu PDIT-EDA 1.0 rozlišujeme čtyři hlavní významové skupiny, a to vztahy časové, kauzální, kontrastivní a navazovací (viz tab. 4.7).

Časové vztahy se ve způsobech realizace liší. Zatímco současnost se vyjadřuje především jako VE (47 % výskytů), nesoučasnost je typicky mezivětná (51 % MI, 28 % ME, 18 % VE, 3 % VI).

Také skupina vztahů s významem kontrastu je po formální stránce nejednotná. Vysoce frekventovaná opozice patří ke dvěma explicitním způsobům vyjádření (ME 42 %, VE 31 %); její varianta, restriktivní opozice, je pak doložena především v jednom z těchto typů, ME (67 %). Naproti tomu pro přípustku je charakteristické ztvárnění typu VE (60 %). Pro rektifikaci je příznačná dvojvrcholová distribuce VE a MI (po 36 %). Zbývající dva vztahy, konfrontace a gradace, jsou pak nevymezené, vyjadřují se častěji třemi způsoby (VI, MI a ME), přičemž jejich distribuce je u gradace rovnoměrná (po 32 %), zatímco u konfrontace převažuje typ MI (47 %). Obecně tedy lze říci, že vztahy s významem kontrastu jsou po formální stránce extrémně nejednotné; lze však u nich pozorovat – s výjimkou vztahu konfrontace a rektifikace – tendenci k explicitnosti.

Také ve významové skupině kauzálních vztahů dochází k různým realizacím jednotlivých sémantických typů. Nejčastější vztah, příčina–důsledek, je po formální stránce nevymezený, se třemi hlavními skupinami vyjádření (MI 37 %, VE 30 %, ME 27 %). Naproti tomu významově poměrně blízká explikace se vyjadřuje především jako MI (68 %). Za poznámku také stojí odlišná distribuce pravé (viz výše) a nepravé příčiny–důsledku, byť tento nepravý vztah je v našich datech doložen celkově jen pěti doklady, a proto je třeba k jejich rozdělení přistupovat se značnou rezervou. Vztah nepravé příčiny–důsledku je doložen především jako ME (80 %), jednou také jako MI (20 %), vnitrovětné zastoupení nemá. Poslední dva vztahy z této skupiny, podmínka a účel, patří k typu VE (podmínka 91 % výskytů, účel 86 %). Z hlediska způsobu vyjádření je tedy významová skupina kauzálních vztahů nejednotná; s výjimkou vztahu explikace u ní můžeme pozorovat tendenci k explicitnosti.

Na rozdíl od předchozích tří skupin se významové vztahy ve skupině navazovacích vztahů vyjadřují poměrně jednotně, a to převážně jako MI (exemplifikace 70 %, specifikace 73 %, ekvivalence 74 %, generalizace 76 %). Další dva vztahy, konjunkce a konjunktivní alternativa, mají dvojvrcholovou distribuci MI a VE, přičemž většinu u nich tvoří také doklady typu MI (konjunkce MI 42 %, VE 37 %; konjunktivní alternativa MI 45 %, VE 30 %). Výjimkou mezi sémantickými typy ve skupině navazovacích vztahů je disjunktivní alternativa, která se vyjadřuje téměř výhradně jako VE (92 %). Celkově jsou tedy tyto vztahy převážně typu MI, v případě konjunkce a konjunktivní alternativy MI a VE. Konjunktivní alternativa má ve způsobu vyjádření blíže ke konjunkci než k disjunktivní alternativě. Disjunktivní alternativa se v této skupině ostře odlišuje.

Souhrnně lze říci, že mezi sémantickými skupinami diskurzních vztahů a způsobem jejich vyjádření není jasná souvislost, s výjimkou vztahů navazovacích, které mají tendenci k realizaci typu MI. Ve skupině vztahů kontrastivních a kauzálních je jistá tendence k explicitnímu vyjádření vztahů; celkově je však typické ztvárnění i blízkých významových typů naprosto různorodé.

Právě různorodost typické formy blízkých sémantických vztahů je překvapivá. Srovnajme např. odlišná ztvárnění konjunktivní a disjunktivní alternativy, u časových vztahů současnosti a nesoučasnosti, dále např. příčiny a explikace. Tyto výsledky nás přivedly na myšlenku ověřit typické způsoby vyjádření některých vztahů také

v psycholingvistickém experimentu a zjistit tak, nakolik jsou zjištění o distribuci sémantických typů v korpusu PDiT-EDA 1.0 platná obecně (viz kap. 6).

5.7 Závěr

Při analýze sémantických typů diskurzních vztahů v korpusu PDiT-EDA 1.0 jsme zkoumali jednak distribuci samotných sémantických typů obecně, jednak jejich vztah k formálnímu ztvárnění. K pěti nejfrekventovanějším sémantickým typům, které tvoří 71,6 % všech diskurzních vztahů v korpusu PDiT-EDA 1.0, patří konjunkce (33,4 %) a dále specifikace, příčina-důsledek, opozice a kontinuační.

Distribuce vztahů v textu je výrazně ovlivňována dvěma parametry: formou jejich vyjádření (vztahy explicitní a implicitní) a strukturní realizací (vztahy mezivětné a vnitrovětné). Na základě těchto dvou hledisek jsme vyčlenili čtyři způsoby vyjádření diskurzního vztahu: VE, ME, VI, MI. Nejčastěji se v našich datech vyskytují vztahy typu MI (47 %) a VE (29 %); naopak vztahy typu VI jsou zastoupeny jen okrajově. Jednotlivé sémantické typy diskurzních vztahů se liší ve vyhraněnosti způsobů vyjádření. Některé sémantické typy mají výrazně vymezený způsob vyjádření (např. generalizace je jasným typem MI), jiné jsou vymezené pouze mírně či vůbec. K nevymezeným typům patří např. vztah konfrontace. U sémantických typů s nevymezenou formou lze rozlišit vztahy jednovrcholové, dvojevrcholové i třívrcholové. Ve všech případech je nejméně zastoupenou skupinou VI. Charakteristickou vícevrcholovou distribucí je kombinace dvou nejčastějších způsobů vyjádření, totiž VE a MI u jednoho sémantického vztahu (např. konjunkce).

Charakterizovali jsme sémantické typy také podle každého z uvedených hledisek zvlášť (forma vyjádření, strukturní realizace) a popsali, pro které typy je nějaké ztvárnění příznačné – např. vztah opozice bývá vyjádřen explicitně. Zároveň jsme vymezili, kterými způsoby se které vztahy naopak nevyjadřují (např. účel, podmínka a přípustka nebývají realizovány způsobem VI).

Z analýzy dokladů pěti nejčastějších sémantických typů (konjunkce, specifikace, příčina-důsledek, opozice a kontinuační) vyplývá, že nemají jeden či dva převažující způsoby vyjádření, naopak, i když každý z nich má nějaké typické ztvárnění, tato ztvárnění se mezi sebou výrazně liší.

Zkoumali jsme také souvislost mezi významem a formou diskurzních vztahů z pohledu významových a formálních skupin. Jak se ukazuje, vztahy navazovací se převážně vyjadřují způsobem MI a naopak, typ MI slouží především pro vyjádření navazovacích vztahů. Dále lze pozorovat, že kontrastivní a kauzální vztahy bývají především explicitní. Nicméně celkově je vazba mezi významem a formou skupin diskurzních vztahů velice volná, vztahy významově blízké se vyjadřují formálně různě a naopak, vztahy formálně stejné či podobné mají nejrozumnější významy.

6

Implicitní vztahy v psycholingvistických experimentech

Výsledky analýzy sémantických typů diskurzních vztahů a jejich explicitní či implicitní formy jsou v několika ohledech překvapivé. Některé sémantické typy vztahů se vyjadřují typicky implicitně, např. kontinuační či generalizační. Jakou roli potom má konektor v jejich explicitních variantách? Napomáhá porozumění, nebo je nadbytečný, či dokonce porozumění brání? Jiné sémantické typy jsou naopak typicky explicitní (disjunktivní alternativa, podmínka). Znamená to, že v implicitní variantě jsou hůře srozumitelné? Dalším překvapivým zjištěním je rozdílná forma vyjádření u blízkých sémantických typů. Proč k němu dochází? Nemůže jít o zvláštnost, která platí pro náš datový soubor, nikoli ale obecněji?

Abychom mohli zodpovědět tyto otázky, provedli jsme několik psycholingvistických experimentů.¹ Během nich jsme respondentům předkládali krátké texty s konektory a bez konektorů a srovnávali jsme, u kterých je porozumění snazší.

Pro analýzu jsme vybrali čtyři sémantické typy, dva kontrastivní (konfrontace, přípustka) a dva časové (současnost, nesoučasnost). Abychom získali srovnatelná data, omezili jsme se na vztahy mezivětné.

V korpusu PDiT-EDA 1.0 se u kontrastivních vztahů ukázalo, že sémantický typ konfrontace je spíše implicitní, zatímco přípustka je v mezivětných vztazích zastoupena vyváženě mezi implicitní a explicitní formou (u vnitrovětných vztahů je výrazně explicitní). Ověřujeme tedy, zda implicitní forma přípustky brání jejímu porozumění více než u konfrontace.

Podobnou distribuci pozorujeme také u mezivětných vztahů časových: nesoučasnost je tu převážně implicitní, zatímco současnost se vyjadřuje rovnoměrně explicitně i implicitně. Také zde verifikujeme tvrzení, že implicitnost u současnosti ztěžuje porozumění více než u nesoučasnosti.

6.1 Metodologie experimentů

Pokusy proběhly metodou čtení vlastním tempem (self-paced reading, Just et al. 1982, Jegerski 2014). Při tomto postupu se čtenáři postupně odkrývají a zase mizí části textu na obrazovce; tempo odkrývání si čtenář určuje sám stisknutím tlačítka. U jednotlivých

¹ Tyto experimenty proběhly ve spolupráci s laboratorii Labels (FF UK a Psychologický ústav AV ČR), kde se na nich podílel doc. Filip Smolík. Jemu také děkujeme za statistické zpracování experimentu.

vých položek se měří čtenářův reakční čas, tj. v jakých intervalech mačká tlačítko. Předpokládá se, že u položek, které jsou obtížnější na porozumění, je reakční čas delší. Toto zdržení se přitom může projevit ne přímo u dané položky (komplikovaného místa v textu), ale také s jistým zpožděním o něco dále v textu, kdy čtenář vidí, že mu přicházející text nedává smysl hned na první pohled a potřebuje čas, aby tento smysl nějak interpretoval. Čtenář se u této metody nemůže vrátit k položkám, které se již zakryly.

Otázky, které si klademe výše, jsme shrnuli do následující výchozí hypotézy: přítomnost explicitního konektoru usnadňuje porozumění textu. Pro respondenty tedy budou snazší na porozumění vztahy explicitní než implicitní. Respondenti proto budou mít u explicitních diskurzních vztahů kratší reakční dobu v úseku na konci druhého uvedeného diskurzního argumentu. U implicitních vztahů budou potřebovat čas na pochopení souvislosti mezi dvěma diskurzními argumenty, a budou proto v závěru sekvence diskurzních argumentů reagovat pomaleji. Nehodnotíme celkovou dobu čtení věty, protože při dávkovaném čtení jde o proces, který neodpovídá přirozenému plynulému čtení. (Navíc je doklad s implicitním vztahem vždy o jedno slovo kratší, totiž o diskurzní konektor.)

Jednotlivé doklady, které jsme respondentům předkládali, mají ustálenou stavbu čtyř vět (srov. vztah konfrontace v příkladu 105). První věta vytváří kontext, v druhé je obsažen první diskurzní argument. Třetí věta obsahuje druhý diskurzní argument, na jehož počátku u explicitních diskurzních vztahů stojí konektor (u implicitních vztahů je vynechán). Čtvrtá věta je opět pouze doplňková, kontextová. Jejím cílem je zabránit tomu, aby se sledované jevy vyskytly v poslední větě, protože čtení posledních úseků mívá specifické rysy, které by mohly výzkum zkreslit.

(105) *V interdisciplinárním programu byl za tu dobu podán pouze jediný, biologicko-historický projekt.*

U historiků vloni jednohlasně prošel.

Zato v biologii potom naprosto propadl.

Finanční prostředky tedy nakonec vyplaceny nebyly.

(Experimentální sada)

Pro každý vztah jsme se zaměřili na funkci právě jednoho konektoru: u konfrontace to bylo *zato*, u přípustky *přesto*, u současnosti *mezitím* a u nesoučasnosti *potom*. Vyloučili jsme tak z výzkumu možnost, že porozumění textu bude ovlivněno přídatnými významy synonymních konektorů, které v tuto chvíli nejsme schopni odhadnout (např. konfrontace *zato* – *naproti tomu*).

Struktura vět v diskurzních argumentech (tj. druhá a třetí věta dokladu) je v rámci každého sémantického typu totožná a zároveň co nejjednodušší. Do dokladů proto nejsou zahrnuty negativní a reflexivní tvary sloves, slovesné tvary složené a reflexiva obecně. Naší snahou bylo také vyjadřovat příslovečná určení času a místa shodně, buď příslovcem, nebo předložkovou frází. Výjimečně se oba způsoby v jednom souboru

dat mísí; v těchto případech jsme dali přednost přirozenému vyznění textu, jehož není takto opakovaně jednoduché dosáhnout. Doklady jednotlivých sémantických typů měly tedy následující větnou stavbu:²

Konfrontace

Věta 2 (Arg1)

- (nevyjádřený podmět v sg.)
- 1. kontrastivní člen (předmět nebo příslovečné určení času či místa; jedno slovo, max. jednoduchá předložková fráze)
- příslovečné určení času či místa
- jiné příslovečné určení (příslavce)
- přísudkové sloveso

Věta 3 (Arg2)

- (konektor)
- (nevyjádřený podmět v sg.)
- 1. kontrastivní člen (předmět nebo příslovečné určení času či místa; jedno slovo, max. jednoduchá předložková fráze)
- příslovečné určení času či místa
- jiné příslovečné určení (příslavce)
- přísudkové sloveso

- (106) *Strýce Martina jsem měl obzvlášť rád.
V soukromí vždy jednal odvážně.
Zato na veřejnosti často působil nesměle.
Naučil mě lyžovat a také se nebát výšek.*

(Experimentální sada)

- (107) *Zajímavý je vývoj počtu žáků.
Na venkově každoročně nepatrně klesá.
V Praze trvale mírně stoupá.
Zřejmě je vyšší nejen o žáky, kteří se stěhují do Prahy, ale také o ty, kteří dojíždějí.*

(Experimentální sada)

Přípustka

Věta 2 (Arg2)

- (nevyjádřený podmět, označení osoby, v sg.)
- příslovečné určení času či místa
- přechodné přísudkové sloveso (minulý čas)
- přívlastek shodný závislý na předmětu (přídavné jméno)
- předmět (neživý)

² U každého sémantického typu uvádíme za popisem větné stavby příklad dokladu s explicitním a implicitním vztahem.

Věta 3 (Arg1)

- (konektor)
- (nevyjádřený podmět, koreferenční s podmětem předchozí věty)
- příslovecné určení času či místa
- přechodné přísudkové sloveso (minulý čas)
- přívlastek shodný závislý na předmětu (přídavné jméno)
- předmět (neživý)

- (108) *Rodiče se rozhodli koupit Milanovi něco na sport.
Dlouho chtěl nové kolo.
Přesto nakonec dostal školní atlas.
Maminku atlas ve výloze upoutal cestou do cyklistického obchodu.*
(Experimentální sada)

- (109) *Petr byl v náročné situaci.
Pořád měl obrovské dluhy.
Dál kupoval drahé zbytečnosti.
Blížil se neodvratný krach.*
(Experimentální sada)

Současnost

Věta 2 (Arg1)

- podmět 1
- přechodné přísudkové sloveso (minulý čas)
- předmět
- příslovecné určení místa (předložková vazba)

Věta 3 (Arg2)

- (konektor)
- příslovecné určení místa (předložková vazba)
- nepřechodné přísudkové sloveso (minulý čas)
- podmět 2 (nekoreferenční s podmětem 1, součást fokusu)

- (110) *Dvorek na chalupě najednou ožil.
Zedník míchal maltu v kolečku.
Mezitím z rozhlasu vyhrávala hudba.
Babička začala velet partě dělníků.*
(Experimentální sada)

- (111) *Bylo tu další prázdninové poledne.
Tereška skákala panáka na chodníku.
V bufetu probíhala hádka.
Opakovalo se to tak každý den.*

(Experimentální sada)

Nesoučasnost

Věta 2 (Arg2)

- podmět 1
- přechodné přísudkové sloveso (minulý čas)
- předmět
- předložková vazba (především příslovečné určení místa)

Věta 3 (Arg1)

- (konektor)
- předložková vazba (příslovečné určení místa)
- nepřechodné přísudkové sloveso (minulý čas)
- podmět 2 (nekoreferenční s podmětem 1, součást fokusu)

- (112) *Po obědě nastal konečně klid.
Velitel prolistoval noviny na pohovce.
Potom na dveře zaklepal nováček.
Podal jen standardní hlášení a byl hned pryč.*

(Experimentální sada)

- (113) *Celé město bylo v napětí.
Meyer zaslechl dunění od náměstí.
Do ulice vjel tank.
Zdálo se, že neznají přesně cestu.*

(Experimentální sada)

Pro každý sémantický typ jsme respondentům v experimentu předkládali položky s explicitním vyjádřením diskurzního vztahu a vztahy implicitní (nikoli tytéž položky, aby se respondentům nezrychloval reakční čas proto, že obsah dané položky už znají). Respondenti byli rozděleni vždy do dvou skupin: doklady, které dostala jedna skupina jako explicitní, byly druhé skupině předloženy jako implicitní a naopak. U každého dokladu proto bylo možné porovnat, jak rychle byl čten jako explicitní a jako implicitní, ale od různých respondentů. Pro každý sémantický typ tak vzniklo několik sad dokladů, viz tab. 6.1.

U experimentů zaměřených na časové vztahy dostali jednotliví respondenti testovací sady s tímto složením: 10 dokladů explicitní současnosti, 10 (jiných) dokladů implicitní současnosti, 10 dokladů explicitní nesoučasnosti, 10 (jiných) dokladů im-

	Respondenti 1	Respondenti 2
Sada A (10 dokladů)	explicitní	implicitní
Sada B (10 dokladů)	implicitní	explicitní

Tabulka 6.1: Rozdělení sad dokladů pro každý sémantický typ mezi respondenty

plicitní nesoučasnosti, 40 dokladů výplňových. Pořadí dokladů bylo vždy náhodné a pro každého respondenta jiné.

U experimentů kontrastivních se pak testovací sada skládala vždy z následujících dokladů: 10 dokladů explicitní přípustky, 10 (jiných) dokladů implicitní přípustky, 10 dokladů explicitní konfrontace, 10 (jiných) dokladů implicitní konfrontace, 40 dokladů výplňových.

Respondenti nevěděli, na co jsou experimenty zaměřeny, nevěděli také, že věty 2 a 3 jsou pro výzkum důležitější než věty 1 a 4. Podstatné totiž bylo, abychom těmito informacemi neovlivnili jejich čtení textu. Ze stejného důvodu byly položky s diskurzními vztahy v experimentech doplněny ještě výplňovými položkami, v nichž se sledované jevy nevyskytovaly. Struktura výplňových položek byla podobná (čtyři věty), větná stavba druhé a třetí věty v nich však nebyla tak striktně omezena jako v diskurzních dokladech a na konektorech ani sémantickém typu diskurzních vztahů v nich nezáleželo (srov. příklad 114).

- (114) *Ivan měl rád jednoduché technické zlepšováky.
V zimě si omotal pneumatiky na kole prádelní šňůrou.
Kola mu tak na sněhu neprokluzovala.
Pyšnil se, že má nejlevnější sněhové řetězy v republice.*
(Experimentální sada)

Bylo nutné také udržet pozornost čtenářů u textů a zajistit, že texty pouze mechanicky neodklikají. Proto následovala za každou položkou jednoduchá kontrolní otázka k obsahu položky, na kterou respondenti odpovídali *ano* / *ne* (otázka k příkladu 114: *Vyplývá z textu, že Ivan jezdil v zimě na kole?*).

Celkově čtenáři neprováděli žádný vědomý jazykový rozbor, nevyslovovali se o tom, jak čemu rozumějí, neanotovali texty pomocí termínů ani jejich náhražek. Jejich úkolem bylo pouze číst texty v tempu, při němž byli schopni textům rozumět, a odpovídat na otázky k textům. Komentáře mohli připojit ve vzkazu výzkumníkům na konci experimentu.

Každého experimentu se zúčastnilo 50–60 (konfrontace 50 a přípustka 50 osob, současnost 60 a nesoučasnost 60 osob). Šlo o dospělé rodilé mluvčí češtiny.

Experimenty byly provedeny v programu *ibex* (<https://spellout.net/ibexfarm/>, cit. 2. 2. 2021).

Implicitní vztahy				Explicitní vztahy		
Pozice ve větě	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	Počet pozorování	Průměrný čas (ms)	Směr. odchylka
1	–	–	–	498	556	193
2	493	615	249	496	516	198
3	497	520	214	491	533	204
4	500	529	203	495	531	209
5	463	841	412	462	770	365

Tabulka 6.2: Konfrontace: výsledky experimentu

6.2 Výsledky experimentů

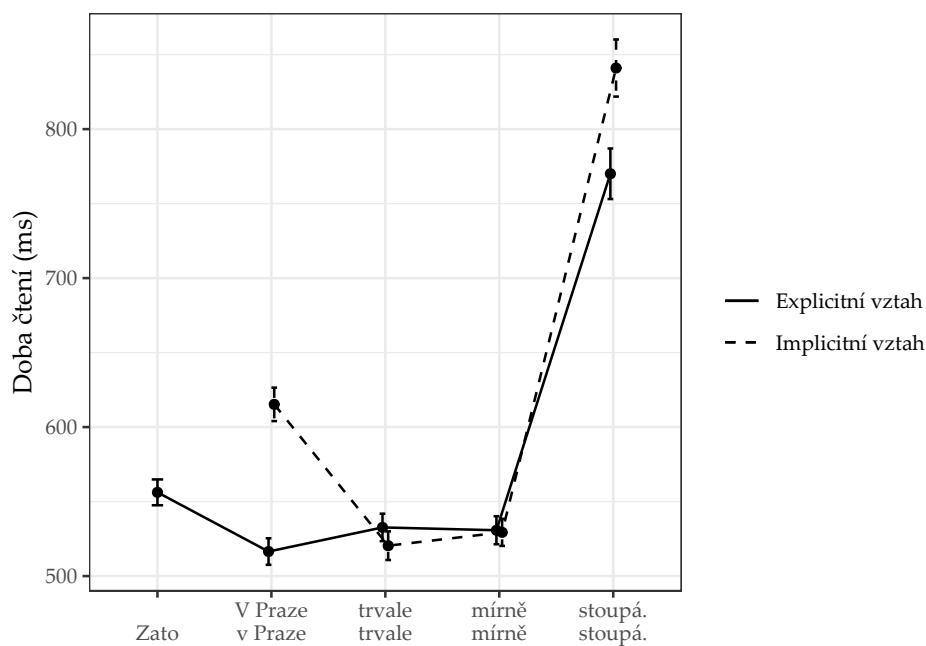
Výsledky byly analyzovány pomocí smíšených regresních modelů (Baayen 2008). Tyto modely vyhodnocovaly sílu závislosti mezi reakční dobou čtení a přítomností / nepřítomností konektoru. Pro každý experiment byly porovnávány časy čtení pro každé slovo ve větě; věty buď byly, nebo nebyly uvozeny spojovacím výrazem, tento fakt sloužil jako fixní prediktor. Zajímala nás zejména rychlost čtení posledního úseku třetí věty (tj. druhé věty v sekci argumentů diskurzivního vztahu). Zohledněny byly také rysy jednotlivých respondentů a položek jako zkřížené náhodné faktory (tj. bylo zohledněno, že každý čtenář čte jinak rychle a že každá položka sama o sobě je jinak náročná na porozumění). Pro vyloučení chybných a podezřelých pozorování byly vyřazeny časy čtení pod 100 ms a nad 2000 ms. Modely byly odhadovány pomocí balíků lme4 and lmerTest v systému R (Kuznetsova et al. 2017, Bates et al. 2015).

6.2.1 Konfrontace

Základní výsledky experimentu jsou shrnuty v tabulce 6.2, jíž odpovídá graf na obrázku 6.1.³ Další tabulka (6.3) pak ukazuje vyhodnocení těchto výsledků pomocí smíšeného regresního modelu. Její výstupy se mohou lišit od předchozí tabulky a grafu, protože jsou očištěné od individuálních efektů (jak efekty slov v jednotlivých dokladech – některá z nich mohou být čtena nápadně jinak od jiných ve stejné pozici, tak efekty jednotlivých respondentů – někdo čte rychleji, někdo pomaleji). Stejným způsobem jsou představeny výsledky i u dalších vztahů.

³ V tomto grafu, stejně jako v následujících, pro lepší přístupnost popisujeme jednotlivé pozice ve větě pomocí výrazů z příkladové věty. Stejně tak bychom mohli napsat *konektor – příslovečné určení* atd. Graf vypovídá o všech dokladech sledovaného vztahu, s různým lexikálním obsazením.

Svislé úsečky v jednotlivých bodech ukazují velikost směrodatné odchylky.



Obrázek 6.1: Konfrontace: doba čtení úseků věty u explicitního a implicitního vztahu

U konfrontačního vztahu se projevil statisticky významný rozdíl ve druhém diskurzním argumentu na dvou místech, jednak u prvního slova po konektoru (*v Praze*), jednak u posledního slova (*stoupá*, srov. tab. 6.3).⁴

Zpomalení u prvního slova ve větách implicitních (druhá pozice) je společné všem čtyřem sledovaným vztahům (kolem 100 ms). Dochází zde k efektu prvního slova, na němž se čtenáři obvykle poněkud zdrží a který je obecný, nesouvisí nijak s diskurzními vztahy. Je vidět, že ve větách implicitních i explicitních potom dochází ke zrychlení u druhého slova, jen u vět implicitních je tento efekt posunutý až na příslovecné určení času, protože právě ono je u nich druhým slovem.

Zároveň je tu patrný rozdíl v rychlosti zpracování prvních slov – v explicitních větách konektoru, v implicitních prvního plnovýznamového členu. Zpracování plnovýznamového členu vyžaduje více času než zpracování konektoru. Vysvětlujeme to tím, že konektor pochází z uzavřené třídy, která se obvykle na počátku věty dá očekávat, zatímco plnovýznamové slovo je nepredikovatelné. Tento rozdíl, tj. v čase čtení prvních úseků v explicitních a implicitních větách, statisticky nehodnotíme –

⁴ Za hranici signifikance považujeme hodnotu $p = 0,01$.

Pozice ve větě	Implicitní vztah čas čtení (ms)	Explicitní vztah čas čtení (ms)	Rozdíl (ms)	p
2	615,57	518,35	-97,22	<0,001
3	520,25	534,18	13,93	0,196
4	529,33	526,86	0,196	0,823
5	865,58	791,22	-74,36	<0,001

Tabulka 6.3: Konfrontace: doba čtení úseků věty u explicitního a implicitního vztahu (výstup smíšeného regresního modelu; p označuje statistickou významnost rozdílu)

vzhledem k různé povaze materiálu (konektor vs. plnovýznamové slovo) by to nemělo smysl.

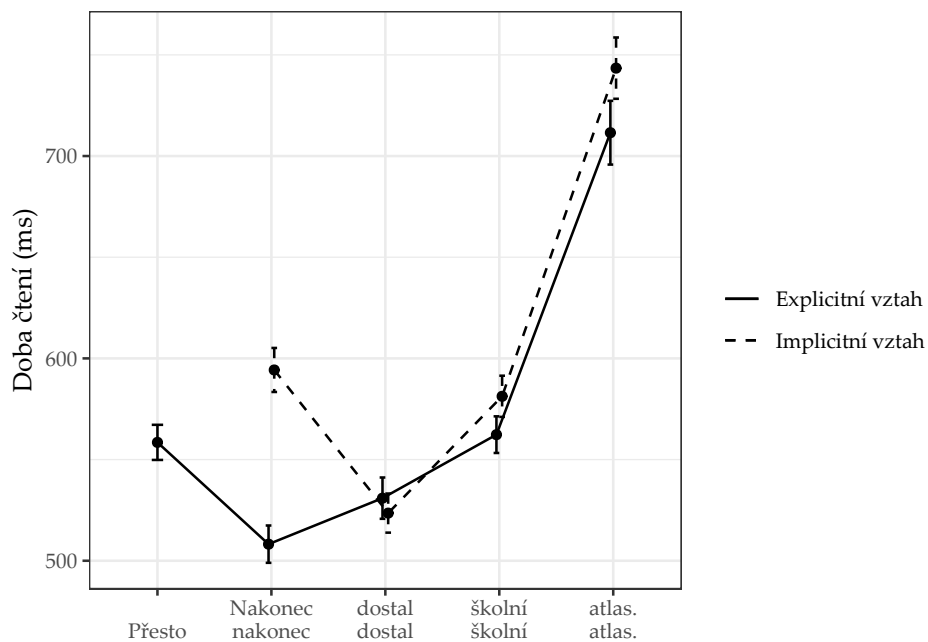
Naopak druhý statisticky signifikantní rozdíl je důležitý. Projevuje se ve čtení posledního úseku věty: ukazuje se tu, že u explicitních vztahů zpracováváme význam konfrontačního vztahu podstatně rychleji.

Implicitní vztahy				Explicitní vztahy			
Pozice ve větě	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	
1	–	–	–	496	558	193	
2	494	594	242	500	508	206	
3	496	524	215	499	531	228	
4	498	581	227	498	562	202	
5	472	743	329	476	712	344	

Tabulka 6.4: Přípustka: výsledky experimentu

6.2.2 Přípustka

Podobný je průběh explicitního a implicitního vztahu u přípustky (srov. tab. 6.4, 6.5 a obr. 6.2). Také zde se liší doba zpracování počátku věty, což je dáno efektem prvního slova a rozdílu mezi konektorem a plnovýznamovým slovem na počátku, jak jsme již uvedli. Tento rozdíl není podstatný. Důležitý je opět rozdíl v době čtení posledního úseku (*atlas*), kde u explicitní věty je čtení rychlejší. Tento rozdíl nedosahuje signifikance $p = 0,01$, ukazuje však zjevnou tendenci ($p < 0,05$). Lze tak dovozovat, že přítomnost konektoru v přípustkovém vztahu usnadňuje porozumění, byť slaběji než u vztahu konfrontace.



Obrázek 6.2: Přípustka: doba čtení úseků věty u explicitního a implicitního vztahu

Pozice ve větě	Implicitní vztah čas čtení (ms)	Explicitní vztah čas čtení (ms)	Rozdíl (ms)	p
2	594,82	508,2	-86,62	<0,001
3	523,9	530,87	6,97	0,555
4	582,03	563,38	-18,65	0,082
5	752,92	721,82	-31,19	0,049

Tabulka 6.5: Přípustka: doba čtení úseků věty u explicitního a implicitního vztahu (výstup smíšeného regresního modelu)

6.2.3 Současnost

U vztahu současnosti nebyly zaznamenány žádné statisticky významné rozdíly kromě druhé pozice, tj. pozice porovnávající první výraz v implicitním vztahu s výrazem bezprostředně po spojovacím výrazu v explicitním vztahu (*v bufetu*). Jak už

Implicitní vztahy				Explicitní vztahy		
Pozice ve větě	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	Počet pozorování	Průměrný čas (ms)	Směr. odchylka
1	–	–	–	585	587	271
2	579	607	289	597	539	253
3	595	548	272	585	551	242
4	548	690	385	544	693	384

Tabulka 6.6: Současnost: výsledky experimentu

Pozice ve větě	Implicitní vztah čas čtení (ms)	Explicitní vztah čas čtení (ms)	Rozdíl (ms)	p
2	612,65	539,97	-72,68	<0,001
3	550,69	554,51	3,82	0,712
4	721,86	720,68	-1,18	0,939

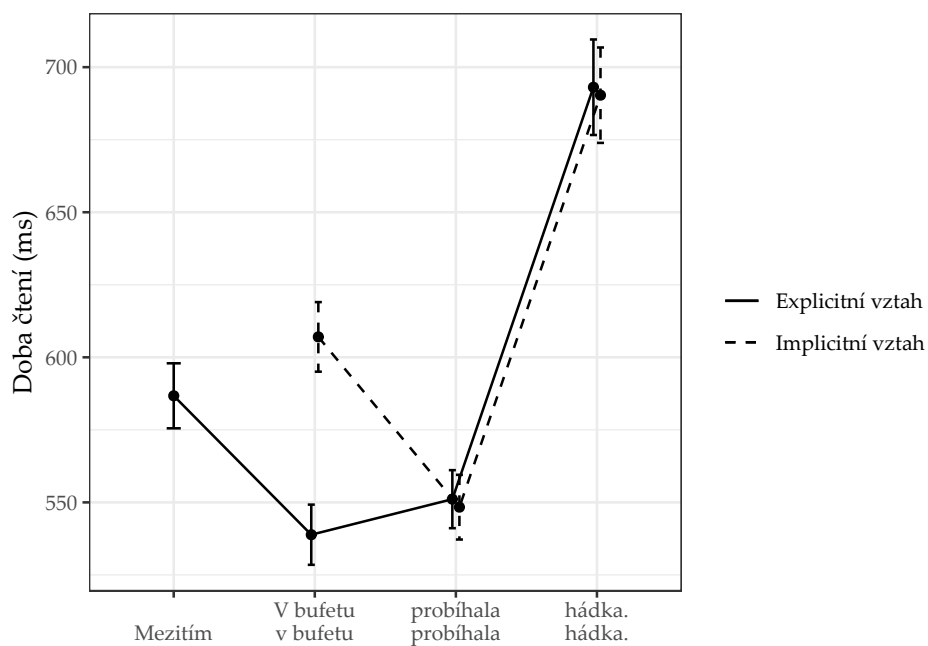
Tabulka 6.7: Současnost: doba čtení úseků věty u explicitního a implicitního vztahu (výstup smíšeného regresního modelu)

bylo uvedeno výše, jde o efekt obecného zpomalení na počátku věty, nesouvisí tedy s explicitností / implicitností diskurzivního vztahu. Rychlost čtení posledního úseku ve větě se u explicitních a implicitních vztahů neliší, viz tab. 6.6, 6.7 a obr. 6.3.

6.2.4 Nesoučasnost

U vztahu nesoučasnosti (viz tab. 6.8, 6.9 a obr. 6.4) byl zaznamenán rychlejší čas čtení v implicitních vztazích u třetí pozice, tj. u druhého slova implicitních vztahů (*zaklepal*). Pravděpodobně se jedná o přesahující efekt z předchozí fáze, kdy pomalejší zpracování prvního slova (efekt počátku věty) poté vede k rychlejšímu zpracování slova dalšího. Shodou okolností jde o druhé (a tedy rychleji čtené) slovo v implicitních vztazích a třetí (tedy již nezrychlené) slovo ve vztazích explicitních. Rozdíl v časech čtení zde proto nepovažujeme za podstatný. Tento jev, byť bez statistické významnosti, lze pozorovat i u všech ostatních sledovaných vztahů.

Naopak podstatná je skutečnost, že u vztahu nesoučasnosti nebyl zaznamenán statisticky významný rozdíl v rychlosti čtení posledního slova (*nováček*).



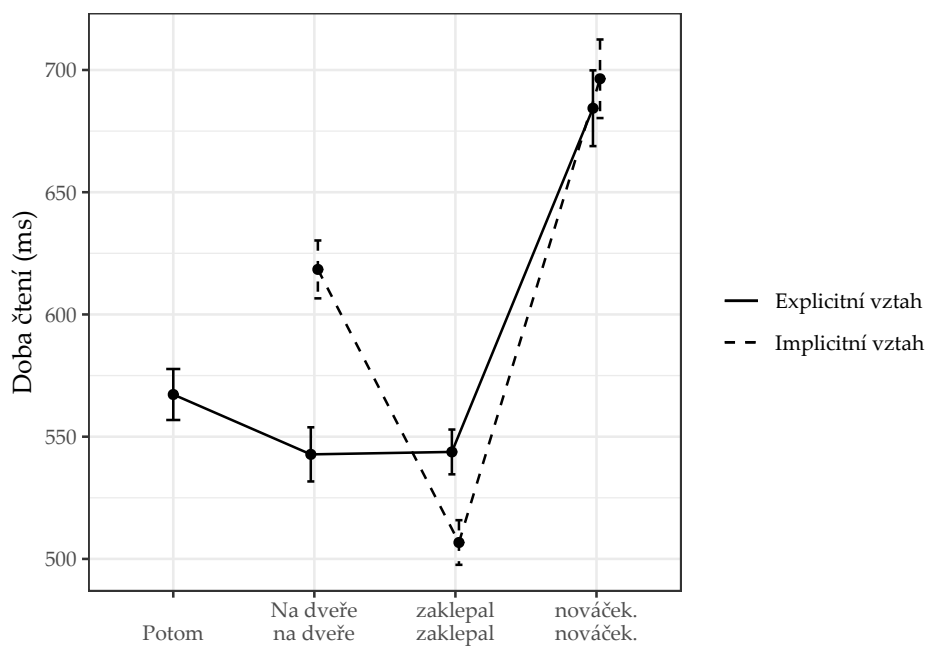
Obrázek 6.3: Současnost: doba čtení úseků věty u explicitního a implicitního vztahu

Implicitní vztahy				Explicitní vztahy			
Pozice ve větě	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	Počet pozorování	Průměrný čas (ms)	Směr. odchylka	
1	–	–	–	588	567	253	
2	578	618	285	593	543	270	
3	596	507	223	588	544	222	
4	559	696	380	548	684	363	

Tabulka 6.8: Nesoučasnost: výsledky experimentu

6.3 Závěr a diskuse

Výsledky experimentů jsou rozdílné pro vztahy kontrastivní a pro vztahy časové. U kontrastivních vztahů přítomnost konektoru významně urychluje čtení posledního segmentu ve třetí větě dokladu ve vztahu konfrontace. U přípustky se tendence



Obrázek 6.4: Nesoučasnost: doba čtení úseků věty u explicitního a implicitního vztahu

Pozice ve větě	Implicitní vztah čas čtení (ms)	Explicitní vztah čas čtení (ms)	Rozdíl (ms)	p
2	620,6	544,19	-76,41	<0,001
3	506,79	544,51	37,72	<0,001
4	718,87	721,19	2,32	0,874

Tabulka 6.9: Nesoučasnost: doba čtení úseků věty u explicitního a implicitního vztahu (výstup smíšeného regresního modelu)

k rychlejšímu čtení explicitního vztahu projevuje také, nicméně toto zrychlení nedosahuje míry statistické významnosti.

Na rychlost čtení časových vztahů nemá přítomnost konektoru vliv, u posledního úseku třetí věty je u nich rychlost čtení shodná.

Z pozorování těchto čtyř vztahů tedy vyplývá, že přítomnost konektoru v žádném případě porozumění nekomplikuje. U časových vztahů nezáleží na explicitnosti či im-

	Anotace PDiT-EDA 1.0		Psycholingvistické experimenty
	Implicitní	Explicitní	
Konfrontace	46	19	Explicitní forma je významně srozumitelnější (silná tendence, $p < 0.001$)
Připustka	10	11	
Současnost	4	4	Explicitní forma je významně srozumitelnější (slabší tendence, $p = 0.049$)
Nesoučasnost	52	29	Na formě nezáleží

Tabulka 6.10: Srovnání vybraných mezivětných vztahů v anotaci PDiT-EDA 1.0 a v psycholingvistických experimentech (počty výskytů)

plicitnosti vztahu, u vztahů kontrastivních jsou implicitní vztahy hůře srozumitelné, což platí zejména pro konfrontaci, slaběji pro připustku.

Porovnejme nyní výsledky experimentů s výsledky anotace v korpusu PDiT-EDA 1.0. Při srovnání je třeba vzít v potaz, že anotace uvedených vztahů v korpusu zahrnuje všechny významové nuance vztahů konfrontace, připustky, současnosti a nesoučasnosti, vyjádřené u explicitních vztahů různými konektory, zatímco experimenty se zaměřovaly na přítomnost a nepřítomnost pouze jednoho konektoru pro každý sledovaný vztah, tj. data z experimentu nevypovídají o sémantice určitého vztahu v celé šíři.

Mezivětný vztah konfrontace je v korpusu PDiT-EDA 1.0 realizován především implicitně, srov. tab. 6.10. (To platí i pro jeho vnitrovětnou variantu.) Z experimentu však vyplývá, že právě implicitní forma konfrontace je hůře srozumitelná. Tento rozpor mezi anotacemi a experimentem by mohl mít několik příčin. Za prvé, jak už jsme uvedli, v anotaci je konfrontace pojata širěji, odpovídá více diskurzním konektorům, které mohou mít různé významové zabarvení, zatímco v experimentu se zkoumají pouze vztahy vyjádřené konektorem *zato*. Za druhé může nesoulad mezi výsledky anotace a experimentu ukazovat na to, že data v anotaci jsou specifická či příliš malá a nevystihují rysy češtiny obecně. Zároveň je třeba znovu uvážit, že anotátoři dosáhli poměrně nízké shody.

Za třetí, anotátoři při rozboru textů v korpusu měli k dispozici širší kontext než respondenti v experimentu, kteří se mohli orientovat pouze na základě jedné věty před vlastním vztahem a jedné věty po něm. Je možné, že pro pochopení vztahu konfrontace je potřeba širší kontext obecně. To by mohlo např. znamenat, že jde o vztah, který se nevyskytuje na počátku textů nebo v jejich prvním odstavci, nýbrž se užívá až ve chvíli, kdy se čtenáři či posluchači dostatečně orientují v situaci, která je v textu představována.

Tuto myšlenku jsme ještě ověřili na datech PDiT-EDA 1.0. Zaměřili jsme se na texty, které obsahují alespoň 18 vět, a zkoumali, zda se v nich vztah konfrontace vyskytuje na počátku textu. Rozdělili jsme prvních 18 vět textu (včetně nadpisu) na třetiny (věta 1–6, 7–12, 13–18) a sledovali, ve které z těchto částí je umístěn pravý argument vztahu konfrontace. Přitom jsme nerozlišovali, zda jde o vztah mezivětný, či vnitrovětný a explicitní, či implicitní. Vztah konfrontace se v první skupině vyskytl 4krát, ve druhé 19krát a ve třetí 22krát. Při srovnání relativních zastoupení u různých sémantických typů (v procentních bodech)⁵ je konfrontace v prvních šesti větách zastoupena ze všech sémantických typů nejméně, pouze v 9 % případů. Pro srovnání, druhý kontrastivní vztah, přípustka, se v této části textu vyskytuje v 17 % případů. Data z PDiT-EDA 1.0 ukazují, že konfrontace se na počátku textů nevyskytuje často; její nízký výskyt ale není relevantní při kontrole χ^2 testem. Domníváme se, že by mohlo záležet také na tom, jaké výrazy se dostávají do kontrastu – mohou se lišit nároky na délku předcházejícího kontextu např. u časových určení typu *vloni* – *letos*, která jsou více nasnadě pro vybavení v konkrétní situaci, a u syntaktických předmětů, které tak nasnadě nejsou (*Kancelář* uklidila – *kuchyňku* vynechala). Tyto úvahy bude třeba podložit dalším výzkumem.

⁵ 100 % tvoří vždy výskyty jednoho sémantického typu ve všech 18 větách, např. všechny výskyty konjunkce.

7

Shrnutí

Tato práce představuje další příspěvek k dlouholetému výzkumu stavby věty a stavby textu v Ústavu formální a aplikované lingvistiky na Matematicko-fyzikální fakultě Univerzity Karlovy. Ve výzkumu výstavby textu jsme v ní překročili hranici vztahů vyjádřených explicitně a začali se zabývat vztahy implicitními. Šlo o vykročení do značné nejistoty: nebylo totiž zřejmé, nakolik je vůbec možné se implicitními vztahy seriózně zabývat, zda interpretace anotátorů nebudou natolik různé, že by toto odpoutání od formy vedlo jen do sféry neověřitelných domněnek.

Nicméně se ukázalo, že díky zkušenosti s předchozími anotacemi jsme schopni anotovat implicitní vztahy s přijatelnou mírou shody. Vznikl tak korpus PDiT-EDA 1.0 (Zikánová et al. 2018), který obsahuje nejen anotace implicitních diskurzních vztahů, ale také vztahů explicitních, což poprvé dává veřejnosti k dispozici úplná data o diskurzních vztazích v češtině.

Tato data jsme podrobně analyzovali. Původně jsme v nich hledali zejména rozdíly mezi vztahy explicitními a implicitními; později jsme zjistili, že důležitým hlediskem je také skutečnost, zda se jedná o vztahy vnitrovětné, nebo mezivětné. Objevili jsme tak důležitost strukturní realizace jako parametru, který už budeme muset vždy brát v potaz, ať se budeme zabývat jakoukoli podmnožinou diskurzních vztahů.

Dalším krokem do neznáma vedle rozhodnutí zabývat se implicitními vztahy bylo pro nás zavedení psycholingvistických experimentů do našeho výzkumu. Bylo zřejmé, že to, co získáme, budou první psycholingvistická data o českém diskurzu, tj. data, k nimž prozatím v češtině neexistuje referenční výzkum. Na referenčních datech bychom mohli např. ověřit, zda se někde mezi oběma výzkumy neobjevují rozpory, které by ukazovaly na chybu. Nicméně první krok v tomto směru je hotov: data představená v této práci mohou sloužit jako referenční pro následující psycholingvistické experimenty na češtině.

7.1 Korpus PDiT-EDA 1.0

Shrňme nyní výsledky, které přinesl náš výzkum implicitních diskurzních vztahů v češtině. Pro výzkum byl vytvořen korpus PDiT-EDA 1.0, který obsahuje 100 textů různých žurnalistických žánrů (sportovní zpravodajství, recenze apod.); jeho celkový rozsah je 2 592 vět. Texty korpusu PDiT-EDA 1.0 pocházejí z korpusu PDiT 2.0. Nový korpus převzal z PDiT 2.0 anotaci syntakticko-sémantické struktury věty (tektogramatické roviny), syntaxe, explicitních diskurzních vztahů, koreference, aktuálního členění ad. K nim doplňuje úplnou anotaci vztahů implicitních.

Při anotování implicitních vztahů byly odlišeny a vyznačeny také další druhy vztahů, které dosud nebyly anotovány: vztahy mezi entitami, vztahy diskurzni makrostruktury, vztahy mezi otázkou a odpovědí a také další diskurzni rysy argumentů (titulek, místo, autor ad.).

7.2 Meziannotátorská shoda

Abychom ověřili spolehlivost naší anotace, změřili jsme shodu dvou anotátorů v paralelní anotaci 12 textů (233 vět) z korpusu PDiT-EDA 1.0. V prvním měření jsme sledovali, nakolik se anotátoři shodnou na prosté existenci diskurzniho vztahu mezi stejnými diskurzni argumenty. Shoda tu dosáhla míry $F1$ 0,54, při podrobnějším měření pak 0,59 pro vnitrovětné vztahy a 0,53 pro vztahy mezivětné. Při analýze míst neshody jsme zjistili tři hlavní oblasti, v nichž k neshodě dochází. V prvním případě anotátoři značí existenci implicitního diskurzniho vztahu, ale rozcházejí se v označení velikosti argumentů, které jsou tímto vztahem spojeny. Za druhé, u vnitrovětných vztahů pramenila neshoda často z rozdílné interpretace vedlejších vět vztazných: podle instrukcí měli anotátoři v těchto případech označovat spojení věty řídící a vztazné pouze tehdy, pokud věta vztazná vyjadřovala nezávislý význam (tzv. nepravá věta vedlejší). Záleželo pak na tom, zda anotátoři vyhodnotí vztaznou větu jako nepravou větu vedlejší. Ve třetím případě typických neshod na existenci implicitního diskurzniho vztahu se anotátoři neshodli na tom, zda pozorovaný vztah mezi argumenty patří ke vztahům implicitním, či jiným. Anotátoři se tu rozcházejí v hodnocení jistých výrazů v textu: jeden ho např. pojímá jako primární konektor (označuje vztah jako explicitní), druhý jako rematizátor (tj. vztah značí jako implicitní); jindy jeden vidí v textu sekundární konektor (explicitní vztah), druhý nikoli (anotuje implicitní vztah).

V druhém měření jsme pak zjišťovali, nakolik se anotátoři u těch vztahů, které shodně označili za implicitní, shodnou také na významovém typu vztahu. Celkově se shodli v 58 % případů, hodnota Cohenovy κ tu dosáhla 0,47. U vnitrovětných vztahů dosáhla shoda 38 %, Cohenova κ 0,09, u vztahů mezivětných činila shoda 64 % a Cohenova κ 0,56. Je třeba podotknout, že v paralelní anotaci vyznačili anotátoři poměrně málo vztahů vnitrovětných celkově, proto čísla o shodě jsou u nich spíše orientační. Nicméně poměr mezivětných a vnitrovětných vztahů a paralelní anotaci odpovídá jejich zastoupení v celkovém rozsahu korpusu PDiT-EDA 1.0.

Rozborem míst neshody u tohoto typu měření jsme odhalili tři hlavní okruhy, kde dochází k záměnám vztahů. Jádrem prvního z nich je vztah konjunkce, tyto výskyty bývají hodnoceny také zejména jako kontinua, konjunktivní alternativa či konfrontace. Druhou oblastí je vztah specifikace, zaměřovaný zejména s konjunkcí a explikací. Třetí typ záměn je spojen se vztahem explikace, jíž v interpretaci konkuruje vztah příčiny-důsledku a vztah ekvivalence. Z dat vyplývá, že záměny typů diskurzni vztahů se mohou odlišovat také podle toho, zda jde o vztahy mezivětné, či vnitrovětné. Např. u vztahů vnitrovětných se konjunkce zaměňuje s konfrontací a konjunktivní alternativou, kdežto v mezivětných vztazích si konkuruje se specifikací a kontinuí.

Typicky mezivětné uspořádání mají také vztahy s různověstím explikace / příčina-důsledek.

Porovnali jsme také míry anotátorské shody u implicitních vztahů s výsledky pro shodu u různých dalších typů anotací koherence textu, které proběhly v pražských korpusech dříve: jde o diskurzí vztahy explicitní, textovou koreferenci a asociační anaforu (srov. Mírovský et al. 2010, s. 777; Poláková et al. 2013, s. 96). U tohoto srovnání je však třeba mít na paměti, že různé typy anotace jsou různě náročné. Shodnout se na existenci jistého vztahu může být snazší tam, kde je vztah signalizován nějakým konkrétním slovem či jinou formou; shodnout se na typu je zase snazší tam, kde se rozlišuje typů jen několik.

Z tohoto srovnání vyplynulo, že anotátoři se shodli na existenci vztahů implicitních častěji než u vztahů explicitních. (F1 u implicitních vztahů byla 0,54, u explicitních vztahů 0,43.) Toto zjištění je překvapivé, spíše bychom čekali, že vyšší shoda bude u vztahů explicitních, protože jsou formálně signalizovány konektorem. Pravděpodobně tu sehrál roli úkol anotátorů implicitních vztahů vyznačit i ostatní druhy nesignalizovaných vztahů (otázka – odpověď apod.), který je přinutil setrvalovat u textu déle. U textové koreference byla anotátorská shoda na existenci vztahu podstatně vyšší než u implicitních vztahů (F1 = 0,72), což odpovídá nižší náročnosti tohoto úkolu (vyhledávání výrazů se stejnou referencí). Naproti tomu anotátorská shoda u anotace asociační anafory je nižší (F1 = 0,46); předpokládáme, že tomu tak je proto, že jednak jde o poměrně náročný úkol – najít výrazy s rozdílnou, avšak sémanticky blízkou referencí, jejichž vztah se nevyjadřuje žádnými formálními znaky, jednak anotátoři mohou text interpretovat různě, z povahy úkolu není možné dosáhnout vysoké shody.

Druhou zkoumanou hodnotou je shoda na druhu vztahu u těch vztahů, kde se anotátoři shodli na jejich existenci v daném místě v textu. Shoda u explicitních a implicitních vztahů se značně liší, u explicitních dosáhla 80 % případů, u implicitních 58 %. Jednoznačnější interpretaci sémantického typu u explicitních vztahů zcela jistě napomáhá význam přítomného diskurzího konektoru. U textové koreference a asociační anafory je shoda na druhu vztahu výrazně vyšší než u diskurzí vztahů (textová koreference 90 %, asociační anafora 92 %). Jejich rozlišení je jednodušší, protože se dělí do méně druhů.

Cohenova κ dosáhla u sledovaných anotací následujících hodnot: implicitní vztahy 0,47, explicitní vztahy 0,74, textová koreference 0,73 a asociační anafora 0,89. Vzájemný poměr těchto hodnot odráží skutečnost, že anotace implicitních diskurzí vztahů je nejnáročnějším typem anotace – nebylo při ní možné využít nějakého formálního vyjádření v textu jako vodítka a zároveň se rozpoznávané vztahy dělily do mnoha druhů (23 sémantických typů). Obecně pak toto srovnání dokládá, že anotace méně založené na formě, příp. anotace vyšších jazykových rovin či jevů dosahují nižší míry shody (srov. Poláková 2014, s. 251–255).

7.3 Výsledky anotace v korpusu PDiT-EDA 1.0

Celkem bylo v korpusu PDiT-EDA 1.0 zachyceno 2 715 diskurzních vztahů, přičemž počet implicitních vztahů lehce převažuje nad počtem vztahů explicitních (1 427 vztahů implicitních, 1 288 vztahů explicitních).

Zkoumali jsme také distribuci sémantických typů diskurzních vztahů a souvislost sémantických typů s explicitností či implicitností, tj. s formou vyjádření diskurzních vztahů. (Připomeňme, že korpus umožňuje zkoumat nejen implicitní vztahy, které v něm byly nově anotovány, ale také celek diskurzních vztahů explicitních i implicitních a jejich vzájemné uspořádání.) Přitom se ukázalo, že podstatným parametrem je také strukturní realizace vztahů (vztahy mezivětné a vnitrovětné), proto jsme tento aspekt přiřadili vedle formy vyjádření jako další hledisko pro zkoumání distribuce sémantických typů diskurzních vztahů.

Celkově jsme v anotaci rozlišovali 23 sémantických typů diskurzních vztahů. Pět nejčastějších sémantických typů, které společně představují 71,6 % všech výskytů diskurzních vztahů, tvoří konjunkce (33,4 % všech výskytů), specifikace, příčina-důsledek, opozice a kontinua. Naopak naprosto marginálně jsou zastoupeny vztahy nepravé (nepravá příčina, nepravá opozice a nepravá podmínka), které mají v korpusu PDiT-EDA 1.0 každý méně než 0,2 % všech výskytů.

Na základě kombinace hlediska formy vyjádření a strukturní realizace jsme vymezili čtyři způsoby vyjádření diskurzních vztahů: vztahy mezivětné explicitní (ME), mezivětné implicitní (MI), vnitrovětné explicitní (VE) a vnitrovětné implicitní (VI). Dvě nejčastěji zastoupené skupiny v našich datech jsou vztahy typu MI (47 %) a VE (29 %), méně jsou zastoupeny vztahy typu ME (18 %), pouze okrajově se pak v korpusu nacházejí vztahy typu VI (6 %).

Jednotlivé sémantické typy inklinují ke konkrétnímu způsobu vyjádření v různé míře, proto jsme rozlišili sémantické typy se způsobem vyjádření výrazně vymezeným, mírně vymezeným a nevymezeným. K sémantickým typům s výrazně vymezeným způsobem vyjádření patří např. vztah disjunktivní alternativy (skupina VE) či generalizace a kontinua (MI). Mezi sémantické typy s mírně vymezeným způsobem vyjádření se řadí např. vztah přípustky (VE), dále např. ekvivalence či specifikace (MI). U vztahů s nevymezeným způsobem vyjádření rozlišujeme podle počtu způsobů typického vyjádření vztahy jednovrcholové, dvojvrcholové a trojvrcholové. Typickou dvojvrcholovou kombinací je spojení skupin MI a VE (charakteristické např. pro konjunkci či příčina-důsledek), u trojvrcholového uspořádání je to pak spojení skupin MI, VE a ME (gradace). Skupina VI je vždy okrajová, nepřevažuje u žádného sémantického typu ani v žádném uspořádání s více vrcholy.

Sledovali jsme také, které způsoby vyjádření se pro které sémantické typy v našich datech nevyužívají či využívají se pouze okrajově: např. generalizace není doložena ve skupině VE.

Pro některé sémantické typy je uvedené členění příliš detailní, může se v něm ztratit fakt, že např. jistý sémantický typ bývá explicitní, bez ohledu na strukturní

realizaci. Proto jsme rozčlenili sémantické typy ještě podle každého hlediska zvlášť (forma vyjádření, strukturní realizace). Vymezili jsme tak vztahy typicky explicitní (např. účel), typicky implicitní (např. exemplifikace), typicky mezivětné (např. nesoučasnost) a typicky vnitrovětné (např. podmínka). Také zde jsme popsali, která ztvárnění se pro který sémantický typ (téměř) nevyužívají, např. ekvivalence nebývá explicitní, kontinuační pak nebývá vnitrovětná.

Zvláštní pozornost jsme věnovali způsobům vyjádření pěti nejčastějších sémantických typů (konjunkce, specifikace, příčina-důsledek, opozice a kontinuační). Ukazuje se, že nemají jeden či dva převažující způsoby vyjádření, naopak, každý sice má nějaký typický způsob, ale tyto způsoby se mezi sebou značně liší.

Sledovali jsme také, zda existuje nějaká vazba mezi skupinami sémantických typů jako celky (vztahy časové, kauzální, kontrastivní a navazovací) a způsoby jejich vyjádření. Z výzkumu vyplynulo, že navazovací vztahy se vyjadřují převážně jako MI a naopak, typ MI se využívá především pro vyjádření navazovacích vztahů. Ve skupině vztahů kontrastivních se pak projevuje tendence k explicitní formě. Nicméně celkově nebylo možné pozorovat nějakou typickou vazbu mezi skupinami sémantických typů a způsoby vyjádření: pro jednotlivé sémantické typy z jedné skupiny jsou charakteristické nejednotlivější způsoby vyjádření. Je zajímavé, že způsobem vyjádření se odlišují i vztahy významově blízké (např. současnost a nesoučasnost, konjunktivní a disjunktivní alternativa). Tuto různorodost pozorujeme i v opačném směru: vztahy vyjádřené jedním způsobem (např. VE) mívají nejednotlivější význam, patří k odlišným skupinám sémantických typů.

7.4 Implicitní vztahy v psycholingvistických experimentech

Některá zjištění z analýzy anotací v korpusu PDiT-EDA 1.0 jsme se rozhodli ověřit také v psycholingvistických experimentech. Zjišťovali jsme v nich, nakolik přítomnost diskurzního konektoru napomáhá u různých sémantických typů porozumění textu. Přitom jsme věnovali pozornost sémantickým typům, které jsou významově blízké, avšak v datech PDiT-EDA 1.0 se lišily ve způsobu vyjádření.

Pro experimenty byly vybrány čtyři sémantické typy, a to dva časové (současnost, nesoučasnost) a dva kontrastivní (přípustka, konfrontace). Doklady s těmito sémantickými typy se omezovaly na vztahy mezivětné. Porozumění textu jsme zjišťovali pomocí metody čtení vlastním tempem, při němž jsme zjišťovali, jak rychle jsou respondenti schopni číst s porozuměním doklady s konektory a bez nich.

Výsledky ukazují, že přítomnost diskurzního konektoru má různou závažnost u vztahů časových a u vztahů kontrastivních. U vztahů časových nemá přítomnost konektoru na porozumění textu vliv. Naproti tomu u kontrastivních vztahů napomáhá diskurzní konektor rychlejšímu zpracování textu, u vztahu konfrontace významně, u přípustky lze tento jev pozorovat jako tendenci, nicméně pod hranicí statistické významnosti. U časových vztahů tedy nezáleží na jejich explicitnosti či implicitnosti, kdežto pro zkoumané vztahy kontrastivní je tento rys podstatný.

Výsledky experimentu jsme porovnali s výstupy anotace v korpusu PDiT-EDA 1.0. U časových vztahů se výsledky obou přístupů shodují pro současnost (explicitní či implicitní forma vyjádření nemá na porozumění tomuto vztahu vliv; v anotaci jsou obě formy zastoupeny rovnoměrně). Mírně se odlišují pro nesoučasnost: v datech PDiT-EDA 1.0 bývá nesoučasnost spíše implicitní, v experimentech vychází, že explicitní i implicitní forma jsou stejně přijatelné, na porozumění nemá formy vyjádření vliv. V obou datových souborech u nesoučasnosti tedy platí, že přítomnost diskurzního konektoru je pro porozumění nepodstatná, ani mu nenapomáhá, ani mu nebrání.

Pokud jde o vztahy kontrastivní, jsou výstupy anotace a experimentu v zásadě ve shodě u vztahu přípustky, zatímco u konfrontace se odlišují. V korpusu PDiT-EDA 1.0 je přípustka v mezivětných vztazích vyjádřena stejně často explicitně jako implicitně. Podle experimentu je explicitní forma přípustky lépe srozumitelná než forma implicitní. Implicitní forma přípustky je tedy možná, ale přítomnost konektoru usnadňuje její pochopení. (V tom se přípustka liší od vztahů časových: u nich je také implicitní forma možná, avšak na přítomnosti konektoru nezáleží, nelze říci, že nějak podstatně napomáhá porozumění.)

U mezivětného vztahu konfrontace je v datech PDiT-EDA 1.0 doložena především implicitní forma, zatímco podle experimentu je implicitní forma podstatně hůře srozumitelná, na přítomnosti diskurzního konektoru tu významně záleží. Pro tento rozpor se nabízejí tři vysvětlení. Za prvé se zde může odrážet skutečnost, že v experimentech jsme pracovali pouze s doklady, kde byl přítomen či nepřítomen konektor *zato*, kdežto v anotacích je konfrontace pojata širěji, zahrnuje různé konfrontační konektory a jejich významové odstíny. Za druhé může tento nesoulad poukazovat na fakt, že data v korpusu jsou příliš specifická či malého rozsahu, neodrážejí rysy češtiny jako celku. Za třetí lze vyslovit domněnku, že vztah konfrontace, a zejména v implicitní formě, vyžaduje pro správné pochopení poměrně rozsáhlý předchozí kontext. U dokladů v experimentu však tvořil předchozí kontext pouze jednu větu, proto se tento vztah mohl v implicitní formě ukázat jako výrazně nesrozumitelný.

Shrňme tedy výsledky srovnání výstupů z obou datových souborů. U mezivětných vztahů časových není pro porozumění podstatné, zda jsou explicitně vyjádřeny konektorem. U mezivětných kontrastivních vztahů přípustky a konfrontace je porozumění významně znesnadněno, pokud je konektor nevyjádřen. U přípustky brání vypuštění konektoru porozumění méně, u konfrontace více. Datové soubory vykazují rozpor v přijatelnosti mezivětné implicitní konfrontace. Nabízí se vysvětlení, že vztah implicitní konfrontace vyžaduje pro správnou interpretaci dostatečně rozsáhlý a návodný předchozí kontext.

7.5 Závěr

Implicitní diskurzní vztahy jsou jedním z mnoha subsystémů koherence. V této práci jsme zkoumali jejich uplatnění a distribuci v češtině, a to ve srovnání se vztahy

explicitními. Zjistili jsme, že většina diskurzních vztahů se realizuje právě implicitně, a proto hrají v koherenci textu významnou roli a zaslouží si pozornost v obecné analýze koherence. Ukazuje se, že implicitní vztahy mohou v přiměřených kontextech vyjadřovat nejrůznější významy (byť některé pouze okrajově), nejsou ohraničeny pouze na omezenou množinu sémantických typů. Na druhou stranu podstatná je právě přiměřenost kontextu a dalších přídatných signálů, které napomáhají porozumění. U některých sémantických typů nepředstavuje vypuštění diskurzního konektoru v porozumění textu překážku (např. vztahy časové), zatímco u jiných může implicitnost bránit správnému vyvození vztahu mezi argumenty.

Poznatky plynoucí z tohoto výzkumu lze prakticky využít v oblastech, jež souvisejí s porozuměním jazyku (např. v adaptaci textů pro děti či pro osoby s nižší znalostí jazyka), a v počítačovém zpracování textů (automatický překlad, automatická summarizace apod.). Tyto poznatky mohou například osvětlit, který význam lze přisuzovat kterému způsobu vyjádření diskurzního vztahu v jazyce. Pro explicitní vztahy je základním lexikálním vodítkem samozřejmě význam konektoru, avšak i pro implicitní vztahy můžeme určit, které významy jsou u nich pravděpodobnější než jiné, a to na základě jednak typických způsobů vyjádření, jednak frekvence sémantických typů. Kupř. u mezivětného implicitního spojení lze určit, že s vyšší pravděpodobností má význam specifikace než význam disjunktivní alternativy.

Výzkum implicitních diskurzních vztahů je také příspěvkem k souhrnu poznatků teoretické lingvistiky. Implicitní vztahy jsou totiž částí v systému textové koherence, který má mnoho složek fungujících ve vzájemné složité souhře. Některé z dalších koherenčních subsystémů byly zachyceny v datech pražských korpusů již dříve (aktuální členění, textová a gramatická koreference, asociační anafora, explicitní diskurzní vztahy PDiT 1.0), jiné byly vyčleněny nově jako specifické okruhy v anotaci PDiT-EDA 1.0 vedle implicitních vztahů – vztah mezi otázkou a odpovědí, vztahy mezi entitami, vztahy diskurzní makrostruktury. Vedle těchto faktorů se na koherenci textu podílejí také prostředky syntaktické, lexikální obsazení a asociativní tematická pole spojená s lexikem. Je zřejmé, že tyto a další faktory koherence se v textech navzájem ovlivňují a že se v této souhře uplatňují jisté pravidelné mechanismy. Lze např. očekávat, že implicitnost v jedné složce (a tedy nutnost rekonstrukce) je vyvažována explicitností, návodností v jiné složce. Implicitní vztahy jsme schopni vyvodit díky jiným signálům koherence; domníváme se však, že různé kompenzační mechanismy v koherenci jsou typické nejen pro implicitní diskurzní vztahy. Předpokládáme také, že existují jisté typické vzorce celých souborů různých koherenčních faktorů, na něž jsme zvyklí, a které proto snadno interpretujeme v jednom smyslu a také je snadno vytváříme. Např. může existovat typická větná stavba s typickým průběhem aktuálního členění, která je typicky argumentem v jistém diskurzním vztahu; u takového uspořádání je pak snazší porozumět jeho významu, i když jsou některé části implicitní.

Vedle oblasti textové koherence obecně nás výzkum implicitních vztahů dovedl také k otázkám jazykové vágnosti (neurčitosti), s nimiž jsme se ostatně setkali již při předchozích anotacích. Tento jev se dobře ukázal při analýze mezinotátorské shody.

Zatímco některé výskyty neshod lze vidět jako jevy víceznačné (např. jedno spojení diskurzních argumentů má dvojí přesné významové čtení), u jiných případů je přesné čtení až nadbytečné: význam diskurzního vztahu je velmi volný, někdy dokonce není identifikovatelný. Identifikovatelná je spíše pouze existence vztahu mezi dvěma argumenty, nikoli jeho sémantická povaha. Je pak otázkou pro další výzkum, jak s neurčitostí nakládat v anotaci.

Protože tato práce představuje první ucelený popis explicitních a implicitních diskurzních vztahů na českých datech a také výsledky prvních psycholingvistických experimentů zaměřených na tuto problematiku, přináší také řadu nových otázek, na které se může zaměřit výzkum v budoucnu. Jednou z těchto oblastí je, jak již bylo uvedeno, odhalení typických vzorců s ustáleným způsobem souhry různých koherenčních faktorů, včetně frekvenční analýzy výskytu těchto vzorců. Sem mohou patřit – v jednoduché formě – typické sekvence několika konkrétních diskurzních vztahů za sebou, např. na konkrétním místě v textu. Právě umístění různých jevů koherence na konkrétním místě v textu může být dalším předmětem výzkumu: viděli jsme u vztahu implicitní konfrontace, že nedostatečný předcházející kontext (tedy pozice na začátku textu) u ní může ztěžovat porozumění. Celkově je textová koherence se souhrou všech jazykových subsystémů natolik složitý jev, že větší část jejího popisu a vysvětlení leží stále ještě před námi.

Seznam obrázků

4.1	Tektogramatický strom	29
4.2	Anotace diskurzních vztahů na tektogramatické rovině	38
5.1	Distribuce explic. a implic. vztahů ve vnitrovětné a mezivětné struktuře . . .	82
5.2	Sémantické typy s výrazně vymezeným způsobem vyjádření	86
5.3	Sémantické typy s mírně vymezeným způsobem vyjádření	88
5.4	Sémantické typy s nevymezeným způsobem vyjádření	89
5.5	Způsoby vyjádření pěti nejčastějších sémant. typů v PDiT-EDA 1.0	94
6.1	Konfrontace: doba čtení úseků věty u explicitního a implicitního vztahu . . .	104
6.2	Přípustka: doba čtení úseků věty u explicitního a implicitního vztahu	106
6.3	Současnost: doba čtení úseků věty u explicitního a implicitního vztahu . . .	108
6.4	Nesoučasnost: doba čtení úseků věty u explicitního a implicitního vztahu . .	109
8.1	Annotation of discourse relations on the tectogrammatical level	129
8.2	Semantic types with a clearly specified way of expression	134
8.3	Semantic types with a moderate tendency towards a spec. way of expression	135
8.4	Semantic types with unspecified way of expression	136

Seznam tabulek

4.1 Významová klasifikace diskurzních vztahů v PDTB 2.0	22
4.2 Funktory jako sémantické ohodnocení syntaktického vztahu závislosti	33
4.3 Funktory pro kořeny souřadných struktur	34
4.4 Funktory pro efektivní kořeny nezávislých klauzí	35
4.5 Funktory atomických uzlů	36
4.6 Funktory pro víceslovné jednotky a cizojazyčné výrazy	36
4.7 Významová klasifikace diskurzních vztahů v Pražském diskurzním korpusu	40
4.8 Rozsah PDiT-EDA 1.0 a PDiT 2.0	50
4.9 Zastoupení textových žánrů v korpusu PDiT-EDA 1.0	51
4.10 Povolené překryvy vztahů v anotaci PDiT-EDA 1.0	56
4.11 Mezipřímá shoda u anotace implicitních vztahů v PDiT-EDA 1.0	59
4.12 Mezipřímá shoda u vnitrovětných a mezivětných implic. vztahů	60
4.13 Mezipřímá shoda u významového zařazení mezivětných vztahů	66
4.14 Mezipřímá shoda u významového zařazení vnitrovětných vztahů	67
4.15 Mezipřímá shoda u různých typů textových vztahů	74
5.1 Zastoupení koherenčních vztahů v korpusu PDiT-EDA 1.0	82
5.2 Zastoupení diskurzních vztahů v korpusu PDiT-EDA 1.0	83
5.3 Zastoupení vztahů diskurzní makrostruktury v korpusu PDiT-EDA 1.0	83
5.4 Zastoupení diskurzních rysů v korpusu PDiT-EDA 1.0	83
5.5 Distribuce sémant. typů explic. a implic. vztahů (vnitrovětné, mezivětné)	85
5.6 Řídké či nedoložené způsoby vyjádření sémant. typů v PDiT-EDA 1.0	92
6.1 Rozdělení sad dokladů pro každý sémantický typ mezi respondenty	102
6.2 Konfrontace: výsledky experimentu	103
6.3 Konfrontace: výstup smíšeného regresního modelu	105
6.4 Přípustka: výsledky experimentu	105
6.5 Přípustka: výstup smíšeného regresního modelu	106

6.6	Současnost: výsledky experimentu	107
6.7	Současnost: výstup smíšeného regresního modelu	107
6.8	Nesoučasnost: výsledky experimentu	108
6.9	Nesoučasnost: výstup smíšeného regresního modelu	109
6.10	Srovnání výsledků anotace v PDiT 1.0 a psycholingv. experimentů	110
8.1	Semantic classification of disc. relations in the Prague Discourse Treebank .	127
8.2	Inter-annotator agreement for implicit disc. relations in the PDiT-EDA 1.0 . .	130
8.3	Distribution of coherence relations in the PDiT-EDA 1.0	131
8.4	Distribution of discourse relations in the PDiT-EDA 1.0	132
8.5	Distribution of the relations of discourse macrostructure in the PDiT-EDA 1.0	132
8.6	Distribution of discourse features in the PDiT-EDA 1.0	132
8.7	Semant. types of explic. and implic. relations (intra-sent., inter-sent.) . . .	133
8.8	Results of the mixed regression model	138

Seznam zkratek

Arg1, 2	argument 1 a 2 v diskurzním vztahu
CzeDLex	Lexicon of Czech Discourse Connectives (Mírovský et al. 2017)
EntRel	vztah založený na entitách (entity-based relation)
EvalD	Evaluator of Discourse (Novák et al. 2019)
GeCzLex	Czech–German Lexicon of Anaphoric Connectives (Poláková et al. 2020)
Inter-Expl	inter-sentential explicit discourse relation
Inter-Impl	inter-sentential implicit discourse relation
Intra-Expl	intra-sentential explicit discourse relation
Intra-Impl	intra-sentential implicit discourse relation
MČ3	Mluvnice češtiny 3. 1987. Praha: Academia.
ME	mezivětný explicitní diskurzní vztah
MI	mezivětný implicitní diskurzní vztah
NoRel	mezera v koherenci (no relation)
PDiT 1.0	Prague Discourse Treebank 1.0 (Pražský diskurzní korpus 1.0)
PDiT 2.0	Prague Discourse Treebank 2.0 (Pražský diskurzní korpus 2.0)
PDiT-EDA 1.0	Enriched Discourse Annotation of PDiT Subset 1.0 (Rozšířená anotace diskurzních vztahů na části Pražského diskurzního korpusu 1.0, Zikánová et al. 2018)
PDT	Prague Dependency Treebank (Pražský závislostní korpus)
PDT 2.5	Prague Dependency Treebank 2.5 (Pražský závislostní korpus 2.5)
PDT 3.0	Prague Dependency Treebank 3.0 (Pražský závislostní korpus 3.0)
PDT 3.5	Prague Dependency Treebank 3.5 (Pražský závislostní korpus 3.5)
PDTB	Penn Discourse Treebank
PDTB 2.0	Penn Discourse Treebank 2.0
PDTB 3.0	Penn Discourse Treebank 3.0
PML-TQ	Prague Markup Language – Tree Query
PZK	Pražský závislostní korpus (Prague Dependency Treebank)
RST	Rhetorical Structure Theory
TrEd	Tree Editor TrEd
VE	vnitrovětný explicitní diskurzní vztah
VI	vnitrovětný implicitní diskurzní vztah

Literatura

- Al-Saif, Amal, Katja Markert. 2010. The Leeds Arabic Discourse Treebank: Annotating discourse connectives for Arabic. In: Calzolari, N. et al. (eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta: European Language Resources Association (ELRA), s. 2046–2053.
- Artstein, Ron, Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34 (4), s. 555–596.
- Asher, Nicholas. 1993. *Reference to Abstract Object in Discourse*. Norwell: Kluwer Academic Publishers.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bates, Douglas, Martin Mächler, Ben Bolker, Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67 (1), s. 1–48.
- Bejček, Eduard, Jan Hajič, Jarmila Panevová, Jiří Mírovský, Johanka Spoustová, Jan Štěpánek, Pavel Straňák, Pavel Šidák, Pavlína Vimmrová, Eva Šťastná, Magda Ševčíková, Lenka Smejkalová, Petr Homola, Jan Popelka, Markéta Lopatková, Lucie Hrabalová, Natalia Klyueva, Zdeněk Žabokrtský. 2011. *Prague Dependency Treebank 2.5*. Data/software. Prague: LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <<http://hdl.handle.net/11858/00-097C-0000-0006-DB11-8>> Cit. 14. 2. 2021.
- Bejček, Eduard, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, Šárka Zikánová. 2013. *Prague Dependency Treebank 3.0*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <<http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3>> Cit. 14. 2. 2021.
- Bendová, Klára, Silvie Cinková. Adaptation of Classic Readability Metrics to Czech. In: *Proceedings of Text Speech and Dialogue 2021*, Olomouc, Czech Republic. (V tisku.)
- Braud, Chloé, Pascal Denis. 2014. Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin: Dublin City University and Association for Computational Linguistics, s. 1694–1705. <<https://aclanthology.org/C14-1160>> Cit. 14. 2. 2021.
- Bunt, Harry, Rashmi Prasad. 2016. ISO DR-Core (ISO 24617-8): Core Concepts for the Annotation of Discourse Relations. In: Bunt, H. (ed.), *Proceedings 12th joint ACL-ISO workshop on interoperable semantic annotation (ISA-12)*. Portorož, s. 45–54. <<https://sigsem.uvt.nl/isa12/ISA12Proceedings.pdf>> Cit. 14. 2. 2021.

- Carlson, Lynn, Daniel Marcu, Mary Ellen Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Traum, D. et al. (eds.), *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*. Aalborg: Association of Computational Linguistics. <<https://aclanthology.org/W01-1605>> Cit. 14. 2. 2021.
- Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1), s. 37–46.
- Cohen's kappa, heslo *wikipedie*, <https://en.wikipedia.org/wiki/Cohen%27s_kappa> Cit. 28. 8. 2020.
- Crible, Ludivine, Ágnes Abuczki, Nijolė Burkšaitienė, Péter Furkó, Anna Nedoluzhko, Giedrė Valūnaitė Oleškevičienė, Sigita Rackevičienė, Šárka Zikánová. 2019. Functions and translations of underspecified discourse markers in TED Talks: a parallel corpus study on five languages. In: *Journal of Pragmatics* 142, Amsterdam: Elsevier, s. 139–155.
- Čermáková, Anna, Lucie Jílková, Zuzana Komrsková, Marie Kopřivová, Petra Poukarová. 2019. Diskurzní markery. In: Hoffmannová, J. et al. (eds.), *Syntax mluvené češtiny*. Praha: Academia, s. 244–351.
- Dai, Zeyu and Ruihong Huang. 2018. Improving Implicit Discourse Relation Classification by Modeling Inter-dependencies of Discourse Units in a Paragraph. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans: Association for Computational Linguistics, s. 141–151. <<https://www.aclweb.org/anthology/N18-1013>> Cit. 14. 2. 2021.
- Daneš, František. 1968. Typy tematických poslopností v textu. *Slovo a slovesnost* 29 (2), s. 125–141.
- Das, Debopam, Maite Taboada. 2013. Explicit and implicit coherence relations: A corpus study. In: *Actes du congrès annuel de l'Association canadienne de linguistique 2013. Proceedings of the 2013 annual conference of the Canadian Linguistic Association*. <https://www.sfu.ca/~mtaboada/docs/publications/Das_Taboada_CLA_2013.pdf> Cit. 14. 2. 2021.
- De Beaugrande, Robert-Alain, Wolfgang Dressler. 1981. *Introduction to Text Linguistics*. London: Longman.
- De Saussure, Ferdinand. 1989. *Kurs obecné lingvistiky*. Přel. F. Čermák. Praha: Academia.
- Degand, Liesbeth, Bert Cornillie, Paola Pietrandrea (eds.). 2013. *Discourse Markers and Modal Particles*. Amsterdam: John Benjamins.
- Dressler, Wolfgang. 1972. *Einführung in die Textlinguistik*. Tübingen: Niemeyer.
- Enriched Discourse Annotation of PDiT Subset 1.0 (PDiT-EDA 1.0). Informační webová stránka. <<https://ufal.mff.cuni.cz/pdit-eda1.0>> Cit. 14. 2. 2021.
- Eysenck, Michael W., Mark T. Keane. 2015. *Cognitive Psychology: A Student's Handbook*. London/New York: Psychology Press.
- Fairclough, Norman. 1995. *Critical Discourse Analysis: The Critical Study of Language*. Longman. London/New York: Longman.
- Fedriani, Chiara, Andrea Sansò (eds.). 2017. *Pragmatic Markers, Discourse Markers and Modal Particles*. Amsterdam/Philadelphia: John Benjamins.

- Fernándezová, Eva M., Helen Smithová Cairns. 2014. *Základy psycholingvistiky*. Přel. J. Chromý. Praha: Karolinum.
- Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
- Fischer, Kerstin (ed.). 2006. *Approaches to Discourse*. Oxford: Elsevier.
- Hajič Jan, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Hajičová, Jiří Havelka, Petr Homola, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Petr Pajas, Jarmila Panevová, Lucie Poláková, Magdaléna Rysová, Petr Sgall, Johanka Spoustová, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, Zdeněk Žabokrtský. 2018. *Prague Dependency Treebank 3.5*. Data/software. Prague: Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. <<http://ufal.mff.cuni.cz/pdt3.5>> Cit. 14. 2. 2021.
- Hajič, Jan, Eduard Bejček, Jaroslava Hlaváčová, Marie Mikulová, Milan Straka, Jan Štěpánek, Barbora Štěpánková. 2020a. Prague Dependency Treebank – Consolidated 1.0. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*. Marseille: European Language Resources Association, s. 5208–5218.
- Hajič, Jan, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Fučíková, Eva Hajičová, Jiří Havelka, Jaroslava Hlaváčová, Petr Homola, Pavel Ircing, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, David Mareček, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Michal Novák, Petr Pajas, Jarmila Panevová, Nino Peterek, Lucie Poláková, Popel, Martin, Jan Popelka, Jan Romportl, Magdaléna Rysová, Jiří Semecký, Petr Sgall, Johanka Spoustová, Milan Straka, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jana Šindlerová, Jan Štěpánek, Barbora Štěpánková, Josef Toman, Zdeňka Urešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, Zdeněk Žabokrtský. 2020b. *Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)*. Data/software, LINDAT-CLARIAH, <<http://hdl.handle.net/11234/1-3185>> Cit. 14. 2. 2021.
- Hajičová, Eva. 2012. What We Learned From Complex Annotation of Topic-Focus Articulation in a Large Czech Corpus. In: Radimský, J. (ed.), *Écho des Études romanes* 8, s. 51–64.
- Hajičová, Eva, Jiří Mírovský, Barbora Štěpánková. 2020. Focalizers and Discourse Relations. *The Prague Bulletin of Mathematical Linguistics* 115, s. 187–197.
- Hajičová, Eva. 1993. *Issues of Sentence Structure and Discourse Patterns*. Praha: Univerzita Karlova, nakladatelství Karolinum.
- Halliday, Michael Alexander Kirkwood, Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Harley, Trevor A. 2014. *The Psychology of Language: From Data to Theory*. London/New York: Psychology Press.
- Harris, Zellig Sabetai. 1952. Discourse Analysis. *Language* 28, s. 1–30.
- Hausenblas, Karel. 1971. *Výstavba jazykových projevů a styl*. Praha: Univerzita Karlova.

- Hirschová, Milada. 2017. *Česká věta na rozhraní mezi gramatikou a pragmatikou*. Praha: Univerzita Karlova, nakladatelství Karolinum.
- Hnátková, Milena, Michal Křen, Pavel Procházka, Hana Skoumalová. 2014. The SYN-series corpora of written Czech. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík: European Language Resources Association (ELRA), s. 160–164. <http://www.lrec-conf.org/proceedings/lrec2014/pdf/294_Paper.pdf> Cit. 14. 2. 2021.
- Hoek, Jet, Jacqueline Evers-Vermeul, Ted J. M. Sanders. 2015. The role of expectedness in the implicitation and explication of discourse relations. In: *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)*. Lisbon: Association for Computational Linguistics, s. 41–46. <<https://jethoek.github.io/publications/Hoeketal2016.pdf>> Cit. 14. 2. 2021.
- Hoek, Jet, Hannah Rohde, Jacqueline Evers-Vermeul, Ted J. M. Sanders. 2021. Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience* 36 (3), s. 382–399.
- Hoek, Jet, Merel Scholman. 2017. Evaluating discourse annotation: Some recent insights and new approaches. In: *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (isa-13)*. Montpellier: Tilburg University, Tilburg Center for Cognition and Communication, s. 1–13. <<https://www.aclweb.org/anthology/W17-7401.pdf>> Cit. 14. 2. 2021.
- Hoek, Jet, Sandrine Zufferey. 2015. Factors influencing the implicitation of discourse relations across languages. In: *Proceedings of the 11th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (isa-11)*. London: Association for Computational Linguistics, s. 39–45. <<https://jethoek.github.io/publications/HoekZufferey2015.pdf>> Cit. 14. 2. 2021.
- Hoffmannová, Jana. 2008. Vágnost, neurčitost, nepřesnost, nejistota... ve vztahu k mluveným projevům a jejich zkoumání. In Jaklová, A. (ed.), *Člověk – jazyk – text*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Filozofická fakulta, Ústav bohemistiky, s. 229–234.
- Hoffmannová, Jana, Jiří Homoláč, Eliška Chvalovská, Lucie Jílková, Petr Kaderka, Petr Mareš, Kamila Mrázková. 2016. *Stylistika mluvené a psané češtiny*. Praha: Academia.
- Hrbáček, Josef. 1994. *Nárys textové syntaxe spisovné češtiny*. Praha: Trizonia.
- Chromý, Jan, Silvie Cinková, Jana Šamánková. 2021. Srozumitelnost českého odborného a úředního textu — proč se jí zabývat a jak ji měřit. *Studie z aplikované lingvistiky* 1. Praha: Univerzita Karlova, Filozofická fakulta, s. 35–52.
- Ibex. <<https://spellout.net/ibexfarm/>> Cit. 5. 2. 2021.
- Jegerski, Jill. 2014. Self-paced Reading. In: Jegerski, J. et al. (eds.), *Research Methods in Second Language Psycholinguistics*. New York/London: Routledge, s. 20–49.
- Just, Marcel A., Patricia A. Carpenter, Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111, s. 228–238.
- Karlík, Petr et al. (eds.). 2016a, b. *Nový encyklopedický slovník češtiny I, II*. Praha: Nakladatelství Lidové noviny.

- Kim, Najoung, Song Feng, Chulaka Gunasekara, Luis Lastras. 2020. Implicit Discourse Relation Classification: We Need to Talk about Evaluation. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, s. 5404–5414. <<https://www.aclweb.org/anthology/2020.acl-main.480>> Cit. 14. 2. 2021.
- Kopečný, František, Vladimír Šaur, Václav Polák. 1980. *Etymologický slovník slovanských jazyků – Slova gramatická a zájmena. Sv. 2, Spojky, částice, zájmena a zájmenná adverbia*. Praha: Academia.
- Kosek, Pavel. 2003. *Spojovací prostředky v češtině období baroka*. Ostrava: Ostravská univerzita, Filozofická fakulta.
- Kuznetsova, Alexandra, Per B. Brockhoff, Rune H. B. Christensen. 2017. lmerTest Package: Tests In: Linear Mixed Effects Models. *Journal of Statistical Software* 82 (13), s. 1–26. <<https://doi.org/10.18637/jss.v082.i13>> Cit. 14. 2. 2021.
- Loureda, Óscar, Adriana Cruz, Martha Rudka, Laura Nadal, Ines Recio, Margarita Borreguero Zuloaga. 2015. Focus Particles in Information Processing: An Experimental Study on Pragmatic Scales with Spanish “incluso”. In: De Cesare, A.M. et al. (eds.), *Linguistik online* 71 (2/15), s. 129–151. <https://www.researchgate.net/publication/281491439_Focus_Particles_in_Information_Processing_An_Experimental_Study_on_Pragmatic_Scales_with_Spanish_incluso#fullTextFileContent> Cit. 14. 2. 2021.
- Mann, William. C., Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8 (3), s. 243–281.
- Mathesius, Vilém. 1947. O tak zvaném aktuálním členění větném. In: Mathesius, V., *Čeština a obecný jazykozpyt*. Praha: Melantrich, s. 234–242.
- Mikulová Marie, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský. 2013. *From PDT 2.0 to PDT 3.0 (Modifications and Complements)*. ÚFAL Technical Report, TR-2013-54. Praha: Universitas Carolina Pragensis.
- Mikulová, Marie, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razimová, Petr Sgall, Jan Štěpánek, Zdeňka Uřešová, Kateřina Veselá, Zdeněk Žabokrtský. 2006. *Anotace na tektogramatické rovině Pražského závislostního korpusu*. Technická zpráva. Praha: Univerzita Karlova, Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky.
- Mírovský Jiří, Pavlína Jínová, Lucie Poláková. 2014. Discourse Relations in the Prague Dependency Treebank 3.0. In: *The 25th International Conference on Computational Linguistics (Coling 2014), Proceedings of the Conference System Demonstrations*. Dublin: Dublin City University (DCU), s. 34–38.
- Mírovský, Jiří, Lucie Mladová, Šárka Zikánová. 2010. Connective-Based Measuring of the Inter-Annotator Agreement in the Annotation of Discourse in PDT. In: *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Beijing: Tsinghua University Press, s. 775–781.
- Mírovský Jiří, Pavlína Synková, Magdaléna Rysová, Lucie Poláková. 2017. CzeDLex – A Lexicon of Czech Discourse Connectives. *The Prague Bulletin of Mathematical Linguistics* 109, s. 61–91.

- Mistrík, Jozef. 1966. *Slovosled a vetosled v slovenčine*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- Mistrík, Jozef. 1989. *Štylistika*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Mluvnice češtiny III. *Skladba*. 1987. Praha: Academia.
- Nedoluzhko, Anna. 2011. *Rozšířená textová koreference a asociační anafora. Koncepte anotace českých dat v Pražském závislostním korpusu*. Praha: Univerzita Karlova, Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky.
- Novák, Michal, Jiří Mírovský, Kateřina Rysová, Magdaléna Rysová, Eva Hajičová. 2019. *EVALLD 4.0 – Evaluator of Discourse*. Data/software, LINDAT/CLARIN digital library. Prague: Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. <<http://hdl.handle.net/11234/1-3065>> Cit. 14. 2. 2021.
- Oza, Umangi, Rashmi Prasad, Sudheer Kolachina, Dipti Misra Sharma, Aravind Joshi. 2009. The Hindi Discourse Relation Bank. In: Stede, M. et al. (eds.), *Proceedings of the third Linguistic Annotation Workshop*. Suntec, Singapur: Association for Computational Linguistics, s. 158–161.
- Pajas, Petr, Jan Štěpánek. 2008. Recent Advances in a Feature-Rich Framework for Treebank Annotation. In: Scott, Donia et al. (eds.), *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester: Coling 2008 Organizing Committee, s. 673–680.
- Pajas, Petr, Jan Štěpánek. 2009. System for Querying Syntactically Annotated Corpora. In: *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*. Suntec, Singapur: Association for Computational Linguistics, s. 33–36.
- Panevová, Jarmila. 1980. *Formy a funkce ve stavbě české věty*. Praha: Academia.
- Pešek, Ondřej. 2011. *Argumentativní konektory v současné francouzštině a češtině*. České Budějovice: Jihočeská univerzita v českých Budějovicích.
- Pitler, Emily, Annie Louis, Ani Nenkova. 2009. Automatic sense prediction for implicit discourse relations in text. In: Su, K. et al. (eds.), *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, vol. 2. Suntec, Singapur: Association for Computational Linguistics, s. 683–691.
- Poláková, Lucie. 2014. K možnostem korpusového zpracování nadvětných jevů. *Naše řeč* 97 (4–5), s. 241–258.
- Poláková Lucie, Pavlína Jínová, Jiří Mírovský. 2014. Genres in the Prague Discourse Treebank. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association, s. 1320–1326.
- Poláková, Lucie, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mírovský, Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová. 2012b. *Manual for Annotation of Discourse Relations in the Prague Dependency Treebank*. Technical report no. 2012/47. Prague: Charles University in Prague, Institute of Formal and Applied Linguistics.
- Poláková, Lucie, Pavlína Jínová, Šárka Zikánová, Eva Hajičová, Jiří Mírovský, Anna Nedoluzhko, Magdaléna Rysová, Veronika Pavlíková, Jana Zdeňková, Jiří Pergler, Radek

- Ocelák. 2012a. *Prague Discourse Treebank 1.0*. Data/software. Prague: LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <<http://hdl.handle.net/11858/00-097C-0000-0008-E130-A>> Cit. 14. 2. 2021.
- Poláková, Lucie, Jiří Mírovský, Anna Nedoluzhko, Pavlína Jínová, Šárka Zikánová, Eva Hajičová. 2013. Introducing the Prague Discourse Treebank 1.0. In: *Proceedings of the 6th International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, s. 91–99.
- Poláková, Lucie, Jiří Mírovský, Pavlína Synková. 2017. Signalling Implicit Relations: A PDTB – RST Comparison. *Dialogue and Discourse* 8 (2), s. 225–248.
- Poláková, Lucie, Jiří Mírovský, Šárka Zikánová, Eva Hajičová. 2021. Discourse Relations and Connectives in Higher Text Structure. *Dialogue and Discourse* 12 (2), s. 1–37.
- Poláková, Lucie, Kateřina Rysová, Magdaléna Rysová, Jiří Mírovský. 2020. GeCzLex: Lexicon of Czech and German Anaphoric Connectives. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*. Marseille: European Language Resources Association, s. 1082–1089. <<http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.137.pdf>> Cit. 14. 2. 2021.
- Prague Discourse Treebank 2.0 (PDiT 2.0)*. Informační webová stránka. <<https://ufal.mff.cuni.cz/pdit2.0>> Cit. 20. 4. 2021.
- Prasad, Rashmi, Nikhil Dinesh, Ala Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, Bonnie Webber. 2008b. The Penn Discourse TreeBank 2.0. In: Calzolari, N. et al. (eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakech: European Language Resources Association (ELRA), s. 2961–2968.
- Prasad, Rashmi, Alan Lee, Nikhil Dinesh, Eleni Miltsakaki, Geraud Campion, Aravind Joshi, Bonnie Webber. 2008a. *Penn Discourse Treebank Version 2.0*. LDC2008T05. Philadelphia: Linguistic Data Consortium. <<https://catalog.ldc.upenn.edu/LDC2008T05>> Cit. 14. 2. 2021.
- Prasad, Rashmi, Susan McRoy, Nadya Frid, Aravind Joshi, Hong Yu. 2011. The Biomedical Discourse Relation Bank. *BMC Bioinformatics* 12, s. 188–205.
- Prasad, Rashmi, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, Bonnie Webber. 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. Technical report IRCS-08-01. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania. <<https://www.seas.upenn.edu/pdtb/PDTBAPI/pdtb-annotation-manual.pdf>> Cit. 14. 2. 2021.
- Prasad, Rashmi, Bonnie Webber, Alan Lee, Aravind Joshi. 2019. *Penn Discourse Treebank Version 3.0*. LDC2019T05. Philadelphia: Linguistic Data Consortium. <<https://catalog.ldc.upenn.edu/LDC2019T05>> Cit. 14. 2. 2021.
- Rohde, Hannah, Alexander Johnson, Nathan Schneider, Bonnie Webber. 2018. Discourse Coherence: Concurrent Explicit and Implicit Relations. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne: Association for Computational Linguistics, s. 2257–2267. <<https://www.aclweb.org/anthology/P18-1210>> Cit. 14. 2. 2021.

- Rysová, Kateřina, Magdaléna Rysová, Jiří Mírovský. 2016. Automatic evaluation of surface coherence in L2 texts in Czech. In: *Proceedings of the 28th Conference on Computational Linguistics and Speech Processing ROCLING XXVIII*. Taipei, Taiwan: The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), s. 214–228. <<http://aclweb.org/anthology/O/O16/O16-1021.pdf>> Cit. 14. 2. 2021.
- Rysová, Magdaléna. 2012. Alternative Lexicalizations of Discourse Connectives in Czech. In: Calzolari, N. et al. (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA).
- Rysová, Magdaléna. 2018. *Diskurzvní konektory v češtině: Od centra k periferii*. Praha: Matematicko-fyzikální fakulta UK, Ústav formální a aplikované lingvistiky.
- Rysová, Magdaléna, Lucie Poláková, Jiří Mírovský, Pavlína Synková. 2018. Describing CzeDLex – a Lexicon of Czech Discourse Connectives. In: *TextLink – Structuring Discourse in Multilingual Europe – Final Action Conference*. Toulouse: University of Toulouse, s. 129–135.
- Rysová, Magdaléna, Pavlína Synková, Jiří Mírovský, Eva Hajičová, Anna Nedoluzhko, Radek Ocelák, Jiří Pergler, Lucie Poláková, Veronika Scheller, Jana Zdeňková, Šárka Zikánová. 2016. *Prague Discourse Treebank 2.0*. Prague: LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <<http://hdl.handle.net/11234/1-1905>> Cit. 14. 2. 2021.
- Sanders, Ted J. M., Vera Demberg, Jet Hoek, Merel C. J. Scholman, Fatemeh Torabi Asr, Sandrine Zufferey, Jacqueline Evers-Vermeul. 2018. Unifying Dimensions in Coherence Relations: How Various Annotation Frameworks Are Related. *Corpus Linguistics and Linguistic Theory* 17 (1), s. 1–71. <<https://www.degruyter.com/document/doi/10.1515/cllt-2016-0078/html>> Cit. 14. 2. 2021.
- Sgall, Petr, Eva Hajičová, Eva Buráňová. 1980. *Aktuální členění věty v češtině*. Praha: Academia.
- Sgall, Petr, Eva Hajičová, Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publishing Company.
- Sgall, Petr, Ladislav Nebeský, Alla Goralčíková, Eva Hajičová. 1969. *A Functional Approach to Syntax in Generative Description of Language*. New York: American Elsevier.
- Schiffrin, Deborah. 1994. *Approaches to Discourse*. Blackwell Textbooks in Linguistics. Oxford: Blackwell.
- Smolík, Filip. 2015. Word order and information structure in Czech 3-and 4-year-olds' comprehension. *First Language* 35 (3), s. 237–253.
- Smolík, Filip, Jiří Lukavský. 2009. Word Order and Case Inflection in Czech: On-line Sentence Comprehension in Children and Adults. In: Taatgen, N. (ed.), *Proceedings of the Annual Meeting of Cognitive Science Society (COGSCI 2009)*. Amsterdam: Cognitive Science Society, s. 1358–1363. <<http://csjarchive.cogsci.rpi.edu/proceedings/2009/papers/302/index.html>> Cit. 14. 2. 2021.
- Speyer, Augustin, Anita Fetzer. 2014. The coding of discourse relations in English and German argumentative discourse. In: Gruber, H. et al. (eds.), *The Pragmatics of Discourse Coherence: Theories and Applications*. Amsterdam/Philadelphia: John Benjamins, s. 87–119.

- Stede, Manfred. Resolving connective ambiguity. 2014. In: Gruber, H. et al. (eds.), *The Pragmatics of Discourse Coherence: Theories and Applications*. Amsterdam/Philadelphia: John Benjamins, s. 121–141.
- Stede, Manfred, Arne Neumann. 2014. Potsdam Commentary Corpus 2.0: Annotation for discourse research. In: Calzolari, N. et al. (eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík: European Language Resources Association (ELRA), s. 925–929.
- Svoboda, Aleš. 1984. České slovosledné pozice z pohledu aktuálního členění. *Slovo a slovesnost* 45, I, s. 22–34; II, s. 88–103.
- Svoboda, Karel. 1972. *Souvěťi spisovné češtiny*. Acta Universitatis Carolinae. Praha: Univerzita Karlova.
- Štícha, František. 2007. Vágnost jazyka. *Naše řeč* 90, s. 99–101.
- Taboada, Maite. 2009. Implicit and Explicit Coherence Relations. In: Renkema, J. (ed.), *Discourse, of course*. Amsterdam: John Benjamins, s. 127–140.
- Tárnyiková, Jarmila. 2012. A Text-level Approach to Sentence Complexes (English-Czech Interface). In: Molnár, O. et al. (eds.), *Text as a Dynamic Interplay of Text Parameters*. Olomouc: Univerzita Palackého v Olomouci, s. 9–21.
- Tonelli, Sara, Giuseppe Riccardi, Rashmi Prasad, Aravind Joshi. 2010. Annotation of Discourse Relations for Conversational Spoken Dialogs. In: Calzolari, N. et al. (eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta: European Language Resources Association (ELRA), s. 2084–2090.
- Torabi Asr, Fatemeh, Vera Demberg. 2012. Implicitness of Discourse Relations. In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*. Mumbai: Association for Computational Linguistics, s. 2669–2684.
- Uhlířová, Ludmila. 1987. *Knížka o slovosledu*. Praha: Academia.
- van der Vliet, Nynke, Gisela Redeker. 2014. Explicit and implicit coherence relations in Dutch texts. In: Gruber, H. et al. (eds.), *The Pragmatics of Discourse Coherence: Theories and Applications*. Amsterdam/Philadelphia: John Benjamins, s. 23–52.
- van Dijk, Teun A. 1980. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.
- Webber, Bonnie, Aravind Joshi. 1998. Anchoring a lexicalized tree-adjoining grammar for discourse. In: Stede, M. et al. (eds.), *Discourse Relations and Discourse Markers: Proceedings of the Conference*. Somerset, New Jersey: Association for Computational Linguistics, s. 86–92.
- Webber, Bonnie, Rashmi Prasad, Alan Lee, Aravind Joshi. 2019. *The Penn Discourse Tree-bank 3.0 Annotation Manual*. <<https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>> Cit. 14. 2. 2021.
- Webber, Bonnie, Hannah Rohde, Anna Dickinson, Annie Louis, Nathan Schneider. 2018. Explicit Discourse Connectives / Implicit Discourse Relations. In: *Proceedings of the Society for Computation in Linguistics*, Vol. 1, Article 40, s. 228–230 <<https://scholarworks.umass.edu/scil/vol1/iss1/4>> Cit. 14. 2. 2021.

- Wolf, Florian, Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Computational Linguistics* 31 (2), s. 249–287.
- Zeyrek, Deniz, Isin Demirşahin, Ayışi B. Sevdik-Çallı's, Ruket Çakıcı. 2013. Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue and Discourse* 4 (2), s. 174–184.
- Zhou, Yuping, Jill Lu, Jennifer Zhang, Nianwen Xue. 2014. *Chinese Discourse Treebank 0.5*. LDC2014T21. Philadelphia: Linguistic Data Consortium. <<https://catalog.ldc.upenn.edu/LDC2014T21>> Cit. 14. 2. 2021.
- Zhou, Yuping, Nianwen Xue. 2012. PDTB-style discourse annotation of Chinese text. In: H. Li et al. (eds.), *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*. Jeju: Association for Computational Linguistics, s. 69–77.
- Zikánová, Šárka, Eva Hajičová, Barbora Hladká, Pavlína Jínová, Jiří Mírovský, Anna Nedoluzhko, Lucie Poláková, Kateřina Rysová, Magdaléna Rysová, Jan Václ. 2015a. *Discourse and Coherence. From the Sentence Structure to Relations in Text*. Praha: Univerzita Karlova, Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky.
- Zikánová, Šárka, Jiří Mírovský, Pavlína Synková. 2019. Explicit and Implicit Discourse Relations in the Prague Discourse Treebank. In: Ekštejn K. (ed.), *Proceedings of the 22nd International Conference on Text, Speech and Dialogue – TSD 2019, Lecture Notes in Computer Science*, vol. 11697. Cham/Heidelberg/New York/Dordrecht/London: Springer, s. 236–248.
- Zikánová, Šárka, Lucie Mladová, Jiří Mírovský, Pavlína Jínová. 2010. Typical Cases of Annotators' Disagreement in Discourse Annotations in Prague Dependency Treebank. In: Calzolari, N. et al. (eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta: European Language Resources Association (ELRA), s. 2002–2006. <<http://www.lrec-conf.org/proceedings/lrec2010/index.html>> Cit. 14. 2. 2021.
- Zikánová, Šárka, Lucie Poláková, Pavlína Jínová, Anna Nedoluzhko, Magdaléna Rysová, Jiří Mírovský, Eva Hajičová. 2015b. Zachycení výstavby textu v Pražském závislostním korpusu. *Slovo a slovesnost* 76, s. 163–197.
- Zikánová, Šárka, Pavlína Synková, Jiří Mírovský. 2018. *Enriched Discourse Annotation of PDiT Subset 1.0 (PDiT-EDA 1.0)*. Data/Software. Prague: LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <<http://hdl.handle.net/11234/1-2906>> Cit. 14. 2. 2021.
- Zikánová, Šárka, Miroslav Týnovský. 2009. Identification of Topic and Focus in Czech: Comparative Evaluation on Prague Dependency Treebank. In: Zybatow, G. et al. (eds.), *Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure. Formal Description of Slavic Languages 7*. Frankfurt am Main: Peter Lang, s. 343–353.
- Zufferey, Sandrine, Pim Mak, Liesbeth Degand, Ted J. M. Sanders. 2015. Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second language research* 31 (3), s. 1–23.

Rejstřík

- abstraktní objekt, 19
- aktuální členění věty, 28
- analytická rovina, 28
- asociační anafora, 28
- atribuce, 20, 56

- čtení vlastním tempem, 97

- diskurz, 9
- diskurzní argument, 10
 - číslování, 20, 23, 37
- diskurzní konektor, 11
 - alternativní lexikalizace, 20
 - primární, 11, 81
 - sekundární, 11, 81
 - v PDiT, 37
- diskurzní vztah, 12
 - explicitní, 12, 19
 - implicitní, 12, 19
 - mezivětný, 52
 - explicitní, 84
 - implicitní, 84
 - vnitrovětný, 52
 - explicitní, 84
 - implicitní, 84

- forma vyjádření diskurzního vztahu, 84
- funkce a forma, 27, 37
- funkční generativní popis, 17, 27
- funktor, 30

- koreference, 28

- makrostruktura, 52, 54
- mezera v koherenci, 20, 55
- morfologická rovina, 28

- otázka a odpověď, 54

- PDiT-EDA 1.0, 49

- Penn Discourse Treebank 2.0, 16, 19
- Pražský závislostní korpus, 17, 27
- princip minimality, 11, 19

- smíšené regresní modely, 103
- strukturní realizace diskurzního vztahu, 84

- tektogramatická rovina, 28
- text, 9
- textový žánr, 50

- Ústav formální a aplikované lingvistiky, 17

- věty vedlejší vztažné, 62
- vztah mezi entitami, 20
- vztahy časové
 - PDiT, 39, 108
 - PDTB, 21
- vztahy kauzální
 - PDiT, 41, 95
 - PDTB, 21
- vztahy kontrastivní
 - PDiT, 43, 108
 - srovnávací, PDTB, 25
- vztahy navazovací
 - PDiT, 45, 94, 95
 - rozšiřující, PDTB, 26
- vztahy nepravé
 - pragmatické, PDTB, 21

A

Výchozí sada dokladů pro experiment s kontrastivními vztahy

V níže uvedených experimentálních sadách uvádíme pouze doklady vztahů explicitních. U implicitních vztahů byly použity tytéž doklady, byl v nich však vypuštěn diskurzní konektor na počátku druhého argumentu.

A.1 Přípustka

- 1 Kontext: Petr byl v náročné situaci.
 1. argument: Pořád | měl | obrovské | dluhy.
 2. argument: Přesto | dál | kupoval | drahé | zbytečnosti.
 Kontext: Blížil se | neodvratný | krach.
 Kontrolní otázka: Je pravda, že Petr se snažil krach odvrátit?
 Odpověď: Ne

- 2 Kontext: Sestra mi vysloveně mizela před očima.
 1. argument: Nakonec | ztratila | veškerou | energii.
 2. argument: Přesto | v létě | zaběhla | budapeštský | maratón.
 Kontext: Od té doby | radši | nesoudím, kdo dokáže co.
 Kontrolní otázka: Dokázala sestra běžet na dlouhou vzdálenost?
 Odpověď: Ano

- 3 Kontext: Mirek měl nejradši svůj klid.
 1. argument: Odjakživa | nenáviděl | rozbouřené | moře.
 2. argument: Přesto | vloni | obeplul | celý | svět.
 Kontext: Prý si to | úžasně | užil.
 Kontrolní otázka: Líbila se Mirkovi cesta kolem světa?
 Odpověď: Ano

- 4 Kontext: Toho bankéře znám, je to blázen.
1. argument: Doma | měl | velké | problémy.
2. argument: Přesto | vždycky | odmítl | nabízenou | pomoc.
Kontext: Myslel si | samozřejmě, | že všechno zvládne sám.
Kontrolní otázka: Měl bankéř problémy?
Odpověď: Ano
- 5 Kontext: Hanuš byl neuvěřitelně tvrdohlavý.
1. argument: V dětství | ztratil | barevné | vidění.
2. argument: Přesto | na zámku | maloval | hostinské | pokoje.
Kontext: Polepil si | barvy | štítky s nápisy a ptal se komorné na radu.
Kontrolní otázka: Vystačil si Hanuš při malování pokojů sám?
Odpověď: Ne
- 6 Kontext: Na Cilku myslím pořád a musím se smát.
1. argument: V Praze | vlastnila | výnosný | podnik.
2. argument: Přesto | pořád | obývala | starou | maringotku.
Kontext: Jednou | kradla | ve svém vlastním bufetu.
Kontrolní otázka: Vyplyvá z textu, že Cilka byla bohatá?
Odpověď: Ano
- 7 Kontext: Irene v detektivce dokáže neuvěřitelné věci.
1. argument: Po vraždě | ztratila | jediný | klíč.
2. argument: Přesto | nakonec | otevřela | zamčené | dveře.
Kontext: Odšroubovala | panty u dveří | pomocí zahradní motyčky.
Kontrolní otázka: Vyplyvá z textu, že Irene dveře odemkla?
Odpověď: Ne
- 8 Kontext: To byste měli vidět naschvály naší malé Aničky.
1. argument: Po výletě | měla | velkou | žízeň.
2. argument: Přesto | hned | vylila | nabízenou | vodu.
Kontext: Láhev | zahodila | a ještě na ni vyplázla jazyk.
Kontrolní otázka: Napila se Anička po výletě?
Odpověď: Ne

- 9 Kontext: Marek, to byl herec za všechny peníze.
 1. argument: V divadle | potřeboval | hudební | nástroj.
 2. argument: Přesto | brzo | prodal | tahací | harmoniku.
 Kontext: To, co dostal, | propil | a pak přemýšlel, jak předvést
 harmoniku pantomimicky.
 Kontrolní otázka: Je pravda, že Marek získané peníze prohrál na automa-
 tech?
 Odpověď: Ne
- 10 Kontext: Andulka si v kadeřnickém velkoobchodě všechno pře-
 dem domluvila.
 1. argument: Na internetu | objednala | drahý | fén.
 2. argument: Přesto | za týden | obdržela | stolní | lampičku.
 Kontext: Bylo to | pro ni | velké zklamání.
 Kontrolní otázka: Podařil se Andulce nákup tak, jak si představovala?
 Odpověď: Ne
- 11 Kontext: Naše teta měla už opravdu namále.
 1. argument: V lese | ztratila | jedinou | baterku.
 2. argument: Přesto | vzápětí | našla | nenápadnou | cestičku.
 Kontext: Vedla | k vojenskému bunkru | a odtamtud do hospody.
 Kontrolní otázka: Byla cesta, kterou teta našla, nenápadná?
 Odpověď: Ano
- 12 Kontext: Mazunga, to byl také zajímavý vládce.
 1. argument: Jednou | vyhlásil | okamžitou | válku.
 2. argument: Přesto | ihned | rozpustil | připravené | vojsko.
 Kontext: Vyšel | nepříteli | v ústrety sám a po cestě se ztratil.
 Kontrolní otázka: Dal Mazunga rozpustit své vojsko?
 Odpověď: Ano
- 13 Kontext: Honza byl na sebe vždycky hrozně opatrný.
 1. argument: Odjakživa | měl | potravinovou | alergii.
 2. argument: Přesto | na vernisáži | jedl | lískové | oříšky.
 Kontext: Teď opuchl | a proklíná | umění.
 Kontrolní otázka: Vyplývá z textu, že Honza jedl na vernisáži krevety?
 Odpověď: Ne

- 14 Kontext: Helena, ta prostě musí být v centru pozornosti.
 1. argument: Vždycky | zbožňovala | hlučnou | společnost.
 2. argument: Přesto | letos | zrušila | narozeninovou | oslavu.
 Kontext: Nevíme, | co se stalo, | možná má nemocného psa.
 Kontrolní otázka: Vyplývá z textu, že Helena měla na oslavě psa?
 Odpověď: Ne
- 15 Kontext: Rodiče se rozhodli koupit Milanovi něco na sport.
 1. argument: Dlouho | chtěl | nové | kolo.
 2. argument: Přesto | nakonec | dostal | školní | atlas.
 Kontext: Maminku | atlas ve výloze | upoutal cestou do cyklistické-
 ho obchodu.
 Kontrolní otázka: Dostal Milan jízdní kolo?
 Odpověď: Ne
- 16 Kontext: Vašek už je pomalu nebezpečný sám sobě.
 1. argument: Letos | plánoval | celkové | zvolnění.
 2. argument: Přesto | v únoru | přijal | náročnou | práci.
 Kontext: Teď je | nevyspalý | a pořád přitom řídí.
 Kontrolní otázka: Je pravda, že si Vašek chtěl odpočinout?
 Odpověď: Ano
- 17 Kontext: Zoška v cestovní kanceláři vysloveně zápasila
 s technikou.
 1. argument: V Polsku | připravovala | velký | zájezd.
 2. argument: Přesto | před cestou | objednala | sedmimístný | mikrobuse.
 Kontext: Vymazala si | totiž | v objednávkovém formuláři nulu
 v čísle sedmdesát.
 Kontrolní otázka: Mělo na zájezd jet více než sedm lidí?
 Odpověď: Ano
- 18 Kontext: Dlouho jsem učil na housle taky malého Miloše.
 1. argument: Denně | cvičil | lehounkou | skladbu.
 2. argument: Přesto | v červnu | pokazil | závěrečnou | zkoušku.
 Kontext: Dneska má | country kapelu | a kupodivu hrají i tuhle
 skladbu, s doprovodem banja.
 Kontrolní otázka: Je pravda, že Miloš před zkouškou téměř necvičil?
 Odpověď: Ne

- 19 Kontext: Adámek s vyplazeným jazykem pomalu nesl tácek s nádobím.
 1. argument: Po cestě | dával | velký | pozor.
 2. argument: Přesto | na schodech | rozbil | mamčinu | sklenici.
 Kontext: Podíval se | na střepy | a utěšil je, ať nepláčou.
 Kontrolní otázka: Vyplyvá z textu, že se Adámkovi sklenice rozbila?
 Odpověď: Ano
- 20 Kontext: Na jaře se Bára zase vrátila do atletického oddílu, do soutěže.
 1. argument: Po nemoci | slibovala | malý | výkon.
 2. argument: Přesto | rovnou | předvedla | úžasné | zlepšení.
 Kontext: Jako by | nepřijela | z nemocnice, ale ze sportovního soustředění.
 Kontrolní otázka: Byla Bára na sportovním soustředění?
 Odpověď: Ne

A.2 Konfrontace

- 1 Kontext: Nakonec se orchestr pokusil o štěstí i v zahraničí.
 1. argument: V Německu | letos | víceméně | zapadl.
 2. argument: Zato | v Anglii | koncem roku | výrazně | prorazil.
 Kontext: V cestování | se chystá | pokračovat i nadále.
 Kontrolní otázka: Navštívil orchestr letos alespoň dvě země?
 Odpověď: Ano
- 2 Kontext: Zajímavý je vývoj počtu žáků.
 1. argument: Na venkově | každoročně | nepatrně | klesá.
 2. argument: Zato | v Praze | trvale | mírně | stoupá.
 Kontext: Zřejmě je | vyšší | nejen o žáky, kteří se stěhují do Prahy, ale také o ty, kteří dojíždějí.
 Kontrolní otázka: Je pravda, že v Praze stoupá počet žáků?
 Odpověď: Ano
- 3 Kontext: Děti se ve třídě hádaly, jestli Petříka přijmou.
 1. argument: Podle Alenky | pořád | nesnesitelně | žaloval.
 2. argument: Zato | podle Emilky | tady | neprávem | trpěl.
 Kontext: Zastávala | se ho a | litovala ho.
 Kontrolní otázka: Byl někdo ve třídě na Petříkově straně?
 Odpověď: Ano

- 4 Kontext: Szándor Nagy se zamýšlí nad vztahem střední Evropy k Vyšegrádské skupině.
1. argument: Maďarsko | většinou | nadšeně | přijímá.
2. argument: Zato | Slovensko | obecně | spíše | odsuzuje.
Kontext: Jeho nová kniha | vzbudila | na Slovensku poměrně vzrušenou diskusi.
Kontrolní otázka: Vyslovuje se autor knihy také k roli Ruska?
Odpověď: Ne
- 5 Kontext: Z nové uklízečky jsem trochu rozpačitá.
1. argument: Kancelář | včera | celkem | uklidila.
2. argument: Zato | kuchyňku | přitom | úplně | vynechala.
Kontext: Uvidíme, | jak to bude | pokračovat.
Kontrolní otázka: Vyplývá z textu, že uklízečka v kanceláři uklidila?
Odpověď: Ano
- 6 Kontext: Dědeček si nenašel cestu k literatuře hned, býval spíš divoch.
1. argument: V mládí | v Sokole | úspěšně | střílel.
2. argument: Zato | po válce | často | doma | psal.
Kontext: Zachovala se | nám od něj | dvoudílná rodová kronika a sbírky básní.
Kontrolní otázka: Vyplývá z textu, že dědeček nikdy nesportoval?
Odpověď: Ne
- 7 Kontext: Projekt pro mateřskou školu navrhla nakonec prestižní architektonická firma Vojáček company.
1. argument: Technikálie | tentokrát | spíše | načrtla.
2. argument: Zato | design | v návrhu | detailně | rozpracovala.
Kontext: Propojuje | budovu | s venkovním prostorem tak, že prvky zahrady slouží zároveň i jako výzdoba při pohledu z interiéru.
Kontrolní otázka: Je firma, která navrhuje projekt pro školku, známá?
Odpověď: Ano

- 8 Kontext: Ostatně výrazná změna ve sportovním oboru se nevyhnula ani např. Janu Ohradilovi.
 1. argument: V mládí | v Třebíči | závodně | běhal.
 2. argument: Zato | nyní | v Dukle | vytrvalostně | plave.
 Kontext: Tento posun | přitom | vidí jako naprosto logický.
 Kontrolní otázka: Je pravda, že Jan Ohradil změnil sportovní obor?
 Odpověď: Ano
- 9 Kontext: Podle agentury DAKO podléhá výkyvům také cena pohonných hmot.
 1. argument: V září | ve světě | prudce | poklesla.
 2. argument: Zato | v říjnu | na trzích | postupně | stoupala.
 Kontext: V současné době | se drží | spíše na vyšší úrovni.
 Kontrolní otázka: Byla cena pohonných hmot na podzim stálá?
 Odpověď: Ne
- 10 Kontext: V interdisciplinárním programu byl za tu dobu podán pouze jediný, biologicko-historický projekt.
 1. argument: U historiků | vloni | jednohlasně | prošel.
 2. argument: Zato | v biologii | potom | naprosto | propadl.
 Kontext: Finanční prostředky | tedy nakonec | vyplaceny nebyly.
 Kontrolní otázka: Je pravda, že projekt uspěl u obou komisí?
 Odpověď: Ne
- 11 Kontext: Vloni jsme přijali do šestého ročníku jednu nerodilou mluvčí.
 1. argument: Zpočátku | všude | téměř | propadala.
 2. argument: Zato | později | u zkoušek | vysloveně | zazářila.
 Kontext: Přispěly | k tomu | také hodiny české konverzace děti z cizího prostředí pořádané školou.
 Kontrolní otázka: Je pravda, že se prospěch nové žákyně zlepšil?
 Odpověď: Ano
- 12 Kontext: Počasí se v průběhu týdne nemění.
 1. argument: V Čechách | prozatím | stále | prší.
 2. argument: Zato | na Moravě | dosud | ještě | mrzne.
 Kontext: Ve čtvrtek | i zde | přejdou sněhové přeháňky v deštivé.
 Kontrolní otázka: Je pravda, že na Moravě se oteplí později než v Čechách?
 Odpověď: Ano

- 13 Kontext: Pokud šlo o zaslané dárky, měl starý kníže svoje pevné zásady.
 1. argument: Většinu | letos | normálně | přijal.
 2. argument: Zato | desetinu | po prohlédnutí | rozhodně | odmítl.
 Kontext: Pocházely | prý ze zemí, | které nikdy nenavštívil.
 Kontrolní otázka: Vyplývá z textu, že kníže přijal všechny dary?
 Odpověď: Ne
- 14 Kontext: Strýce Martina jsem měl obzvlášť rád.
 1. argument: V soukromí | vždy | jednal | odvázně.
 2. argument: Zato | na veřejnosti | často | působil | nesměle.
 Kontext: Naučil mě | lyžovat | a také se nebát výšek.
 Kontrolní otázka: Je pravda, že strýc Martin působil na veřejnosti neohroženě?
 Odpověď: Ne
- 15 Kontext: Sníh zatím z našich hor nezmizel.
 1. argument: Na Šumavě | na svazích | pomalu | taje.
 2. argument: Zato | v Krkonoších | všude | rychle | přibývá.
 Kontext: Lyžařská sezóna | je tu stále | v plném proudu.
 Kontrolní otázka: Je pravda, že na Šumavě už sníh vůbec není?
 Odpověď: Ne
- 16 Kontext: Ten nádherný prsten měl jednu zvláštní vlastnost.
 1. argument: Na slunci | vždycky | zářivě | zbělal.
 2. argument: Zato | ve tmě | všude | dočista | ztmavl.
 Kontext: A když se | setkal | s pravou láskou, rudě se rozzářil.
 Kontrolní otázka: Mohl být prsten aspoň někdy bílý?
 Odpověď: Ano
- 17 Kontext: Andrej si tady nakonec zvykl, trochu mu v tom pomohl sport.
 1. argument: Fotbal | v Česku | spíše | opomíjí.
 2. argument: Zato | hokeji | nyní | dobře | rozumí.
 Kontext: Komentuje ho | typickými | ironickými poznámkami.
 Kontrolní otázka: Zajímá Andreje fotbal stejně jako hokej?
 Odpověď: Ne

- 18 Kontext: Po prázdninách jsme tenkrát dostali nového třídního učitele.
 1. argument: Dívky | většinou | spíš | vystrašil.
 2. argument: Zato | chlapcům | zatím | celkově | vyhovoval.
 Kontext: Byl dost | rázný a | dával najevo, že si nenechá nic líbit.
 Kontrolní otázka: Vyhovoval nový učitel spíše chlapcům?
 Odpověď: Ano
- 19 Kontext: Maruška slušně poděkovala a posadila se na pařez.
 1. argument: Jablíčko | na paloučku | hned | snědla.
 2. argument: Zato | koláč | zatím | raději | nechala.
 Kontext: Kdo ví, zda | potká | další dobré lidi?
 Kontrolní otázka: Snědla si Maruška jablíčko a koláč najednou?
 Odpověď: Ne
- 20 Kontext: Vtom Ulrich zaslechl na chodbě kroky uklízečky.
 1. argument: Porcelán | tam | opatrně | vrátil.
 2. argument: Zato | hodinky | hned | bezostyšně | sbalil.
 Kontext: Rozevřel | noviny | a zatvářil se roztržitě.
 Kontrolní otázka: Vrátil Ulrich alespoň něco?
 Odpověď: Ano

A.3 Výplňové položky 1

- 1 Kontext: Festival úplně zaplavil celé město.
 1. argument: Ulicemi | spěchali | herci | nalíčení na vystoupení.
 2. argument: Muži | v ženských šatech, | klauni, | muzikanti.
 Kontext: Nemít na sobě | aspoň zlaté | šupinaté kalhoty bylo najednou nepatřičné.
 Kontrolní otázka: Je pravda, že bylo možné potkat muže v ženských šatech?
 Odpověď: Ano
- 2 Kontext: Erika na kole projížděla loužemi na rozbité cestě.
 1. argument: Nejdřív | se rozjela | obrovskou | rychlostí.
 2. argument: Vší | silou | stiskla | řídítka.
 Kontext: A zkusila, | jak bude | louže hluboká a jestli přece jen projede.
 Kontrolní otázka: Projížděla Erika loužemi pomalu?
 Odpověď: Ne

- 3 Kontext: Bylo tak příjemné, když přestalo pršet.
1. argument: Najednou | jsme mohli | vyrazit | na cestu.
2. argument: Přece jen | už jsme se | pod přewisem | docela nudili.
Kontext: On slovní fotbal | se také nedá | hrát věčně.
Kontrolní otázka: Je pravda, že skupina vyrazila na cestu ráda?
Odpověď: Ano
- 4 Kontext: Vycházející Měsíc nenadále osvítil sochu Ježíše Krista.
1. argument: Člověk | by se | až | lekl.
2. argument: Přece jen | ve skalách v noci | nečekáte | polonahou postavu.
Kontext: Ale koneckonců, | on je s námi | všude a pořád.
Kontrolní otázka: Vyplývá z textu, že se socha objevila nenadále u rybníka?
Odpověď: Ne
- 5 Kontext: Uličkou v letadle se blížili stewardi s vozíkem s nápoji.
1. argument: Za nimi | podupával | nervózní | pán.
2. argument: Nakonec | se pokusil | protáhnout se | okolo vozíku.
Kontext: Sedl si přitom | na klín | jedné paní a převrhl na sebe víno.
Kontrolní otázka: Počkal pán, až stewardi s vozíkem odjedou?
Odpověď: Ne
- 6 Kontext: Vypravili se s tetou do bazénu.
1. argument: Při převlékání | Péťa | nemohl najít | plavky.
2. argument: Rozhodl se | tedy | plavat | v kraťasech.
Kontext: Na cestu | domů | se omotal ručníkem.
Kontrolní otázka: Plaval Péťa omotaný ručníkem?
Odpověď: Ne
- 7 Kontext: Zoologické zahradě se podařilo odchovat ptáčata vzácného papouška.
1. argument: První | se vylíhla | vloni | v květnu.
2. argument: Ošetřovatelé | zajistili | klid | u hnízda.
Kontext: Návštěvníci | je mohli vidět | až o dva měsíce později.
Kontrolní otázka: Mohli návštěvníci vidět ptáčata hned po vylíhnutí?
Odpověď: Ne

- 8 Kontext: Jeník dostal za úkol najít vzácný prsten.
 1. argument: Prý ho | spolkla | stříbrná | ryba.
 2. argument: I vydal se | pro radu | do hvozdu | ke stařečkovi.
 Kontext: Jel tři dny | a tři noci, | až dojel k chaloupce.
 Kontrolní otázka: Jel Jeník několik dní a nocí?
 Odpověď: Ano
- 9 Kontext: Jana říkala, že žít na severu je pro Čechy obtížné.
 1. argument: Velkou | část roku | bývá | krátký den.
 2. argument: Také | zdejší speciality | jsou pro nás | nezvyklé.
 Kontext: Ryby | a brusinky | se jednoho dne přejedí.
 Kontrolní otázka: Vyplývá z textu, že život na severu je pro Čechy mimořádně příjemný?
 Odpověď: Ne
- 10 Kontext: Odbavení na letišti probíhalo plně automaticky.
 1. argument: Po pásu | přijížděly | nádoby | na příruční zavazadla.
 2. argument: Po vyprázdnění | samy zajížděly | někam | pod zem.
 Kontext: Chybí | už jen automatizace | obouvaní.
 Kontrolní otázka: Bylo odbavení na letišti automatizované?
 Odpověď: Ano
- 11 Kontext: Na univerzitě probíhal cvičný požární poplach.
 1. argument: Rozezněly se | sirény | a zavřely se | protipožární dveře.
 2. argument: Jako by tu | náhle | vzniklo | bludiště s překážkami.
 Kontext: Projít | s čajem | z kuchyňky do kanceláře bylo obtížné.
 Kontrolní otázka: Vyplývá z textu, že na univerzitě došlo k požáru?
 Odpověď: Ne
- 12 Kontext: V chodbě mi stojí rotoped už několik měsíců.
 1. argument: Půjčil mi ho | Běda | na trénink | na závod.
 2. argument: Bez velkého | auta | ho teď nemám | jak vrátit.
 Kontext: Asi budu muset | odjet | přímo na něm.
 Kontrolní otázka: Mám rotoped pouze vypůjčený?
 Odpověď: Ano

- 13 Kontext: V Edinburghu jsme viděli spoustu pouličních umělců.
 1. argument: Nikomu | přitom | nevadil | častý hustý déšť.
 2. argument: Diváci | měli | pestrobarevné deštníky | a přšipláště.
 Kontext: A vystupující | používali | voděodolné mikrofony.
 Kontrolní otázka: Je pravda, že divákům déšť nevadil?
 Odpověď: Ano
- 14 Kontext: S vánočními perníčky zápasí Markéta už přes dvacet let.
 1. argument: Vždycky | jsou | hrozně | tvrdé.
 2. argument: Přidává | k nim | do krabice | jablko.
 Kontext: Ale pokaždé | se ve výsledku | musejí máčet v kakau.
 Kontrolní otázka: Daří se Markétě upéci měkké perníčky?
 Odpověď: Ne
- 15 Kontext: Petra přistupuje ke kuchyni systematicky.
 1. argument: Vytvořila si | promyšlené | vyvážené | jídelníčky.
 2. argument: Podle nich | pořizuje | týdenní | nákupy.
 Kontext: Má také | koření | srovnané podle abecedy, tedy vždy mleté za celým.
 Kontrolní otázka: Pořizuje Petra nákupy každý týden?
 Odpověď: Ano
- 16 Kontext: Doprava se dala pomalu do pohybu.
 1. argument: Za tu dobu | už | všichni | otupěli.
 2. argument: Bylo | opravdu | strašně | pozdě.
 Kontext: Kolona se | monotónně | posouvala v dešti do tmy.
 Kontrolní otázka: Vyplývá z textu, že na vozy svítilo slunce?
 Odpověď: Ne
- 17 Kontext: Hanka s Jitkou se ve škole sešly na zadním schodišti.
 1. argument: Šeptem | probíraly | nové | spolužáky.
 2. argument: Pak | přešly | na učitele | a sourozence.
 Kontext: O patro výš | si zase šeptaly | dvě učitelky o manželích.
 Kontrolní otázka: Šeptaly si Hanka s Jitkou o manželích?
 Odpověď: Ne

- 18 Kontext: Zařizujeme teď dětský pokoj.
 1. argument: Děti | chtějí | trampolínu | a nějaký tunel.
 2. argument: Úplně | nás | jejich nápady | nadchly.
 Kontext: Necháme jim | ještě udělat | horolezeckou stěnu.
 Kontrolní otázka: Přejí si děti do dětského pokoje tunel?
 Odpověď: Ano
- 19 Kontext: Při sjíždění Ohře se převrátili a Jitka si namočila telefon.
 1. argument: Pokoušela se | ho pak | vysušit | i v sáčku s rýží.
 2. argument: Rýže | do sebe | natáhla | všechnu vlhkost.
 Kontext: Obojí z Číny | a jak si to pomáhá, | smály se ostatní posádky.
 Kontrolní otázka: Vysušila Jitka telefon pomocí rýže?
 Odpověď: Ano
- 20 Kontext: Babička neměla ráda přesazování květin.
 1. argument: Nechala si | vždycky v březnu | přivést | hlínu.
 2. argument: Dlouhé týdny | se na ni | se vzdycháním | dívala.
 Kontext: K přesazování | se dostala | tak v srpnu.
 Kontrolní otázka: Mívala babička na jaře květiny čerstvě přesazené?
 Odpověď: Ne
- 21 Kontext: Někdo se třeba zamiluje do někoho jiného kvůli tomu, jakou mluví řečí.
 1. argument: Líbí se | mu | ta intonace | a výslovnost.
 2. argument: Intenzivně se | také | tu řeč | učí.
 Kontext: Pak mu | ten druhý | odejde a zůstane mu jen složitý systém časů.
 Kontrolní otázka: Vyplývá z textu, že se někdo může do někoho zamilovat z jazykových důvodů?
 Odpověď: Ano
- 22 Kontext: Studenti se v září vraceli na koleje.
 1. argument: Přivázeli si | zásoby | od maminek | z domova.
 2. argument: Vyměňovali | mezi sebou | sekanou | a koláče.
 Kontext: Hned | zahájili sezónu | pořádným večírkem.
 Kontrolní otázka: Vypovídá text o tom, co se dělo v září?
 Odpověď: Ano

- 23 Kontext: Marek odvezl auto do servisu.
1. argument: Bylo potřeba | rychle | opravit | klimatizaci.
2. argument: Pouštěla | do auta | horký | vzduch.
Kontext: Na cestu | do Itálie | to bylo nepředstavitelné.
Kontrolní otázka: Potřebovalo auto opravit brzdy?
Odpověď: Ne
- 24 Kontext: Náš papoušek je záludný škodič.
1. argument: Schválně | nám | zahazuje | tužky.
2. argument: Ze žárlivosti | nás | klove | do ucha.
Kontext: Oklovává květiny | a pozoruje nás, | co my na to.
Kontrolní otázka: Je pravda, že náš papoušek žárlí?
Odpověď: Ano
- 25 Kontext: Žofinka milovala plyšové hračky.
1. argument: Nejradši | měla | opičku | Hopinku.
2. argument: Vozila | pro ni | s sebou | opičí krmení.
Kontext: Hopinka | dokázala věci, | vyskočila třeba až do okapu na střeše.
Kontrolní otázka: Vyplývá z textu, že Hopinka byla šikovná?
Odpověď: Ano
- 26 Kontext: Zimní stanování má svoje specifika.
1. argument: Mohou | zamrznout | spoje | u skládacích tyček.
2. argument: V mrazu | také | křehnou | umělohmotné součástky.
Kontext: Je dobré | mít s sebou | lepicí pásku na opravy.
Kontrolní otázka: Liší se zimní stanování od letního?
Odpověď: Ano
- 27 Kontext: Andělka se těšila na seminář jógy.
1. argument: Cvičitel | všechny | oslovoval | „vážení jogíni“.
2. argument: Propagoval | čistou | mysl | a srdce.
Kontext: A před Vánoci | přidával | čistý kvartýr.
Kontrolní otázka: Vyplývá z textu, že cvičitel před Vánoci propagoval čistou mysl, čisté srdce a čistý kvartýr?
Odpověď: Ano

- 28 Kontext: Družstvo se připravuje na výměnu oken.
 1. argument: Podle harmonogramu | bude | do prázdnin | hotová.
 2. argument: Realisticky | počítáme | se zpožděním | několika měsíci.
 Kontext: To | se dá | vydržet.
 Kontrolní otázka: Byla už výměna oken v domě dokončena?
 Odpověď: Ne
- 29 Kontext: Česká města v posledních letech vysloveně prokoukla.
 1. argument: Mívají už | nádherná | opravená | náměstí.
 2. argument: Otevřela se | spousta | zajímavých | kaváren.
 Kontext: Najednou | člověk zvažuje, | proč jezdit do ciziny.
 Kontrolní otázka: Vyplývá z textu, že česká města začala v posledních letech vypadat lépe?
 Odpověď: Ano
- 30 Kontext: Blížila se zkouška ze starořečtiny.
 1. argument: Alena si | s povzdechem | připravila | slovník.
 2. argument: Do učení | se jí | vůbec | nechtělo.
 Kontext: Bude si muset | vymyslet | nějakou pořádnou odměnu.
 Kontrolní otázka: Čekala Alenu zkouška ze starořečtiny?
 Odpověď: Ano
- 31 Kontext: Z okna byla vidět prezidentská standarta na Hradě.
 1. argument: Zuřivě se | třepotala | v prudkém | vichru.
 2. argument: Lidé | v rychlosti | zavírali | okna.
 Kontext: Z jednoho parapetu | se na zem zřítil | truhlík s muškáty.
 Kontrolní otázka: Bylo klidné počasí?
 Odpověď: Ne
- 32 Kontext: Tereza ráda zkoušela nové sporty.
 1. argument: Letos se | například | přihlásila | na kickbox.
 2. argument: Samotný | sport | nebyl | tak těžký.
 Kontext: Těžké bylo | navléci si | druhou rukavici.
 Kontrolní otázka: Vyzkoušela Tereza kickbox?
 Odpověď: Ano

- | | | |
|----|---|---|
| 33 | Kontext:
1. argument:
2. argument:
Kontext:
Kontrolní otázka:
Odpověď: | Ivan měl rád jednoduché technické zlepšováky.
V zimě si omotal pneumatiky na kole prádelní šňůrou.
Kola mu tak na sněhu neprokluzovala.
Pyšnil se, že má nejlevnější sněhové řetězy v republice.
Vyplývá z textu, že Ivan jezdil v zimě na kole?
Ano |
| 34 | Kontext:
1. argument:
2. argument:
Kontext:
Kontrolní otázka:
Odpověď: | Ukázalo se, že koupit stůl do pracovny je velký úkol.
Nejdřív vystěhujeme tábornické vybavení.
Pro něj pořídíme skříň na chodbě.
Na to je ale potřeba vyklidit chodbu.
Je v současné době v pracovně tábornické vybavení?
Ano |
| 35 | Kontext:
1. argument:
2. argument:
Kontext:
Kontrolní otázka:
Odpověď: | Mariana milovala dobré čaje.
V zimě si o ně psala Ježíškovi.
Kupovala je také v čajovnách.
Bohužel jí je kolegové vždycky rychle vypili.
Dostávala Mariana čaje pouze darem?
Ne |
| 36 | Kontext:
1. argument:
2. argument:
Kontext:
Kontrolní otázka:
Odpověď: | Marcela v kostele v neděli nevěděla přesně, co má dělat.
Sledovala při mši ostatní okolo.
Vstávala a sedala vždycky podle nich.
Ale pěkně si zazpívala.
Zpívala Marcela v neděli v kostele?
Ano |
| 37 | Kontext:
1. argument:
2. argument:
Kontext:
Kontrolní otázka:
Odpověď: | Profesor psal na tabuli čímkoliv, co mu přišlo pod ruku.
Fixy určené pro papír nešly smazat.
Uklížečky se s tabulí marně trápily.
Vepsal se zkrátka nesmazatelným písmem do výuky syntaxe.
Přidělával profesor uklížečkám práci?
Ano |

- 38 Kontext: Dědeček usnul a Jaroušek se pustil do průzkumu.
1. argument: Ukousl | mamince | polovinu | rtěnky.
2. argument: Rtěnku | nacpal | do plastové lahve | s vodou.
Kontext: Vodu pak | vylil | a sedl si do ní.
Kontrolní otázka: Poškodil Jaroušek mamčinu rtěnku?
Odpověď: Ano
- 39 Kontext: Vůbec nejhorší nápad byl umýt v myčce svíčky z dortu.
1. argument: Vosk se | v horké | vodě | rozpustil.
2. argument: Dokonale | pokryl | všechno nádobí | i vnitřek myčky.
Kontext: Zato knoty | byly bělostné | a perfektně čisté.
Kontrolní otázka: Zachovaly se svíčky v myčce nepoškozené?
Odpověď: Ne
- 40 Kontext: Silva otevřela kufr od auta.
1. argument: Na nohy jí | vychrstla | vlna | malířské barvy.
2. argument: Ze dveří od kufru | jí barva | kapala | na hlavu.
Kontext: V údivu si sáhla | do vlasů | a umazala si ještě ruce.
Kontrolní otázka: Měla Silva v autě vylitou barvu?
Odpověď: Ano

B

Výchozí sada dokladů pro experiment s časovými vztahy

B.1 Současnost

- 1 Kontext: Dvorek na chalupě najednou ožil.
 1. argument: Zedník | míchal | maltu | v kolečku.
 2. argument: Mezitím | z rozhlasu | vyhrávala | hudba.
 Kontext: Babička | začala | velet partě dělníků.
 Kontrolní otázka: Bylo na dvorku ticho?
 Odpověď: Ne

- 2 Kontext: Tatínek se natáhl do lehátka.
 1. argument: Pepík | lízal | zmrzlinu | na terase.
 2. argument: Mezitím | po pláži | chodil | uklížeč.
 Kontext: Začínala | příjemná | siesta.
 Kontrolní otázka: Vyplývá z textu, že na pláži někdo byl?
 Odpověď: Ano

- 3 Kontext: Pověsila jsem záclony na zahradu.
 1. argument: Děda | vařil | oběd | v kuchyni.
 2. argument: Mezitím | na zahrádku | svítilo | sluníčko.
 Kontext: Než dovaříme, | mohlo by | prádlo uschnout.
 Kontrolní otázka: Vařila oběd babička?
 Odpověď: Ne

- 4 Kontext: Přemýšleli jsme, jak mohlo dojít ke krádeži.
 1. argument: Marie | myla | výlohu | v květinářství.
 2. argument: Mezitím | v prodejně | procházel | zloděj.
 Kontext: Zřejmě vešel | a odešel | zadem.
 Kontrolní otázka: Došlo ke krádeži?
 Odpověď: Ano

- 5 Kontext: Nedělo se nic zvláštního.
1. argument: Květa | sledovala | draka | ve výšce.
2. argument: Mezitím | na nebi | letělo | letadlo.
Kontext: Náhle se ale | u letadla | zablýsklo.
Kontrolní otázka: Podařilo se, aby drak létal?
Odpověď: Ano
- 6 Kontext: Odpoledne se nám konečně vyčáslilo na fotografování.
1. argument: Andrea | pózovala | Petrovi | u zábradlí.
2. argument: Mezitím | do Prahy | přijížděla | reportérka.
Kontext: Večer | poskládáme | fotografie i text dohromady.
Kontrolní otázka: Bylo odpoledne dobré počasí?
Odpověď: Ano
- 7 Kontext: Věrka vystříhovala panenky z papíru.
1. argument: Jiřík | dokončoval | obrázek | v pokojíčku.
2. argument: Mezitím | na střechu | bubnoval | děšť.
Kontext: Děti se | zabavily | a alespoň se nepraly.
Kontrolní otázka: Malovala Věrka obrázek?
Odpověď: Ne
- 8 Kontext: V pokoji hrál gramofon.
1. argument: Babička | pletla | svetr | na gauči.
2. argument: Mezitím | venku | parkoval | strejda.
Kontext: Pěkně sněží, | možná si | uvaříme punč.
Kontrolní otázka: Byl v pokoji slyšet nějaký zvuk?
Odpověď: Ano
- 9 Kontext: Bylo tu další prázdninové poledne.
1. argument: Tereška | skákala | panáka | na chodníku.
2. argument: Mezitím | v bufetu | probíhala | hádka.
Kontext: Opakovalo | se to tak | každý den.
Kontrolní otázka: Byla Tereška v bufetu?
Odpověď: Ne
- 10 Kontext: Od bazénu se blížila mladá dvojice.
1. argument: Dívka | držela | župan | za poutko.
2. argument: Mezitím | pod tobogánem | plakalo | dítě.
Kontext: Dívka | podala župan | mladíkovi a běžela dítěti pomoci.
Kontrolní otázka: Přišla od bazénu více než jedna osoba?
Odpověď: Ano

- 11 Kontext: Holčičky se zastavily před domem.
 1. argument: Johanka | ukazovala | fotografii | na mobilu.
 2. argument: Mezitím | za plotem | štěkal | pes.
 Kontext: Vadilo mu, | že si ho | nevšímají.
 Kontrolní otázka: Vyplyvá z textu, že pes na sebe štěkotem upozorňoval?
 Odpověď: Ano
- 12 Kontext: Chystali jsme se do moře.
 1. argument: Honza | nafukoval | matraci | na pláži.
 2. argument: Mezitím | v chatce | uklízela | posluhovačka.
 Kontext: Čekalo nás | dobrodružství | a nemuseli jsme se o nic starat.
 Kontrolní otázka: Chystal se Honza do města?
 Odpověď: Ne
- 13 Kontext: Dopoledne jsme chystali náradí na tábor.
 1. argument: Vojta | piloval | násadu | v dílně.
 2. argument: Mezitím | na dvoře | pracoval | Vašek.
 Kontext: Pustil se | do opravy | vozíku.
 Kontrolní otázka: Pracoval Vojta na dvorku?
 Odpověď: Ne
- 14 Kontext: Druhý víkend na chalupě probíhal mírumilovně.
 1. argument: Vítek | nosil | dřevo | do stodoly.
 2. argument: Mezitím | v zahradě | zpíval | kos.
 Kontext: Maminka už | viděla, | jaká tu bude pohoda.
 Kontrolní otázka: Vyplyvá z textu, že maminka byla spokojená?
 Odpověď: Ano
- 15 Kontext: Rodiče po obědě odjeli.
 1. argument: Strýc | připravoval | grilování | na večer.
 2. argument: Mezitím | na návsi | cinkal | kolotoč.
 Kontext: Děti | šly | zatím najít kapesné.
 Kontrolní otázka: Připravoval se někdo na večer?
 Odpověď: Ano

- 16 Kontext: Pod Blaníkem si parta udělala příjemnou přestávku na svačinu.
1. argument: Kos | tahal | žížalu | z hlíny.
2. argument: Mezitím | ve stínu | zurčel | potok.
Kontext: Nakonec se | rozhodli | tu zůstat přes noc.
Kontrolní otázka: Líbilo se partě pod Blaníkem?
Odpověď: Ano
- 17 Kontext: Bylo právě poledne.
1. argument: Dodávka | přivázela | oběd | do léčebny.
2. argument: Mezitím | na věži | vyzváněl | zvon.
Kontext: Za chvíli | půjdou | děti ze školy.
Kontrolní otázka: Ozvala se v poledne siréna?
Odpověď: Ne
- 18 Kontext: V domě bylo rušno.
1. argument: Maminka | čistila | mrkev | u stolu.
2. argument: Mezitím | v koupelně | prala | pračka.
Kontext: Kocour | obhlédl | situaci a proklouzl ven.
Kontrolní otázka: Byl v domě klid?
Odpověď: Ne
- 19 Kontext: Po vyučování nastal ve škole klid.
1. argument: Školník | opravoval | světlo | na chodbě.
2. argument: Mezitím | v aule | zněl | zpěv.
Kontext: To se sbor | připravoval | na Vánoce.
Kontrolní otázka: Opravoval školník světlo?
Odpověď: Ano
- 20 Kontext: Byl obyčejný listopadový den.
1. argument: Mlha | přikrývala | nábřeží | pod vrchem.
2. argument: Mezitím | na chodník | padalo | listí.
Kontext: Brzo se | začne | šerit.
Kontrolní otázka: Byl čas Vánoc?
Odpověď: Ne

B.2 Nesoučasnost

- 1 Kontext: Rámus vrcholil, všichni se prali.
 1. argument: Jožka | vykopl | glóbus | na skříň.
 2. argument: Potom | do třídy | přišel | ředitel.
 Kontext: Strnuli jsme | a čekali, | co bude.
 Kontrolní otázka: Je pravda, že Jožka kopl glóbus za skříň?
 Odpověď: Ne

- 2 Kontext: Koně jsem nakonec nechal dole ve vsi.
 1. argument: Pes | přečkal | mráz | u ovcí.
 2. argument: Potom | v horách | nastalo | jaro.
 Kontext: Chystali jsme se | zase | na cestu.
 Kontrolní otázka: Přežil pes zimu?
 Odpověď: Ano

- 3 Kontext: Letos nás nemoci a léky neminuly.
 1. argument: Martina | užívala | antibiotika | na chřipku.
 2. argument: Potom | v nemoci | nastal | obrat.
 Kontext: Ustoupila jí | teplota, | ale dostala silný kašel.
 Kontrolní otázka: Je pravda, že Martina se zcela uzdravila?
 Odpověď: Ne

- 4 Kontext: Po obědě nastal konečně klid.
 1. argument: Velitel | prolistoval | noviny | na pohovce.
 2. argument: Potom | na dveře | zaklepal | nováček.
 Kontext: Podal jen | standardní hlášení | a byl hned pryč.
 Kontrolní otázka: Vyplývá z textu, že si velitel nechal nováčka předvolat?
 Odpověď: Ne

- 5 Kontext: Scházeli jsme se na svatbu postupně.
 1. argument: Maminka | vytřela | podlahu | v kuchyni.
 2. argument: Potom | do dveří | vstoupil | Aleš.
 Kontext: Zul se | do ponožek, | a boty v ruce, začal se ptát po strýci.
 Kontrolní otázka: Vyplývá z textu, že strýc chodil po kuchyni v ponožkách?
 Odpověď: Ne

- 6 Kontext: Táborový večer, jeden z vrcholů roku.
1. argument: Oddíl | nachystá | dřevo | na oheň.
2. argument: Potom | nad obzorem | vyjde | Večernice.
Kontext: Padne | rosa | a do večerního oparu začne stoupat tenká spirála dýmu.
Kontrolní otázka: Je pravda, že oddíl připraví dřevo?
Odpověď: Ano
- 7 Kontext: Celé město bylo v napětí.
1. argument: Meyer | zaslechl | dunění | od náměstí.
2. argument: Potom | do ulice | vjel | tank.
Kontext: Zdálo se, | že neznají | přesně cestu.
Kontrolní otázka: Slyšel Meyer hluk ještě před příjezdem tanku?
Odpověď: Ano
- 8 Kontext: Přes drobné chyby v organizaci dopadl večer nakonec skvěle.
1. argument: Moderátor | přivítal | diváky | v sále.
2. argument: Potom | na scénu | dorazil | sponzor.
Kontext: Dostalo se mu | velkých ovací | a nikde nebyla znát nervozita.
Kontrolní otázka: Vyplývá z textu, že se večer vydařil?
Odpověď: Ano
- 9 Kontext: Strýček rozsvítil svíčky.
1. argument: Děti | zpozorovaly | dárky | pod stromečkem.
2. argument: Potom | ke stromečku | přišla | teta.
Kontext: Zapálila | prskavky | a postupně je na něj navěšela.
Kontrolní otázka: Je pravda, že prskavky hořely dříve než svíčky?
Odpověď: Ne
- 10 Kontext: Rozčilovalo mě to, klasická maloměstská fraška.
1. argument: Starosta | předvedl | úsměv | z tribuny.
2. argument: Potom | na trávníku | proběhl | zápas.
Kontext: Všechno | jako vždy | a všechno k ničemu.
Kontrolní otázka: Vyplývá z textu, že trenér mužstva nejednal poctivě?
Odpověď: Ne

- 11 Kontext: Seděli, límce přitažené co nejtěsněji.
 1. argument: Lenka | rozfoukala | uhlíky | v ohništi.
 2. argument: Potom | v údolí | padla | tma.
 Kontext: Nahoře | se ještě | přelévaly pastelové barvy a zezdola voněl les.
 Kontrolní otázka: Vyplyvá z textu, že ohniště bylo úplně vyhaslé?
 Odpověď: Ne
- 12 Kontext: Po všerejším zápase jsme se scházeli na fotbale úplně skleslí.
 1. argument: Prohra | zkazila | náladu | v mužstvu.
 2. argument: Potom | na hřiště | přijel | trenér.
 Kontext: Začal | zdlouhavý | rozbor nepovedené hry.
 Kontrolní otázka: Pustil se tým do rozboru hry?
 Odpověď: Ano
- 13 Kontext: V létě jsem se s naší rodinou moc neviděl.
 1. argument: Bratr | opravoval | chalupu | na venkově.
 2. argument: Potom | na horách | udeřily | mrazy.
 Kontext: Teplá sezóna | skončila | a začali jsme spolu chodit na squash.
 Kontrolní otázka: Je pravda, že bratr byl v létě na služební cestě?
 Odpověď: Ne
- 14 Kontext: Na ulici duněla od časného rána sbíječka.
 1. argument: Bagr | shrnul | asfalt | na hromadu.
 2. argument: Potom | ve škole | začala | výuka.
 Kontext: Hluk | ustal | a celá čtvrť si oddechla.
 Kontrolní otázka: Je pravda, že bagr vytvářel z asfaltu hromadu?
 Odpověď: Ano
- 15 Kontext: Se zájezdem jsme podle plánu vyrazili ve městě na večeri.
 1. argument: Vedoucí | vyhlásila | program | na zítra.
 2. argument: Potom | v divadle | začalo | představení.
 Kontext: Někdo | šel raději | na procházku, přece jen jsme byli unavení.
 Kontrolní otázka: Je pravda, že divadelní představení se konalo až po večeri?
 Odpověď: Ano

- 16 Kontext: Bylo to takové zapadlé letiště, působilo téměř domácí.
1. argument: Stařík | vytrhával | trávu | na dráze.
2. argument: Potom | na asfaltu | přistálo | letadélko.
Kontext: Mělo | oprýskané nápisy | a vystoupili z něj horolezci, závan jiného světa.
Kontrolní otázka: Je pravda, že na letišti nikdo nebyl?
Odpověď: Ne
- 17 Kontext: S probíranou látkou jsme byli trochu ve skluzu.
1. argument: Třída | nabrala | zpoždění | v matematice.
2. argument: Potom | ve škole | vypukla | chřipka.
Kontext: Všechno | se zhoršilo, | nepostupovali jsme s učivem vůbec kupředu.
Kontrolní otázka: Je pravda, že se ve třídě nedařilo dodržovat časový plán výuky?
Odpověď: Ano
- 18 Kontext: Na výročí konce války dorazila i skupinka amerických veteránů.
1. argument: Primátor | přijal | delegaci | na radnici.
2. argument: Potom | na náměstí | proběhla | oslava.
Kontext: Vyhrávala se | známá polka | a točilo se pivo.
Kontrolní otázka: Pilo se při oslavách pivo?
Odpověď: Ano
- 19 Kontext: Noční obloha byla paradoxně plná barev.
1. argument: Světlo | ozářilo | obzor | na východě.
2. argument: Potom | nad horizontem | vyšel | Měsíc.
Kontext: Byl | krvavě rudý | a zdál se obrovský.
Kontrolní otázka: Byla noční obloha pouze černá?
Odpověď: Ne
- 20 Kontext: Cesta ubíhala v poklidu.
1. argument: Milan | vyfotil | krajinu | za oknem.
2. argument: Potom | do vagónu | vešla | průvodčí.
Kontext: Oznámila, | že vlak | nabral zpoždění, nebylo to nic podstatného.
Kontrolní otázka: Byla Milanova cesta vlakem plná dobrodružství?
Odpověď: Ne

B.3 Výplňové položky 2

- 1 Kontext: U sousedů vyběhl na zahradu pes Dingo.
 1. argument: Kousal | plastovou | láhev, | až praskala.
 2. argument: Vzbudilo to | starou | sousedku | odvedle.
 Kontext: Zamračeně | vyhlédla | z okna.
 Kontrolní otázka: Je pravda, že pes vzbudil sousedku?
 Odpověď: Ano

- 2 Kontext: Letos se urodilo hodně třešní.
 1. argument: Byli jsme | skoro | pořád | na stromě.
 2. argument: Vytvořili jsme si | lanovku | na košíky | a peckový terč.
 Kontext: Ten | jsme umístili | k sousedům za plot.
 Kontrolní otázka: Byl terč na naší zahradě?
 Odpověď: Ne

- 3 Kontext: Máme v malém teráriu afrického šneka.
 1. argument: Je | kupodivu | dost | chytrý.
 2. argument: Při cestě | vlakem | rejdí | po skle.
 Kontext: Poznává, | že se děje | něco zajímavého.
 Kontrolní otázka: Vyplývá z textu, že se šnek v neznámém prostředí bojí?
 Odpověď: Ne

- 4 Kontext: V noci se nad obchodním domem točí neónová bílá labuť.
 1. argument: Ale | ve dne | není vidět | ani její obrys.
 2. argument: Jako by | nad ránem | někam | odlétala.
 Kontext: Pražská | bílá labuť | je noční pták.
 Kontrolní otázka: Je labuť v noci vidět?
 Odpověď: Ano

- 5 Kontext: Naší Sašence bylo letos jedenáct.
 1. argument: Pořád | jenom | leží | v knihách.
 2. argument: Teď si | půjčila | 451 stupňů | Fahrenheita.
 Kontext: Ta se bude | divit, | že někdo knížky pálí.
 Kontrolní otázka: Vyplývá z textu, že Sašenka ráda čte?
 Odpověď: Ano

- 6 Kontext: Zuzana má radost, že začala o přestávce v práci chodit běhat.
1. argument: Mnohem lépe | jí to | pak | myslí.
2. argument: Jenom | to prádlo | není potom | kde sušit.
Kontext: Nechává ho | proto | přehozené přes počítač.
Kontrolní otázka: Suší Zuzana prádlo v práci?
Odpověď: Ano
- 7 Kontext: Drahoš rád chodil do práce v sobotu.
1. argument: Byl | tam | úžasný | mír.
2. argument: Ve společné | kanceláři | si dával | nohy na stůl.
Kontext: Na dlouhých chodbách | zkoušel | hvězdy jako kdysi.
Kontrolní otázka: Bylo v sobotu v práci u Drahoše rušno?
Odpověď: Ne
- 8 Kontext: Na kříži na kostele seděl obrovský holub.
1. argument: Čistil | si | zobákem | peří.
2. argument: Byla | to | mírumilovná | chvilka.
Kontext: Jako by se | vprostřed dne | zastavil čas.
Kontrolní otázka: Byl holub na kostele velký?
Odpověď: Ano
- 9 Kontext: Ota si vybírá jídlo v restauraci podle množství alergenů.
1. argument: Více | alergenů | prý znamená | lepší chuť.
2. argument: Jiné | sloupečky | v jídelníčku | ani nechte.
Kontext: Údajně | dosud nikdy | nelitoval.
Kontrolní otázka: Vyhýbá se Ota alergenům v jídle?
Odpověď: Ne
- 10 Kontext: V deset hodin troubí na věži trubači znělku.
1. argument: V našem | týmu | se začíná | svačit.
2. argument: Výjimkou | je | jen | Jana.
Kontext: Ta | teprve | přichází do práce.
Kontrolní otázka: Je pravda, že Jana přichází do práce později než ostatní?
Odpověď: Ano

- 11 Kontext: Na chodbě stojí elegantní černé kolo.
 1. argument: Má | rám | z karbonových | trubek.
 2. argument: Jezdí | na něm | náš | nový kolega.
 Kontext: Říkáme mu | pan | Karbon.
 Kontrolní otázka: Je kolo na chodbě černé?
 Odpověď: Ano
- 12 Kontext: Zapomněli jsme, že má Jirka dovolenou.
 1. argument: V kanceláři se | konala | oslava | jeho narozenin.
 2. argument: Eliška | nakrájela | meloun | pro kolegy.
 Kontext: Poslali jsme | Jirkovi | aspoň fotografii.
 Kontrolní otázka: Účastnil se Jirka oslavy svých narozenin?
 Odpověď: Ne
- 13 Kontext: Máša jezdí na prázdniny do Ruska.
 1. argument: Často se | jí | stávají | zajímavé příhody.
 2. argument: Minule | byla | na kajacích | v Karélii.
 Kontext: Omylem | zapálila | nedopalkem cigarety dřevěný mostek nad řekou.
 Kontrolní otázka: Vyplývá z textu, že Máša zapálila v Karélii domek?
 Odpověď: Ne
- 14 Kontext: V poslední době je možné si ve městech půjčovat kola.
 1. argument: Často | mívají | různé směšné | názvy.
 2. argument: Pěkné | je | například | Metelesku Blesku.
 Kontext: Zato jezdit | na Titaniku | by se člověk bál.
 Kontrolní otázka: Vyplývá z textu, že je možné si půjčovat kola s různými názvy?
 Odpověď: Ano
- 15 Kontext: Pavel vybíhal schody do čtvrtého patra v naučeném rytmu.
 1. argument: Na dvanáct | kroků | vyběhl | jedno schodiště.
 2. argument: V mezipatře | se na jeden krok | otočil | okolo zábradlí.
 Kontext: Tím se mu | vystřídaly | nohy.
 Kontrolní otázka: Je pravda, že Pavel do schodů běhal?
 Odpověď: Ano

- 16 Kontext: Česání švestek je příjemná věc.
 1. argument: Švestky | jsou | nádherně | modré.
 2. argument: Taky | to | rychle | přibývá.
 Kontext: Je to | jedna | z nejkrásnějších chvil roku.
 Kontrolní otázka: Přibývají švestky při česání rychle?
 Odpověď: Ano
- 17 Kontext: Markéta jezdila do kostela na kole, ale nebylo to jednoduché.
 1. argument: Chodit do kostela | v cyklistickém | jí bylo | hloupé.
 2. argument: Převlékala se | tedy | za kostelem | do nedělního.
 Kontext: To zas | musela vybírat | takové, aby se cestou v brašně moc nepomačkalo.
 Kontrolní otázka: Chodila Markéta do kostela v cyklistickém?
 Odpověď: Ne
- 18 Kontext: Bětuška strávila v knihkupectví čtyři hodiny.
 1. argument: Nakonec | ji | odtamtud | vyhodili.
 2. argument: Četla si | novou | Hru | o trůny.
 Kontext: Má ji | rozečtenou | doma a v knihkupectví se jí nedařilo odtrhnout se od děje.
 Kontrolní otázka: Četla Bětuška v knihkupectví tutéž knihu, kterou má doma?
 Odpověď: Ano
- 19 Kontext: Na náměstí vystupoval silák z Třince.
 1. argument: Nechal | na sobě | štípat | špalek.
 2. argument: Byla | to | hrůzná | podívaná.
 Kontext: Sekyra | se mohla | snadno smeknout.
 Kontrolní otázka: Vyplývá z textu, že se silák nechal seknout?
 Odpověď: Ne
- 20 Kontext: Cukrář natočil Evičce cukrovou vatu na špejli.
 1. argument: Zářivě se | přitom smál | na její | maminku.
 2. argument: Evička | za obrovskou | vatou | nebyla ani vidět.
 Kontext: Všichni tři | měli | v letním dni velikou radost.
 Kontrolní otázka: Dostala Evička cukrovou vatu v létě?
 Odpověď: Ano

- 21 Kontext: Děti si na zahradě nacvičily představení na trampolíně.
 1. argument: Martinka | byla | přísná | učitelka skoků.
 2. argument: Její | žačka | nakonec | jako omdlela.
 Kontext: Říkaly | tomu | nervové zřícení.
 Kontrolní otázka: Předváděly děti představení na trampolíně?
 Odpověď: Ano
- 22 Kontext: Být českým turistou je v něčem těžké.
 1. argument: Máme | totiž | vynikající | turistické značení.
 2. argument: Problém | pak | míváme | v cizině.
 Kontext: Tam | býváme | často dezorientováni.
 Kontrolní otázka: Jsou čeští turisté zvyklí na návodné turistické značení?
 Odpověď: Ano
- 23 Kontext: Svět pobaltské mytologie je fantastický.
 1. argument: Nad lesy | létají | hrozné | čarodějnice.
 2. argument: Používají | k tomu | obvykle | stoupu.
 Kontext: Nezřídka | mívají | místo nohy kládu.
 Kontrolní otázka: Je pravda, že pobaltské čarodějnice mívají místo nohy kládu?
 Odpověď: Ano
- 24 Kontext: Na Hlavním nádraží stojí černý klavír.
 1. argument: Vždycky | na něj | někdo | hraje.
 2. argument: V tom | velikém ruchu | se zastaví | čas.
 Kontext: Bývá | to | milé zpomalení.
 Kontrolní otázka: Je na Hlavním nádraží umístěný klavír?
 Odpověď: Ano
- 25 Kontext: Sofinka se po návratu ze školky zlobila.
 1. argument: Školní | rok | trvá | deset měsíců.
 2. argument: Je | to | velká | nespravedlnost.
 Kontext: Prázdniny | prý trvají | jenom čtyři měsíce.
 Kontrolní otázka: Byla Sofinka s délkou prázdnin spokojená?
 Odpověď: Ne

- 26 Kontext: Karel je pořád unavený.
1. argument: Ve firmě | probíhá | náročná | reorganizace.
2. argument: Také | doma | ho čeká | spousta práce.
Kontext: Asi si | půjde | zastřílet z luku.
Kontrolní otázka: Odpočine si Karel doma?
Odpověď: Ne
- 27 Kontext: Hrát šachy je velice náročné.
1. argument: Velmistři | často | končí hru | úplně vyčerpání.
2. argument: Musejí | dbát | na duševní | hygienu.
Kontext: Věnují se | například | meditaci.
Kontrolní otázka: Je pravda, že hra v šachy může vést k vyčerpání?
Odpověď: Ano
- 28 Kontext: V protějším okně někdo sušil kalhoty.
1. argument: Nebezpečně | se nakláněly | nad rušnou | ulici.
2. argument: V Petrovi se | úplně | zastavilo | srdce.
Kontext: Jako by tam | někdo | měl už už vypadnout.
Kontrolní otázka: Sušil Petr kalhoty v okně?
Odpověď: Ne
- 29 Kontext: Létat balónem musí být krásné.
1. argument: Seshora | je vidět | na všechny | strany.
2. argument: Balón | letí | poměrně | pomalu.
Kontext: Je to | tempo | časů Julese Verna.
Kontrolní otázka: Vyplývá z textu, že balón létá rychle?
Odpověď: Ne
- 30 Kontext: Honza vybíral jméno pro syna.
1. argument: Jedno | mu připadalo | hezčí | než druhé.
2. argument: V podchodu | koupil | manželce | hyacint.
Kontext: Koneckonců, | i to | je jméno.
Kontrolní otázka: Je pravda, že jméno pro syna vybírala manželka?
Odpověď: Ne
- 31 Kontext: Pan učitel ořezával odpoledne pastelky.
1. argument: Měl | dnes | dozor | v družině.
2. argument: Začínalo | ho | bolet | zápěstí.
Kontext: Bude muset | koupit | elektrické ořezávatko.
Kontrolní otázka: Má pan učitel elektrické ořezávatko?
Odpověď: Ne

- 32 Kontext: Zdeněk se těšil do letního kina.
 1. argument: Pozve | tam | dávnou | kamarádku.
 2. argument: A vezme | s sebou | pro ni | bundu.
 Kontext: Ona | si ji totiž | vždycky zapomene a pak mrzne.
 Kontrolní otázka: Vyplyvá z textu, že v letním kině může být chladno?
 Odpověď: Ano
- 33 Kontext: Na výletě jsme našli spoustu hub.
 1. argument: Neměli jsme | je už | do čeho | dávat.
 2. argument: Naplnili jsme | i bundy | a | kalhoty.
 Kontext: Bylo nám | líto, | kolik jich tam necháváme.
 Kontrolní otázka: Dávali jsme houby i do oblečení?
 Odpověď: Ano
- 34 Kontext: Uprostřed mostu stála bělostná labuť.
 1. argument: Kousek dál | se připravoval | městský policista | v černém.
 2. argument: Pomalu si | natáhl | pevné | rukavice.
 Kontext: Došel | k labuti | a řekl jí: pojd', holka.
 Kontrolní otázka: Měl policista modrý stejnokroj?
 Odpověď: Ne
- 35 Kontext: Hladina Vltavy se oslnivě leskla.
 1. argument: Po proudu | přijížděla | postava | na paddleboardu.
 2. argument: Najednou | přišly | vlny | od parníku.
 Kontext: Postava | zavrávala | a spadla do vody.
 Kontrolní otázka: Udržela se postava na paddleboardu?
 Odpověď: Ne
- 36 Kontext: Jezdící schody monotónně hučely.
 1. argument: Zajížděly | pod zuby | krytu | na konci.
 2. argument: Tam dál | se | člověk | nedostane.
 Kontext: Možná je | na druhé straně | nějaký zajímavý svět.
 Kontrolní otázka: Vyplyvá z textu, že na opačnou stranu jezdících schodů nemůže člověk zajet?
 Odpověď: Ano

- 37 Kontext: Svého času byly jediné dveře s automatickým otevíráním v Praze na letišti.
1. argument: Děti | tenkrát | nemohly uvěřit | svým očím.
2. argument: Vypadalo | to | jako | zázrak.
Kontext: Zkoušely | dveře | znovu a znovu.
Kontrolní otázka: Byly v Praze na letišti dveře s automatickým otevíráním?
Odpověď: Ano
- 38 Kontext: Jízda pater nosterem bývá napínavá.
1. argument: Je | tu | skryté | nebezpečí.
2. argument: Možná se | nepodaří | včas | vystoupit.
Kontext: A kdo ví, | co přesně | se děje v místě obratu.
Kontrolní otázka: Vyplývá z textu, že jízda pater nosterem je naprosto bezpečná?
Odpověď: Ne
- 39 Kontext: Filip opatrně najel autem na stojan.
1. argument: Technik | na něj | souhlasně | mávl.
2. argument: V dílně | bylo | pekelně | dusno a kluzko.
Kontext: A technik | vypadal | jako čert.
Kontrolní otázka: Bylo v dílně dusno?
Odpověď: Ano
- 40 Kontext: Na sběr divizny je potřeba dobré počasí.
1. argument: Ve vlhku | se květy | totiž | zavírají.
2. argument: Teď však | máme | slunečno | a bezvětrí.
Kontext: Dneska | by to | šlo.
Kontrolní otázka: Je možné sbírat květy divizny za dobrého počasí?
Odpověď: Ano