

**Univerzita Karlova v Praze**

**Filozofická fakulta**

Ústav českého jazyka a teorie komunikace

Filologie – Český jazyk

**M a g d a l é n a R y s o v á**

**Diskurzní konektory v češtině  
(Od centra k periferii)**

**Discourse Connectives in Czech  
(From Centre to Periphery)**

Teze disertační práce

Vedoucí práce: prof. PhDr. Eva Hajičová, DrSc.

2015

# Abstrakt

V práci se zabýváme popisem a analýzou diskurzních (tj. textových) konektorů v češtině v širším smyslu, tedy tím, jakými jazykovými prostředky je možné vyjadřovat v textu sémanticko-pragmatické vztahy. Výzkum přitom neomezujeme na předem stanovenou skupinu výrazů (danou například příslušností k určitým slovním druhům, jako jsou spojky či částice strukturující text), ale snažíme se nalézt a obecně popsat všechny konektivní prostředky v češtině, které mají schopnost spojovat jednotlivé úseky textu v jeden koherentní celek. Zaměřujeme se především na méně probádané víceslovné konektivní struktury typu *to je důvod, proč; kvůli těmto skutečnostem; z těchto důvodů* atd., pro které užíváme označení **sekundární konektory** (za primární konektory považujeme především konektivní sysnématika typu *však, nebo, a, ale, proto* apod.).

## Klíčová slova

diskurz, diskurzní konektory, diskurzní vztahy, Pražský závislostní korpus, text

## 1 Cíle práce, jazykový materiál

Diskurzní vztahy mohou být v textu vyjadřovány různými způsoby. Obecně můžeme vyčlenit dva základní typy diskurzních vztahů – implicitní (nevyjádřené konkrétními jazykovými prostředky, ale pouze vyplývající z obsahu daných výpovědí či vět, srov. příklad 1) a explicitní (vyjádřené konkrétními jazykovými prostředky, srov. příklady 2–13). Škála těchto jazykových prostředků je přitom značně široká, viz následující příklady.

- (1) *Petr je nemocný a nemůže jet. Výlet se ruší.*
- (2) *Petr je nemocný a nemůže jet. Proto se výlet ruší.*
- (3) *Petr je nemocný a nemůže jet. Výlet se tedy ruší.*
- (4) *Petr je nemocný a nemůže jet. Výlet se tak ruší.*
- (5) *Petr je nemocný a nemůže jet. Výlet se tím ruší.*
- (6) *Petr je nemocný a nemůže jet. Výlet se tímto ruší.*
- (7) *Petr je nemocný a nemůže jet. Na základě toho se výlet ruší.*
- (8) *Petr je nemocný a nemůže jet. Z tohoto důvodu se výlet ruší.*
- (9) *Petr je nemocný a nemůže jet. Kvůli tomu se výlet ruší.*
- (10) *Petr je nemocný a nemůže jet. Kvůli této skutečnosti se výlet ruší.*
- (11) *Petr je nemocný a nemůže jet. Kvůli jeho nemoci se výlet ruší.*
- (12) *Petr je nemocný a nemůže jet. Kvůli němu se výlet ruší.*
- (13) *Petr je nemocný a nemůže jet. Důsledkem je zrušení výletu.*

Ve všech uvedených příkladech je důsledkový vztah vyjádřený pokaždé jiným způsobem, pomocí různých výrazových prostředků: *proto, tedy, tak, tím, tímto, na základě toho, z tohoto důvodu, kvůli tomu, kvůli této skutečnosti, kvůli jeho nemoci, kvůli němu, důsledkem je*.

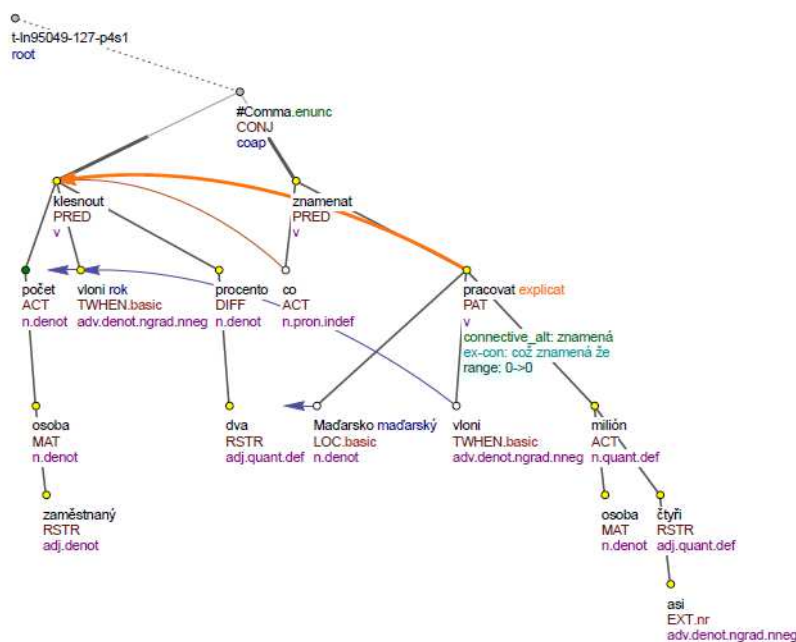
V předkládané disertační práci se zabýváme výrazovými prostředky, které v textu vyjadřují diskurzní vztah explicitně. Ne u všech je ale stejná míra explicitnosti vyjádření, resp. míra jeho konkrétnosti (srov. *proto* vs. *kvůli jeho nemoci*). Explicitně vyjádřené výrazové prostředky signalizující diskurzní vztahy tedy chápeme jako škálu s různými stupni abstraktnosti/obecnosti. V práci se zabýváme tím, které typy výrazů ve škále můžeme označovat jako diskurzní konektory a proč (např. zda jsou diskurzním konektorem výrazové prostředky *proto* i *kvůli jeho nemoci*), na základě jakých kritérií či principů je můžeme třídit a

klasifikovat a v jaké míře se jednotlivé typy těchto výrazových prostředků podílejí na celkové koherenci textu.

Cílem práce je tedy přispět k popisu výstavby textu v češtině, a to především k popisu jazykových prostředků, které mají schopnost vyjadřovat diskurzní vztahy v textu, tedy k popisu diskurzních konektorů. Práce se přitom zaměřuje na výrazy morfologicky, lexikálně i syntakticky neomezené (typu *důsledkem je, příčinou bylo, to je v rozporu s tím* atd., dané výrazy nazýváme sekundární konektory), které v mnohých pojetích (srov. např. Stede a Neumann, 2014; Halliday a Hasanová, 1976; Hrbáček, 1994 atd.) nejsou mezi diskurzní konektory řazeny.

Dílním cílem práce je srovnat různé teorie a přístupy k diskurzu a k diskurzním konektorům v českém i světovém měřítku a přispět novými poznatky (na základě analýzy rozsáhlých korpusových dat) k obecné definici těchto výrazů.

Výzkum byl uskutečněn na ručně anotovaných datech Pražského závislostního korpusu (PDT 3.0) (tj. téměř na 50 000 vět), který v současné době obsahuje i (dosud nepublikovanou) rozšířenou anotaci diskurzních vztahů, tedy vztahů vyjádřených právě sekundárními konektory, tj. výrazy typu *podmínkou je; to je důvod, proč* atd. Dílním cílem je tuto anotaci (provedenou pro účely naší práce – viz obrázek 1) krátce představit a srovnat její výsledky s výsledky dřívější diskurzní anotace primárních konektorů (srov. Poláková a kol., 2012a), tj. především jednoslovných konektorů typu *ale, proto, zatímco* apod. V práci se tedy pokusíme poukázat na možnosti, které zpracování diskurzních vztahů ve velkém měřítku nabízí, uvést teoretické otázky, které mohou na základě anotací vyvstat, a představit způsoby, kterými je možné je řešit.



**Obrázek 1:** Ukázka závislostního stromu z Pražského závislostního korpusu (PDT) s anotací diskurzního vztahu vyjádřeného sekundárním konektorem. Strom znázorňuje souvětí: *Počet zaměstnaných osob klesl loni o dvě procenta, což znamená, že v Maďarsku loni pracovaly asi čtyři milióny osob.*

V daném příkladě je mezi diskurzními argumenty (argumentem 1 je *počet zaměstnaných osob klesl loni o dvě procenta*; argumentem 2 je *v Maďarsku loni pracovaly asi čtyři milióny osob*) diskurzní vztah explikace (vysvětlení) – tj. obsah druhého argumentu vysvětluje obsah prvního argumentu. Diskurzní vztah je explicitně vyjádřen sekundárním konektorem *což znamená, že*. V anotaci je vztah zachycen oranžovou šipkou (ostatní šipky na obrázku zachycují vztahy koreference).

## 2 Postup práce

1. V první části práce se zabýváme teoretickým srovnáním českých i zahraničních přístupů ke zpracování diskurzu. Stežejní otázkou je vymezení a definice diskurzních konektorů.

V práci vycházíme z českých i zahraničních studií, které se ve vymezení diskurzních konektorů značně liší. Někteří autoři (srov. např. Halliday a Hasanová, 1976; Martin, 1992; Knott, 1996; Stede a Neumann, 2014) definují diskurzní konektory v užším smyslu – jako uzavřenou skupinu výrazů, které patří mezi neohebné slovní druhy, především spojky, příslovce a částice (např. *protože, i, nebo, také, poté* atd.), jiní (srov. např. Hoffmannová, 1983, 1984; Prasadová a kol., 2010; Asrová a Dembergová, 2013; Rysová 2013b, 2015) vymezují konektory v širším smyslu, tj. jako výrazy morfologicky, lexikálně i syntakticky neomezené – zdůrazňují tedy především funkci daných výrazů v textu, ne jejich slovnědruhovou příslušnost.

2. V práci dále představujeme, jakým způsobem jsme zpracovávali sekundární konektory v Pražském závislostním korpusu. Na základě studia české i zahraniční literatury (především jednotlivých pojetí diskurzních konektorů) jsme provedli nejprve anotaci vzorku dat PDT, na kterém jsme sledovali chování sekundárních konektorů v reálných textech. Na základě zkušeností s autentickými texty jsme stanovili obecné anotační principy, podle kterých jsme sekundární konektory anotovali v celém PDT (tj. na přibližně 50 000 vět; anotace prováděli tři anotátoři). Následně proběhla kontrola anotací a měření anotátorské shody.

3. Na získaných datech PDT jsme provedli vlastní výzkum diskurzních konektorů v češtině, zaměřený na konektory sekundární.

Předkládaná disertační práce by tak měla na základě podrobné analýzy současného jazykového materiálu, a to psaného i mluveného (kromě PDT jsme využívali i Český národní korpus a korpus mluvené češtiny DIALOG), přinést nové poznatky o možnostech vyjadřování diskurzních vztahů a celkově přispět k prohloubení popisu výstavby textu v češtině.

## 3 Závěry

### 3.1 Definice diskurzních argumentů a konektorů

Na základě analýzy rozsáhlého jazykového materiálu jsme se pokusili vymezit pojem diskurzní argumenty a konektory. Diskurzní argumenty (tj. úseky textu, mezi nimiž je diskurzní vztah) vymezujeme především sémanticky – diskurzní argument je abstraktní objekt, který může vyjadřovat stavy, události, situace atd. Argumenty bývají v textu realizovány především jednotlivými větami (či posloupností několika vět) nebo nominalizovanými frázemi. Z hlediska syntaxe tak mají nejčastěji podobu vět (jejichž základem je sloveso v určitém tvaru, a proto je nazýváme slovesnými argumenty) a jmenných frází (jejichž základem bývá nejčastěji slovesné podstatné jméno; tyto argumenty nazýváme nominálními).

V textech mohou být diskurzní argumenty realizovány například větou, souvětím, samostatnou vedlejší větou, posloupností několika vět i celým odstavcem.

V některých případech mohou být diskurzními argumenty i objekty vyjádřené jinak než slovně, tj. situace, k nimž mohou odkazovat konektory (např. v počátečních výpovědích typu *Tak jsem doma.*).

Diskurzí argumenty, tj. objekty, mezi kterými existují textové vztahy, tedy mohou být:

i. **vyjádřené slovně;**

- **slovesně** (realizované jednoduchou větou, souvětím, samostatnou vedlejší větou, posloupností několika vět i celým odstavcem či textem);
- **jmenně** (realizované jmennými frázemi);

ii. **situační<sup>1</sup>.**

Diskurzí argumenty mohou stát v textu bezprostředně vedle sebe nebo od sebe mohou být „odděleny“ jiným textem, který není součástí ani jednoho z argumentů.

Diskurzí argumenty jsou propojeny vztahy, které mohou být vyjádřené nebo nevyjádřené. Textové vztahy mohou být vyjádřeny diskurzími konektory (primárními a sekundárními) či dalšími spojovacími prostředky. Hranici mezi diskurzími konektory a ostatními konektivními prostředky je možné stanovit na základě toho, zda příslušné výrazy fungují v textu jako indikátory diskurzí vztahů (ve velké míře) univerzálně či příležitostně (tj. v závislosti na kontextu) – srov. různé varianty předložkových frází obsahující sekundární předložku *díky* a anaforický výraz odkazující k předchozímu diskurzímu argumentu: *díky tomu, díky jeho chorobě, díky této skulince, díky všem těmto opatřením, díky pomoci příbuzných a přátel z USA, díky této preciznosti* aj. či různé realizace konektivních struktur obsahující slovo *důvod*: *jako důvod uvedla, jako důvod odstoupení komuniké ČSOB uvádí, důvodem je, důvodem poklesu je, důvodem tohoto návrhu je* (uvedené příklady pocházejí z PDT).

**Daná spojení se podle našeho názoru liší tím, zda jsou kontextově závislá** (např. *díky jeho chorobě*) **nebo nezávislá** (*díky tomu*) – jinými slovy, zda je můžeme použít (do značné míry) **univerzálně pro vyjádření daného sémantického typu diskurzího vztahu** (v tomto případě vztahu příčinně-důsledkového). **Z tohoto hlediska mohou být spojení *díky jeho chorobě* či *důvodem poklesu je* použita pouze v omezeném množství kontextů** (tj. v prvním případě v textech o chorobách, ve druhém v textech o poklesu), **zatímco *díky tomu* či *důvodem je* nejsou kontextem limitována**. Ve spojení *důvodem je* není anaforický výraz na povrchu vyjádřen (oproti *důvodem tohoto návrhu* aj.), a může proto být použit k vyjádření příčinně-důsledkového vztahu (téměř) nezávisle na kontextu<sup>2</sup>. Obdobně ve spojení *díky tomu* si *tomu* jako deiktický výraz vyhledá své sémantické vztahy po zasazení do jakéhokoli kontextu. Slova jako *pokles, návrh, choroba, skulinka, odstoupení* aj. tuto schopnost nemají, tj. nemohou svůj význam zcela naplnit až po dosazení do konkrétního textu. Tato skutečnost je podle našeho názoru pro diskurzí konektory zásadní. Pro závislost či nezávislost konektivních spojení na kontextu budeme nadále používat označení **princip univerzálnosti**.

**Princip univerzálnosti tedy hodnotí konektivní struktury podle míry jejich vhodnosti použití pro různé kontexty**, tj. zda fungují v textu jako ukazatele daného diskurzího vztahu spíše univerzálně nebo příležitostně. Jinými slovy tento princip určuje – pokud máme několik odlišných kontextů, ve kterých se vyskytuje určitý sémantický typ diskurzího vztahu – do kolika těchto kontextů se daná konektivní struktura (se schopností vyjádřit tento konkrétní typ vztahu) hodí, tj. do jaké míry je „univerzální“. Z tohoto hlediska chápeme například spojení *díky tomu* jako obecnější (obecněji v textech využitelné) než *díky*

<sup>1</sup> Situačním argumentem může být například hluk ve třídě. Slovní reakce *Tak ticho!* obsahuje druhý (slovně vyjádřený) diskurzí argument (*ticho*) a konektor (*tak*), který oba argumenty spojuje.

<sup>2</sup> V úvahu nyní nebereme vhodnost daných konektivních prostředků v různých funkčních stylech, jde nám pouze o odlišení spojení typu *důvodem je* a *důvodem poklesu je* apod. ve funkci diskurzí konektorů.

*jeho chorobě*. Univerzální platnost<sup>3</sup> (či vhodnost použití ve více kontextech) chápeme jako klíčovou vlastnost diskurzních konektorů. Proto jsme se rozhodli stanovit hranici těchto výrazů právě zde. Termín **diskurzní konektor** tedy budeme nadále používat pro **výrazy či spojení, které slouží jako univerzální indikátory příslušných diskurzních vztahů (nezávisle na jejich slovnědruhové příslušnosti), tj. např. výrazy jako ale; nebo; naopak; to je důvod, proč; kvůli tomu; příčinou je; to znamená aj.** Ostatní spojení, která fungují jako konektivní prostředky v závislosti na kontextu (tj. neuniverzálně, pouze příležitostně) budeme nazývat **volné konektivní fráze** (srov. např. *díky jeho chorobě; kvůli tomuto zvýšení; díky této skulině; díky této preciznosti; důvodem poklesu je; důvodem tohoto návrhu je*).

### 3.1.1 Charakteristika primárních a sekundárních konektorů

Vzhledem k tomu, že daná kritéria pro diskurzní konektory splňují značně různorodé jazykové prostředky (srov. např. jednoslovné výrazy *a, také, totiž* a víceslovná spojení typu *to je důvod, proč; příčinou bylo* atd.), vyčleňujeme v rámci kategorie diskurzních konektorů dvě skupiny, pro které zavádíme označení primární a sekundární konektory.

Jako **primární konektory** označujeme především symsémantika s primární spojovací funkcí, tj. výrazy patřící mezi spojky (např. *a, ale, však, i*) a částice strukturující text nebo jeho části (dělení slovních druhů přebíráme podle *Mluvnice češtiny 2* (1986) – srov. např. *totiž, přitom, navíc, pak, potom, nakonec, například*), některá vztažná zájmena (*což*) a vztažná zájmenná příslovce (*proto, přičemž*). Za primární konektory současně považujeme také některé grafické značky (dvojtečku a pomlčku), číselné a písmenné výrazy sloužící ke strukturaci textu, nejčastěji k vyjádření výčtu (např. *1., 2., 3.; a), b), c); i), ii), iii)*) a větnou negaci, pokud přispívá k vyjádření diskurzního vztahu (*Nebudu se jenom dívat na televizi, ale půjdu i na chvíli ven*).

U zmíněných příslovcí a zájmen se spojovací funkcí dochází často k přechodu ve spojky či strukturující částice (srov. např. *potom, pak, což*), dané výrazy tedy ztrácejí větněčlenskou platnost. Proces přechodu v jiný slovní druh ovšem nemusí být plně dokončen, při určování slovních druhů u zmíněných konektivních prostředků tak může docházet k neshodám.

Vzhledem ke zmíněnému překrývání v rámci slovních druhů (především mezi spojkami, částicemi a příslovci) a ke skutečnosti, že zejména u mnohých příslovcí se spojovací funkcí dochází k přechodu k jinému slovnímu druhu, chápeme primární konektory především jako symsémantika, tj. výrazy bez větněčlenské platnosti.

Primární konektory jsou tedy většinou jednoslovné (některé ovšem mohou tvořit korelativní páry), gramatikalizované a lexikalizované výrazy, ve většině případů nedovolující modifikaci (nelze například říci *\*obecně a, \*jednoduše ale*, výjimku tvoří výrazy typu *jednoduše/obecně/hlavně proto, že*). Primární konektory bývají v různých studiích užívány jako typičtí zástupci diskurzních konektorů. Frekvence primárních konektorů v textech je mnohem vyšší než frekvence konektorů sekundárních (v Pražském závislostním korpusu vyjadřují primární konektory 94,6 % diskurzních vztahů vyjádřených explicitně). Z tohoto důvodu si můžeme **primární konektory** představit **jako centrum množiny diskurzních konektorů** a **sekundární konektory jako její periferii**.

**Sekundární konektory** jsou univerzální konektivní struktury obsahující autosémantickou složku (jejíž lexikální význam ale bývá oslaben). Jedná se o struktury, které zatím neprošly gramatikalizací (tj. v plném rozsahu – proces gramatikalizace u nich ovšem probíhá; u různých typů sekundárních konektorů je zároveň proces gramatikalizace v různém stadiu). Sekundární konektory jsou na rozdíl od primárních většinou víceslovné a v některých

<sup>3</sup> Univerzálnost diskurzních konektorů ale zpravidla nikdy nedosahuje úplných 100 %.

případech bývá obtížné určit jejich rozsah (při určování rozsahu sekundárních konektorů, tj. např. zda je sekundárním konektorem spojení *podmínkou je* nebo *podmínkou je, že*, bereme jako obecné kritérium možnost náhrady primárním konektorem<sup>4</sup>).

Jednotlivá slova, která dané spojení tvoří, nemívají spojovací funkci ze své podstaty (jako spojky či strukturující částice), ale platnost konektoru nabývá až celá struktura. Ve větě mají sekundární konektory větněčlenskou platnost (některé mohou fungovat i jako větné modifikátory). Sekundární konektory ve formě samostatných vět do syntaxe diskurzních argumentů v tomto smyslu nezasahují. Sekundární konektory jsou zpravidla modifikovatelné (srov. *hlavním/jediným/prvním/významným/rozumným... důvodem je, jasně to odůvodnil* apod.). Všechny obsahují implicitní anaforickou referenci k celému prvnímu argumentu, tato reference se podle typu sekundárního konektoru může objevit i v povrchové realizaci konektoru. Sekundární konektory tak můžeme rozdělit do skupin podle toho, zda anaforickou referenci na povrchu vyjádřit mohou (*příkladem je* vs. *příkladem toho je*), nemohou (*stručně řečeno*) nebo musí (*kvůli tomu*).

Jedinečnost sekundárních konektorů spočívá v tom, že mohou (na rozdíl od konektorů primárních) stát syntakticky výše než druhý diskurzí argument, tj. formálně mohou mít povahu hlavní věty (srov. příklady typu *To je důvod, proč tam nepůjdu.*). Některé sekundární konektory mohou tvořit samostatné věty, tj. syntakticky i sémanticky zcela úplné úseky textu (srov. příklad *Důvod je jednoduchý.*). Struktura některých sekundárních konektorů dovoluje nominalizaci druhého argumentu (srov. např. *Důsledkem je stárnutí obyvatelstva.*) – viz tabulka 1.

<b>Primární konektory</b>	<b>Sekundární konektory</b>
syntaktická	struktury s autosémantickou složkou
univerzální	univerzální
gramatikalizované	negramatikalizované
většinou jednoslovné	většinou víceslovné
bez větněčlenské platnosti	větné členy, větné modifikátory či samostatné věty
nemodifikovatelné (s výjimkami)	modifikovatelné (s výjimkami)
	anaforická reference k prvnímu argumentu (explicitní či implicitní)
	jedinečnost některých struktur: a) syntakticky výše než druhý argument b) forma samostatné věty c) nominalizace druhého argumentu

Tabulka 1: Primární vs. sekundární konektory

Díky tomuto idiosynkratickému chování mají sekundární konektory v jazyce jedinečné postavení a diskurzí analýza, která tyto jazykové prostředky nezohledňuje, je podle našeho názoru neúplná. Z tohoto důvodu je zapotřebí brát označení primárních konektorů za centrum množiny diskurzí konektorů a sekundárních konektorů za její periferii především z frekvenčního hlediska, nikoli jako označení hodnotící důležitosti či významu obou skupin.

Primární a sekundární konektory se zároveň mohou v textu objevovat vedle sebe (např. *a z tohoto důvodu; a to je důvod, proč* atd.). V sousedství sekundárních konektorů se

<sup>4</sup> Příklad se sekundárním konektorem: *V rámci projektu se může stát promítačem kdokoli a promítat může kdekoli. Podmínkou je, že se na projekci bezplatně zapůjčeného filmu nebude vybírat vstupné.* Po náhradě primárním konektorem: *V rámci projektu se může stát promítačem kdokoli a promítat může kdekoli, pokud se na projekci bezplatně zapůjčeného filmu nebude vybírat vstupné.*

nejčastěji vyskytují spojky *a*, *ale*, *nebo*. Tyto spojky se ale také často objevují i ve spojení s jiným primárním konektorem (např. *a proto*, *ale také*, *nebo dokonce*, *a tudíž*). Ve spojování sekundárních konektorů s konektory primárními můžeme vidět také analogii k historickému vývoji některých primárních konektorů – srov. např. konektory *ale*, *avšak* apod. (které vznikly ze dvou komponentů *a + -le*; *a + však*).

V naší práci tedy kromě primárních (gramatikalizovaných) a sekundárních (dosud ne zcela gramatikalizovaných) konektorů rozlišujeme ještě konektory **složené** a **nesložené**, **modifikované** a **nemodifikované**.

**Nesložený konektor** v sobě zahrnuje (ze synchronního hlediska) konektor pouze jeden (ať už primární nebo sekundární), např. *a*; *ale*; *avšak*; *z tohoto důvodu*; *za těchto podmínek*. Z příkladů je patrné, že nesložený konektor může být jednoslovný i víceslovný. **Složený konektor** v sobě obsahuje více než jeden konektor (ať už jde o kombinaci dvou primárních konektorů nebo kombinaci primárního a sekundárního konektoru), např. *a tak*; *i když*; *a proto*; *a z tohoto důvodu*. Složené konektory jsou vždy víceslovné.

V některých případech může vyvstat otázka, zda je dané slovní spojení složeným konektorem nebo není. U složených konektorů (typu *a proto*, *i když*) dochází podle našeho názoru k oslabení jedné významové složky (často se jedná o konektory *a*, *i* s konjunktivním významem), tj. složený konektor vnímáme jako jazykový prostředek sloužící k vyjádření primárně jednoho typu významového vztahu (např. tedy složený konektor *a proto* vyjadřuje „silněji“ či zřetelněji vztah příčiny a důsledku než vztah konjunkce, *i když* vztah přípustky apod.). U daných konektorů tedy můžeme pozorovat postupný proces gramatikalizace – *a*, *i* v těchto případech postupně ztrácejí svůj původní význam a stávají se součástí konektoru vyjadřujícího jiný typ diskurzního vztahu. Z historického hlediska můžeme podobný jev pozorovat u již plně gramatikalizovaných, jednoslovných konektorů, jako jsou *ale*, *avšak*, *anebo*. V současné době tedy tyto konektory vyjadřují v textu zcela zřetelně vždy pouze jeden vztah – *ale*, *avšak* vztah opozice, *anebo* vztah disjunktivní či konjunktivní alternativy. Významový vztah konjunkce byl v těchto případech zcela potlačen. Jako složený konektor tedy chápeme výrazy skládající se z více konektorů, u kterých dominuje jeden významový vztah. Pokud si každý ze společně se vyskytujících konektorů zcela zřetelně ponechává svůj původní význam (tj. na daném místě textu silně vnímáme dva různé významové vztahy), považujeme je za navzájem různé diskurzní konektory a u každého z nich označujeme daný významový typ a dosah diskurzního vztahu zvlášť (srov. např. *ale potom*).

Některé konektory mohou být zároveň **modifikované** – srov. např. *hlavně proto*; *právě kvůli tomu*; *z jednoduchých důvodů*; *vlastně proto*, *že*; *především pak*; *jediným důvodem bylo*; *s jasným cílem*, *možná díky tomu* apod. Modifikace se přitom (vzhledem k jejich lexikálně-syntaktické povaze) vyskytuje častěji u sekundárních konektorů.

V některých případech pak může být obtížnější určit, zda se u víceslovného konektoru jedná o konektor složený či modifikovaný. Rozlišovacím kritériem je tedy to, že složené konektory se skládají z více konektorů (tj. všechny části složeného konektoru jsou samy o sobě konektorem – srov. *a tak*, *a proto*, *ale naopak*, *a dokonce*), modifikované konektory naopak obsahují výraz, jehož hlavní funkce není spojovací, ale modifikační, tj. svým významem upřesňuje příslušný diskurzní vztah (říká například, že je daná podmínka hlavní, příčina jednoduchá, cíl jasný apod.). Význam těchto výrazů je tedy především hodnotící (srov. *hlavně*, *jednoduchý*, *jediný*, *jasný*...) či modální (např. *možná*, *vlastně*...).



V naší práci tedy vymezujeme **tři základní opozice**, podle kterých je možné diskurzí konektory členit. Konektory mohou být:

**a) primární** (*však, a, nebo, ale, proto...*) **či sekundární** (*důvodem je, kvůli tomu, z tohoto důvodu*) – hlavním rozlišovacím kritériem je v tomto případě míra gramatikalizace daného výrazu;

**b) složené** (*a tudíž, a kvůli tomu, ale naopak*) **či nesložené** (*nebo, ale, z tohoto důvodu*) – složené konektory se skládají z více konektorů, nesložené se skládají (ze synchronního hlediska) z jednoho konektoru (nesložené konektory mohou být jednoslovné (*ale*) i víceslovné (*z tohoto důvodu*)); složené konektory se v textu mohou vyskytovat vedle sebe (*a proto*), nebo odděleně (*bud'nebo, když\_tak, v případě, že\_pak*), v tomto případě tvoří tzv. korelativní páry;

**c) modifikované** (*hlavně proto, s jasným cílem, jedině kvůli tomu, především vinou toho*) **či nemodifikované** (*proto, s cílem, kvůli tomu, vinou toho*) – modifikované konektory obsahují hodnotící či modální výraz, který upřesňuje význam příslušného diskurzího vztahu.

Uvedené aspekty se mohou vzájemně kombinovat – tj. např. složený konektor může být zároveň modifikovaný apod. (srov. *a hlavním důvodem bylo, a právě proto*). Ve složený konektor se mohou spojovat konektory primární (*a proto*) i primární a sekundární (*a z tohoto důvodu*).

### 3.1.2 Hranice mezi primárními a sekundárními konektory

Sekundární konektory můžeme obecně chápat jako jazykové prostředky, které jsou z jedné strany ohraničené univerzální platností ukazatelů diskurzí vztahů v textu (čímž se odlišují od volných konektivních frází) a z druhé strany (ne)gramatikalizací (čímž se liší od gramaticky i lexikálně ustálených konektorů primárních).

Jednotlivé konektivní prostředky si tak můžeme představit na následující škále či ose zobrazující implicitní a explicitní diskurzí vztahy:

(univerzálnost)

**implicitnost** → volné konektivní fráze → sekundární konektory → primární konektory  
(gramatikalizace)

Jeden pól osy tvoří primární (tj. ustálené, gramatikalizované) konektory, druhý výrazy bez konektivní funkce (diskurzí vztah je pouze implicitní). Mezi těmito dvěma póly stojí sekundární konektory (negramatikalizované, univerzální struktury) a volné konektivní fráze (negramatikalizované, neuniverzální struktury).

Zároveň je podle našeho názoru hranice mezi primárními a sekundárními konektory prostupná. Mezi sekundárními konektory můžeme vysledovat různé podskupiny, které jsou svou podstatou a chováním blízké konektorům primárním. Některé z nich mohou tuto hranici překročit a stát se konektory primárními.

Velkou skupinu sekundárních konektorů tvoří předložkové fráze, tj. struktury obsahující předložku (např. *díky, kvůli, vinou* atd.) a anaforický výraz odkazující k prvnímu diskurzímu argumentu. Jako sekundární konektory tedy v textu působí spojení typu *díky tomu, kvůli tomu, vinou toho* atd. Tyto sekundární konektory jsou podle našeho názoru blízké konektorům primárním.

Podobnou strukturu má z historického hlediska např. primární konektor *proto* vzniklý z předložky *pro* a zájmena *to*. Původně byl tedy tento konektor dvouslovný, posléze prošel gramatikalizací a ustálil se na jednoslovné formě *proto* s primární konektivní funkcí. Produktivnost tohoto tvoření dokládají protějšky českého *proto*, které vznikly podobným způsobem – srov. např. anglický konektor *therefore* (vzniklý spojením slova *there* a předložky *fore*, staroanglické a středoanglické formy předložky *for*; používaný ve významu ‚v důsledku toho‘) či podobný vývoj u konektorů *daarfoor* z holandštiny, *dafür* z němčiny či *derfor* z dánštiny. Obdobný vývoj tohoto konektoru tedy proběhl v několika jazycích, a není tudíž vázaný pouze na jeden konkrétní jazyk. Podobný vývojový proces by tedy mohl nastat i u dalších konektivních struktur, především u těch, které mají obdobnou strukturu, jako měl původně konektor *proto*.

Z tohoto důvodu nechápeme primární a sekundární konektory jako dvě striktně oddělené kategorie, ale jako kategorie s prostupnou hranicí.

### 3.1.3 Vznik primárních konektorů

Vzhledem k předchozím úvahám o prostupnosti hranic mezi primárními a sekundárními konektory a procesu ustalování konektorů jsme se dále zabývali otázkou vzniku primárních konektorů obecně. Na základě prací J. Zubatého a J. Bauera (zabývajících se vznikem vybraných spojek a částic v češtině, např. *a*, *ale*, *avšak*, *přece* apod.) vymezujeme primární konektory jako výrazy vzniklé původně z jiných slovních druhů. Například *a* bylo původně deiktické citoslovce s významem blízcím se novočeskému *hle* (Bauer, 1963: 25). *Ale* se z historického hlediska skládá ze dvou komponentů: *a* (s citoslovečným původem) a *-le* (s významem omezovacího příslovce *jen*), podobně se *a* spojovalo s *však* ve spoju *avšak* (Bauer, 1963: 26), přičemž samotné *však* mělo původně příslovečný význam podobný německému *jedenfalls* s významem ‚vždy‘ (Bauer, 1963: 25). Konektory *proto*, *přesto* jsou ustrnulé předložkové pády, jejichž spojovací funkce se vyvinula z funkce anaforické (tyto výrazy jsou podle Bauera na přechodu mezi příslovci a spojkami; Bauer, 1962: 30). *Tedy* má původ v časovém *tehdy*; *tudíž* v místním (a přeneseně způsobovém) *tudy*. *Přece* vzniklo ze spojení *před se* s původním významem ‚dále kupředu‘ (Bauer, 1962: 31). Jinými slovy, dané výrazy či spojení nabývaly spojovací funkci postupně – v procesu vznikání primárních konektorů docházelo ke ztrátě či oslabování původního lexikálního významu příslušného slova (či jednotlivých složek daného spojení).

Podstatou vzniku primárních konektorů je tedy podle našeho názoru postupné ustalování formy i funkce dané přechodem těchto výrazů či spojení (patřících zpravidla mezi jiné slovní druhy) ke spojkám či strukturujícím částicím a jejich následná gramatikalizace, přičemž tento proces není u některých z nich zatím dokončen.

Podobný proces formování primárních konektorů můžeme sledovat i v jiných jazycích – srov. např. anglický konektor *because* (*protože*) vzniklý ze spojení „by case“ (doslovně ‚podle / vzhledem k případu‘), konektor *nevertheless* (*nicméně*) z víceslovné fráze „never the less“ (ve významu ‚o nic méně‘, analogické k frázi „never the worse“ (ve významu ‚o nic horší/hůře‘)) či primární konektor *for* (*nebot*) vzniklý pravděpodobně ze staroanglických frází typu „for þon þy“ (lit. ‚for the (reason) that‘ – ‚z toho důvodu, že‘) apod. (viz Harper, 2001).

Z tohoto hlediska bychom se mohli dívat na sekundární konektory (tj. výrazy typu *to je důvod*, *proč*) jako na jazykové prostředky, u kterých můžeme sledovat proces postupného ustalování, který ovšem zatím není v takové fázi, že by mohlo dojít k úplné gramatikalizaci. Diskurzívní konektory tedy nechápeme jako uzavřenou množinu výrazů, ale jako škálu jazykových prostředků mapující proces jejich gramatikalizace.

### 3.1.4 Oslabování lexikálního významu u sekundárních konektorů

V rámci procesu ustalování jistého výrazu či spojení jako primárního konektoru dochází k oslabování či úplné ztrátě jeho původního lexikálního významu (*přece* ztratilo význam ‚dále kupředu‘, *a* význam ‚hle‘ apod.). Postupným procesem gramatikalizace tedy dané výrazy či spojení ztrácejí svoji plnovýznamovost (jak jsme popsali výše, spojky a částice, které považujeme za primární konektory, vznikaly z jiných slovních druhů – především příslovcí a citoslovcí, z předložkových frází apod.). Původně se tedy jednalo o plnovýznamová slova či o spojení s plnovýznamovou složkou. Lexikální význam byl ovšem v těchto případech oslaben a dané výrazy se v jazyce ustálily jako synsémantika (příp. jsou stále v procesu ustalování, viz některá příslovce či relativní zájmena s konektivní funkcí).

K postupnému oslabování lexikálního významu dochází také u sekundárních konektorů, tj. postupně se stírají rozdíly mezi plnovýznamovostí a neplnovýznamovostí. Jednotlivé složky sekundárních konektorů postupně ztrácejí svůj původní věcný význam a nový (tj. konektivní) nabývá celá struktura. Jinými slovy, u příslušných autosémantických složek je oslabován jejich původní význam ve prospěch celého spojení. Některé z nich se v jazyce ustalují natolik, že jsou významově zcela nahraditelné primárními konektory – srov. např. sekundární konektory *z toho důvodu*, *z té příčiny*, *pro tu příčinu* a primární konektor *proto* – daná slovní spojení uvádí např. *Slovník spisovného jazyka českého* (1989) jako významové ekvivalenty slova *proto*. Příslušné autosémantické složky (*důvod*, *příčina* aj.) samy o sobě konektivní funkci nemají a v textu mohou vystupovat samostatně v roli plnovýznamových slov.

V kategorii sekundárních konektorů tedy dochází k postupnému ustalování formy, k ustalování konektivní funkce a k oslabování vlastního věcného významu jednotlivých složek příslušné struktury. Tyto struktury pak mají jako celek v textu platnost sekundárních diskurzních konektorů.

## 3.2 Výsledky a evaluace anotací PDT – primární konektory, sekundární konektory, volné konektivní fráze

Výsledky anotace ukázaly, že Pražský závislostní korpus obsahuje v současné době 21 416 diskurzních vztahů vyjádřených explicitně (tj. pomocí konkrétních jazykových prostředků). Těmito prostředky jsou z 94,6 % primární konektory, z 5,4 % sekundární konektory. Z výsledků je patrné, že označení primární a sekundární je pro dané výrazy v tomto ohledu příznačné – k vyjadřování diskurzních vztahů v textu slouží primárně (tj. z 94,6 %) primární konektory. Počet sekundárních konektorů ale není zanedbatelný. Sekundární konektory obohatily diskurzní anotaci o 1 161 nových vztahů, z toho je patrné, že anotace postrádající sekundární konektory by nebyla úplná.

Podle výsledků současné anotace obsahuje PDT dále 151 výskytů volných konektivních frází. Obecně se tedy ukazuje, že pro mluvčího je patrně jednodušší použít pokud možno co nejstálenější (tj. gramatikalizovaný) a většinou také kratší, mnohdy jednoslovný výraz. Mluvčí má tedy (ve většině případů) na výběr různé typy konektivních prostředků (s různou mírou ustálenosti a lexikálního „naplnění“). Nejčastěji přitom volí ten, který je z jeho pohledu při výběru nejjednodušší, tj. nejméně namáhavý při tvorbě textu. Zároveň však mluvčí volí konektivní prostředky s ohledem na adresáta, tj. aby čtenář/posluchač co nejlépe a nejsnadněji interpretoval obsah celého sdělení, tedy i příslušné diskurzní vztahy. Existují případy, kdy je mluvčí právě naopak nucen užít konektivní prostředek s konkrétnějším lexikálním naplněním, který do jisté míry opakuje část předchozího kontextu (např. pokud jsou od sebe diskurzní argumenty v textu příliš vzdáleny

či odděleny jiným textem<sup>5</sup>), aby čtenář lépe pochopil, k jakému úseku textu daný konektivní prostředek odkazuje, a aby se tak zabránilo víceznačností. Volba konektivního prostředku se tedy v první řadě řídí snahou o to, aby bylo sdělení srozumitelné pro adresáta.

Spolehlivost anotace sekundárních konektorů v PDT (která proběhla především pro účely této práce) ukazuje míra anotátorské shody. Shoda na rozpoznání diskurzního vztahu vyjádřeného sekundárním konektorem dosáhla 0,7 míry F1 a shoda na rozpoznání sémantického typu diskurzního vztahu 82 % (resp. 0,78 Cohenovy kappy). Výsledky mezianotátorské shody pro sekundární konektory jsou srovnatelné s obdobným měřením pro primární konektory (viz Poláková a kol., 2013) – srov. 0,83 F1 pro rozpoznání diskurzního vztahu a 77 % (resp. 0,71 C. k.) pro rozpoznání sémantického typu vztahu.

Výsledky anotace sekundárních konektorů a volných konektivních frází v PDT pro češtinu jsme srovnali s anotací podobných výrazů (nazývaných souhrnně „altlexy“<sup>6</sup>) pro angličtinu v korpusu Penn Discourse Treebank (PDTB). Zatímco v PDT tvoří sekundární konektory a volné konektivní fráze 6,1 % (v rámci všech explicitně vyjádřených diskurzních vztahů), v PDTB pouze 3,3 %. Rozdílný počet těchto jazykových prostředků zřejmě není způsoben odlišností či specifičností češtiny a angličtiny, ale patrně do jisté míry odráží skutečnost, že dané výrazy nejsou v PDTB anotovány zcela konzistentně (srov. Prasadová a kol., 2010). Sami autoři očekávají, že se počet některých „altlexů“ po detailnější anotaci může dvakrát až třikrát zvýšit. (Ke srovnání českých sekundárních konektorů a anglických altlexů viz Rysová, 2012.)

### 3.3 Nejčastěji používané primární a sekundární konektory v PDT

Z vyhodnocení výsledků anotace vyplývá, že výskyty jednotlivých primárních konektorů značně převyšují výskyty sekundárních konektorů. Sekundární konektory se v reálných textech navíc objevují v mnoha variantách, tj. ne v jedné ustálené formě jako konektory primární (srov. např. *to je důvod, proč; to je důvodem; je to důvodem toho, že* apod.). Zřetelně se zde tedy projevuje skutečnost, že primární konektory již prošly gramatikalizací, tj. jsou formálně i funkčně ustálené, zatímco konektory sekundární jsou zatím stále v procesu ustalování.

Výsledky ukazují, že nejčastěji se vyskytujícím primárním konektorem v PDT je konektor *a* se 5 765 výskyty, nejčastěji se vyskytujícím sekundárním konektorem výraz *dodal* se 121 výskyty. Sloveso *dodat/dodávat* tvoří spolu s dalšími slovesy říkání (tj. slovesy uvozujícími (značenou i neznačenou) přímou či nepřímou řeč) typu *upřesnit/upřesňovat, doplnit/doplňovat, pokračovat* apod. zcela unikátní kategorii. Od ostatních primárních i sekundárních konektorů se liší tím, že v sobě spojují dvě složky – diskurzni argument, tj. to, že někdo něco říká, a diskurzni konektor – tj. to, že to někdo říká v reakci na něco (srov. *dodal = a řekl*). Obecně tedy můžeme říci, že daná slovesa říkání odkazují v textu k jinému slovesu říkání (*řekl – dodal, řekl – upřesnil, řekl – pokračoval* apod.) či jinak vyjádřenému uvození přímé či nepřímé řeči (*podle jeho slov – doplnil* atd.), přičemž mezi vlastními obsahy

<sup>5</sup> Srov. *Po každé směně jsme je [hlavy, M. R.] museli ostříit. Po 14 dnech nepřetržitého dolad'ování byl stroj již k použití. Úspora, vzniklá koupí našich hlav, se však změnila v pasivum, neboť technik seřizoval stroj místo tří dnů čtrnáct. Kvůli častějšímu broušení hlav vzrostly i náklady na provoz zařízení.* Jako konektivní prostředek je užito spojení *kvůli častějšímu broušení hlav*, díky kterému čtenář příslušný diskurzni vztah snáze interpretuje (pokud bychom užili konektory *kvůli tomu* nebo *proto*, přicházela by v úvahu i následující interpretace – *Technik seřizoval stroj místo tří dnů čtrnáct, kvůli tomu / proto vzrostly i náklady na provoz zařízení.*)

<sup>6</sup> Alternativní lexikální vyjádření konektorů.

příslušných přímých či nepřímých řečí je další, samostatný diskurzivní vztah. (Více o slovesech říkání typu *dodat/dodávat, upřesnit/upřesňovat, doplnit/doplňovat, pokračovat* viz Rysová, 2014.)

### 3.4 Srovnání sémantických typů diskurzivních vztahů pro primární a sekundární konektory

Výsledky anotace dále ukázaly, že rozložení jednotlivých sémantických typů diskurzivních vztahů je u primárních a sekundárních konektorů velice podobné. Diskurzivní vztahy s vysokým počtem výskytů (u primárních i sekundárních konektorů) jsou vztahy konjunkce, příčiny a důsledku, podmínky, přípustky, nesoučasnosti (předčasnosti a následnosti) a konfrontace. Naopak vztahy s nejnižším (či velice nízkým) počtem jsou u obou typů konektorů tzv. **nepravé (pragmatické) vztahy**, tj. vztahy nepravé příčiny a důsledku, nepravé podmínky a výsledku podmínky a nepravý kontrast.

Nižší čísla výskytů pragmatických vztahů ovšem neznamenají, že by se v reálných textech vyskytovaly s tak nízkou frekvencí, ale spíše to, že jim v PDT nebyl zatím věnovaný dostatečný prostor (klasifikace diskurzivních vztahů se prozatím zaměřovala spíše na vztahy sémantické). Dalším krokem při výzkumu koherence textu by tak mohl být pohled z vyšší perspektivy, z pohledu komunikačních strategií autora.

**Primární a sekundární konektory se naopak liší u vztahů opozice a explikace.** Zatímco vztah opozice je u primárních konektorů na druhém místě (tj. s druhým nejvyšším počtem výskytů), u sekundárních konektorů se tento typ diskurzivního vztahu objevuje velice zřídka. Na druhou stranu vztah explikace je u sekundárních konektorů čtvrtým nejpočetnějším typem diskurzivního vztahu, zatímco u konektorů primárních patří mezi vztahy s nižším počtem výskytů.

**Obecně můžeme shrnout, že všechny sémantické typy diskurzivních vztahů jsou v PDT vyjadřovány častěji primárními konektory** (většinou ve více než 90 % všech výskytů) – ani u jednoho typu vztahu tedy nepřevážil výskyt sekundárních konektorů. Nepatrně vyšší procentuální zastoupení sekundárních konektorů (ve srovnání s ostatními typy vztahů) mají pouze vztahy explikace (sekundární konektory tvoří 22,7 %), exemplifikace (16,9 %) a generalizace (16,7 %), i když i v těchto případech zastoupení primárních konektorů výrazně převažuje.

#### 3.4.1 Možnosti rozšíření klasifikace diskurzivních vztahů vyjadřovaných sekundárními konektory

Klasifikace sémantických typů diskurzivních vztahů pro sekundární konektory byla převzata z anotace primárních konektorů (srov. Poláková a kol., 2013). V průběhu anotace se ovšem ukázalo, že sekundární konektory vyjadřují i jiné typy vztahů, než se objevují ve zmíněné klasifikaci. Anotátoři upozornili celkem na tři: a) vyplývání (či vyvození výsledků – vyjadřované sekundárními konektory *z toho vyplývá, že; v každém případě; výsledkem bylo; výsledkem je*), b) vyvození závěru (*závěr zněl, že; došel k závěru, že; jednoznačný závěr je, že; dospělo k tomuto závěru*), c) zřetel (*v tomto směru, se zřetelem k tomu*).

Klasifikace diskurzivních vztahů by tak (na základě zkušenosti s reálnými texty) měla být pro sekundární konektory rozšířena.

### 3.5 Srovnání diskurzních konektorů v rámci mezivětných a vnitrovětných vztahů

V rámci analýzy primárních a sekundárních konektorů jsme dále zjišťovali, jaký je u nich poměr mezivětných a vnitrovětných vztahů, tj. zda primární a sekundární konektory, případně některé sémantické typy diskurzních vztahů tíhnou spíše k mezivětnému či vnitrovětnému vyjádření. Z analýzy jsme vyvodili následující závěry.

1) **Primární konektory vyjadřují častěji vnitrovětné diskurzní vztahy** než vztahy mezivětné (srov. 70 % vs. 30 %), naopak **sekundární konektory vyjadřují častěji vztahy mezivětné** (srov. 63 % vs. 37 %).

Pokud je tedy diskurzní vztah vyjádřen například spojkou *a*, je její častější užití *Mám rýmu a nepůjdu do školy.* než *Mám rýmu. A nepůjdu do školy.* Naopak pokud je diskurzní vztah vyjádřen například spojením *z tohoto důvodu*, je častěji užito ve sledu vět typu *Mám rýmu. Z tohoto důvodu nepůjdu do školy.* než v souvětí typu *Mám rýmu, z tohoto důvodu nepůjdu do školy.*

Obecně (v absolutních číslech) jsou ale diskurzní vztahy vnitrovětné i mezivětné častěji vyjadřovány konektory primárními. Existují však výjimky, viz následující bod 2).

2) Celkově jsou mezivětné i vnitrovětné diskurzní vztahy vyjadřovány především primárními konektory (celkový počet primárních konektorů v PDT je více než 17krát vyšší než celkový počet sekundárních konektorů). I přesto **pro vyjádření vztahu podmínky dominuje podle dat PDT užití sekundárního konektoru** (v absolutních počtech výskytů).

Podle PDT se tak například častěji užívá *Půjdu ven. V případě, že nebude pršet.* než *Půjdu ven. Jestliže nebude pršet.*

3) **U některých typů diskurzních vztahů vyjádřených sekundárními konektory může mít vliv na to, zda budou realizovány mezivětně či vnitrovětně, také syntaktická povaha daných diskurzních argumentů.** U vztahu restriktivní opozice (signalizované např. spojením *výjimkou je* či *výjimku tvoří*) mezi dvěma slovesnými argumenty převažuje vyjádření vnitrovětné nad mezivětným. U vztahů restriktivní opozice s jedním nominálním diskurzním argumentem naopak převládá vyjádření mezivětné. (Ačkoli počet nalezených výskytů v PDT je v tomto případě poměrně nízký, popisovaný jev je tedy třeba chápat pouze jako možnou tendenci.)

Podle PDT se tak například častěji se užívá *Fotbalové mužstvo nedostalo žádné nové posily. Výjimkou je příchod nadějného záložníka.* než *Fotbalové mužstvo nedostalo žádné nové posily, výjimkou je příchod nadějného záložníka.*

Ale zároveň se podle PDT častěji používá například *Fotbalové mužstvo nedostalo žádné nové posily, výjimkou je, že přišel nadějný záložník.* než *Fotbalové mužstvo nedostalo žádné nové posily. Výjimkou je, že přišel nadějný záložník.*

### 3.6 Srovnání sekundárních konektorů v psané a mluvené češtině

Na základě analýzy vybraných konektivních struktur v Pražském závislostním korpusu (obsahujícím psané texty) a v korpusu mluvené češtiny DIALOG můžeme vyvodit následující závěry.

1) **Počet vybraných struktur s konektivní funkcí je vyšší v Pražském závislostním korpusu, tj. v psaných textech** (PDT obsahuje 252 výskytů daných struktur s konektivní funkcí, tj. 49 % z celkového počtu; DIALOG 180 výskytů, tj. 27 %). Vysvětlení může

spočívat v tom, že některé konektivní struktury typu *toto počínání je důvodem, proč; své chování odůvodnil tím, že; důvodem poklesu je* apod. mohou působit formálněji (než například neutrální konektory, jako jsou *proto, protože* aj.), a proto se objevují častěji v psaných textech.

2) **V psané češtině je obecně vyšší tendence k nominalizaci (či kondenzaci) druhého diskurzního argumentu než v češtině mluvené.** Tj. příklady typu *důvodem je jeho napadení ředitele* (oproti slovesnému vyjádření *důvodem je, že napadl ředitele*) jsou častější v psaných textech.

3) V psané i mluvené češtině se častěji vyskytují univerzální konektivní struktury (tj. sekundární konektory) než neuniverzální (tj. volné konektivní fráze). V mluvené češtině je přitom rozdíl v počtu univerzálních a neuniverzálních struktur zcela jednoznačný – v průměru se v korpusu DIALOG vyskytuje 96 % univerzálních a 4 % neuniverzálních struktur (obsahujících zkoumaná slova), v korpusu PDT je poměr poněkud vyrovnanější – 62 % vs. 38 %.

Obecně tedy můžeme říci, že **konektivní struktury typu *důvodem pro propuštění bylo, příčinou poklesu je, toto počínání je důvodem, proč* jsou typičtější pro psané texty.** Psaná čeština se tedy v tomto ohledu vyznačuje větší lexikální rozmanitostí než čeština mluvená.

4) Obdobně se konektivní struktury se slovem *příčina* vyskytují v korpusu DIALOG pouze ve spojení se slovesem *být* (tj. např. *hlavní příčinou je; příčina je jednoduchá; příčiny jsou nejrůznější; příčiny jsou ty, že* apod.), zatímco v PDT i s jinými slovesy (srov. *vidět příčinu v tom, že; hledat příčinu v tom, že; označit za příčinu*). Z tohoto hlediska je tedy **spektrum konektivních struktur v psaných textech rozmanitější než v textech mluvených.**

5) **Konektivní struktury obsahující sloveso *souviset* (typu *s tím souvisí, to souvisí s* apod.) se v mluvené češtině používají především** (v korpusu DIALOG v 90 % vztahů vyjádřených konektivní strukturou se slovesným základem *souviset*) **k prostému, volnějším navazování dvou různých témat, přičemž význam daného diskurzního vztahu je spíše vágní, zamlžený** – srov. např. *s tím souvisí i druhá věc, to souvisí s celou řadou dalších otázek, to souvisí ještě s jednou věcí, s tím souvisí třetí problém* apod.

**V psané češtině jsou diskurzní vztahy vyjádřené slovesnými frázemi typu *s tím souvisí* naopak ve většině případů „pevnější“, významově jasnější či zřetelnější** (v korpusu PDT jsou tyto vztahy zastoupeny 75 %). Nevyjadřují tedy prosté navazování, ale mohou se významově blížit např. vztahům příčiny a důsledku či explikace.

### 3.7 Testování vhodnosti sekundárních konektorů pro různé kontexty – princip univerzálnosti

Diskurzní konektory vymezujeme v této práci na základě jejich vhodnosti (univerzálnosti) jako konektivních prostředků v různých kontextech, tj. na základě toho, zda daný výraz může být použit jako konektivní prostředek v textu z velké míry univerzálně nebo je závislý na určitém specifickém kontextu. Pro testování univerzálnosti jsme vybrali dvě skupiny strukturálně odlišných sekundárních konektorů – ty, jejichž jádro tvoří sekundární předložky *díky, kvůli, na základě, vzhledem k, vinou*, a ty, jejichž jádrem je podstatné jméno *důvod*. Jedná se přitom o konektivní spojení, která signalizují diskurzní vztah příčiny a důsledku.

**Pro testování jsme zvolili sto různých kontextů z PDT, ve kterých se vyskytuje diskurzní vztah příčiny a důsledku. Do těchto kontextů jsme posléze dosazovali**

**jednotlivé konektivní struktury a zjišťovali, pro jaké procento kontextů jsou vhodnými ukazateli diskurzního vztahu.**

Zvolili jsme přitom padesát kontextů z PDT, ve kterých se v originálním textu objevil sekundární konektor (příp. volná konektivní fráze), a padesát kontextů s primárními konektory *proto* a *a tak*.

**Na základě výsledků výzkumu jsme pro vymezení diskurzních konektorů stanovili hranici 75 % univerzálnosti**, která se ukázala být významným zlomem oddělujícím výrazy typu *z tohoto důvodu; vzhledem k této skutečnosti* apod. od výrazů jako *díky této skulince, díky jeho chorobě* apod. **Nejvyššího stupně univerzálnosti dosáhly v našem výzkumu sekundární konektory z uvedeného důvodu / z uvedených důvodů (99 %), z tohoto důvodu / z těchto důvodů (98 %) a kvůli tomu (93 %)**. Zároveň je ovšem zapotřebí zohlednit skutečnost, že jsme příslušné výrazy testovali na publicistických textech a že na užívání a chování konektivních prostředků může mít vliv právě funkční styl zkoumaných textů.

U jednotlivých výrazů či spojení jsme zároveň zkoumali jejich **frekvenci v textu v korpusu SYN2013PUB, tj. zaměřili jsme se na to, zda při určování příslušného jazykového prostředku jako diskurzního konektoru můžeme zohledňovat také jeho frekvenci v reálných textech, příp. jaký je vztah frekvence a univerzálnosti.**

Výsledky analýzy ukázaly, že nejvyšší frekvenci v reálných textech mají primární konektory – *proto* (733 198 výskytů), *a tak* (228 532) a dále předložkové fráze obsahující zájmeno *ten* – např. *kvůli tomu* (53 755), *díky tomu* (40 376), *vzhledem k tomu* (39 497). Všechny tyto výrazy mají zároveň vysokou míru univerzálnosti, tj. vhodnosti užití jako konektivního prostředku ve velkém množství různých kontextů.

Většina spojení, která dosáhla velmi nízké univerzálnosti, se zároveň v korpusu SYN2013PUB vyskytuje s malou frekvencí (např. spojení *díky této skulince* má univerzálnost 1 % a zároveň nemá žádný výskyt v SYN2013PUB). V těchto případech je tedy univerzálnost daných jazykových prostředků v souladu s jejich frekvencí v reálných textech.

Na druhou stranu se některá spojení s vysokou mírou univerzálnosti vyskytují v korpusu SYN2013PUB s poměrně nízkou frekvencí (ve srovnání např. s primárními konektory; spojení *vzhledem k této skutečnosti* dosáhlo univerzálnosti 88 %, v korpusu SYN2013PUB se vyskytuje ve 145 výskytech), příp. se v uvedeném korpusu nevyskytují vůbec (srov. spojení *(právě) tato skutečnost se stala jedním z důvodů, kvůli kterým* s univerzálností 85 %).

V korpusu SYN2013PUB jsme zároveň pro ilustraci vyhledali spojení *kvůli své nemoci*, které má pravděpodobně (analogicky k testovanému *díky jeho chorobě*) nízkou univerzálnost – v reálných textech korpusu se objevilo ve 106 výskytech (tj. v textech se vyskytuje s podobnou frekvencí jako např. *vzhledem k této skutečnosti*).

**Z těchto důvodů nebereme frekvenci jazykových prostředků v reálných textech jako hlavní kritérium pro určování hranice diskurzních konektorů (jako základní kritérium chápeme jejich univerzálnost pro použití v různých kontextech). Nicméně frekvence příslušných výrazů vypovídá o míře jejich ustálenosti, tj. v rámci kategorie diskurzních konektorů frekvence supluje hranici mezi konektory primárními a sekundárními** (jak jsme popsali výše, v PDT je mezi všemi výskyty diskurzních konektorů 94,6 % primárních konektorů a 5,4 % sekundárních).

Vzhledem k tomu, že sekundární konektory často obsahují explicitní anaforickou referenci, zkoumali jsme dále to, která jména se na této pozici mohou objevit, aby měla celá struktura v textu funkci diskurzního konektoru. Zaměřili jsme se přitom na **obecnější jména typu skutečnost, činnost, opatření** apod., protože významově konkrétnější jména typu



*skulinka* či *choroba* se v rámci konektivních struktur ukázala zcela jednoznačně jako kontextově závislá (tj. s univerzálností 0–2 %). Vybraná obecná jména jsme testovali ve spojení se sekundární předložkou *na základě* (tj. např. *na základě této skutečnosti, na základě této činnosti, na základě tohoto opatření* apod.), která v našem výzkumu dosáhla ve spojení se zájmenem *ten* univerzálnosti 94 %, tj. nejvíce mezi ostatními obdobnými předložkovými frázemi *kvůli tomu* (93 %), *díky tomu* (77 %), *vzhledem k tomu* (77 %) a *vinou toho* (34 %) (použití spojení *vinou toho* je limitováno úzkým sémantickým vymezením předložky *vinou* vázanou na negativní kontext, přesto dané spojení považujeme za diskurzni konektor – vzhledem k jeho analogické struktuře ke konektorům typu *kvůli tomu*).

Výsledky výzkumu ukázaly, že konektivními prostředky s obecnými jmény, které dosahují nejvyšší univerzálnosti, jsou spojení ***na základě toho* (94 %), *na základě tohoto faktu / těchto faktů* (93 %) a *na základě této skutečnosti / těchto skutečností* (91 %)**. Tato spojení jako jediná mezi ostatními zkoumanými strukturami s obecnými jmény přesahují hranici 75 % univerzálnosti, tj. použitelnosti ve funkci konektivního prostředku v různých kontextech. **Z tohoto důvodu považujeme (v rámci zkoumaného vzorku dat) za diskurzni konektory pouze takové struktury s explicitně vyjádřenou anaforickou referencí, na jejíž pozici stojí anaforické výrazy *ten, tento fakt* či *tato skutečnost* (případně *vnitrovětné což, viz níže*), tj. např. *kvůli tomu; vzhledem k těmto skutečnostem; díky tomuto faktu; to je důvod, proč; tyto skutečnosti jsou hlavní příčinou toho, že* apod.**

Výsledky výzkumu dále poukázaly na skutečnost, že některé konektivní prostředky jsou vázané syntakticky – výraz *na základě čehož* byl vhodný pouze pro 21 % vybraných kontextů. V testovaných kontextech však dominovaly mezivětné diskurzni vztahy. Když jsme daný výraz testovali znovu pouze na vztazích vnitrovětných, výsledek vzrostl na 81 %. Z tohoto důvodu považujeme spojení *na základě čehož* za sekundární diskurzni konektor vyjadřující diskurzni vztahy primárně uvnitř věty, podobně jako některé primární konektory typu *což* nebo *příčemž*.

Z výsledků výzkumu je tedy zřejmé, že vedle univerzálnosti diskurzni konektorů (tj. jejich vhodnosti v rámci velkého vzorku kontextů) je při výběru konektivních prostředků zapotřebí zohledňovat i další faktory.

Výsledky naší analýzy ukázaly, že jedním z nich může být také vzdálenost diskurzni argumentů, tj. zda jsou dané argumenty v textu vedle sebe nebo zda je mezi ně vložen další text, a je tedy zapotřebí užít (pokud hrozí nesprávná interpretace daného vztahu) konektivní prostředek s vyšší lexikální naplněností (který jednoznačně odkazuje k příslušnému diskurzni argumentu).

**Na základě našeho výzkumu tedy můžeme shrnout hlavní faktory uplatňující se při výběru vhodných konektivních prostředků. Jsou jimi vzájemná vzdálenost diskurzni argumentů, sémanticko-pragmatické aspekty konektorů (tj. např. míra jejich subjektivity), syntaktické chování konektorů (tj. vázanost na vnitrovětné či mezivětné vztahy), funkční styl příslušného textu (tj. zohlednění stylistické platnosti konektoru) a míra lexikální naplněnosti konektoru. Některé z uvedených faktorů jsou přitom vzájemně provázané. Zároveň právě kvůli těmto různým faktorům a omezením většinou nedosahují primární ani sekundární konektory úplné univerzálnosti 100 %.**

## Závěrem

V předkládané disertační práci jsme se zaměřili na jazykové prostředky, které se významně podílejí na strukturaci textu – tzv. diskurzí konektory. Ukázali jsme, že diskurzí konektory představují široké spektrum jazykových výrazů, od plně gramatikalizovaných po dosud neustálené. Proces jejich ustalování probíhá v jazyce již několik staletí, z tohoto hlediska nejsou diskurzí konektory uzavřenou množinou jazykových prostředků, ale jako celek procházejí neustále vývojem. Diskurzí konektory si tak můžeme představit jako množinu výrazů, která má své více méně neměnné centrum (tzv. primární konektory) a svoji periferii (tzv. sekundární konektory). Provedený výzkum těchto prostředků realizovaný na rozsáhlém jazykovém materiálu by mohl přispět ke zlepšení porozumění vztahům v rámci textu, a tedy k prohloubení obecného povědomí o celkové koherenci textu.

## Literatura (výběr)

- ADAMEC, Přemysl. Konektivní částice a jiné textově propojovací výrazy v současné češtině. In *Přednášky z 37. a 38. běhu LŠSS*. Praha: Univerzita Karlova, 1995, s. 59–64.
- ASHER, Nicholas a kol. Modelling Strategic Conversation: The STAC Project. In BROWN-SCHMIDT, S.; GINZBURG, J.; LARSSON, S. (eds.). *Proceedings of 16th Workshop on the Semantics and Pragmatics of Dialogue*, 2012, s. 145–146.
- ASHER, Nicholas. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer Academic Publishers, 1993.
- ASHER, Nicholas; LASCARIDES, Alex. *Logics of Conversation*. Cambridge University Press, 2003. ISBN 9780521659512.
- ASR, Fatemeh Torabi; DEMBERG, Vera. On the Information Conveyed by Discourse Markers. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistic*. Association for Computational Linguistics, Sofia, Bulgaria, 2013, s. 84–93. ISBN 978-1-937284-61-9.
- BAJZÍKOVÁ, Eugénia. *Úvod do textovej syntaxe*. Bratislava, 1979.
- BAUER, Jaroslav. Podíl citoslovcí na vzniku českých spojek. *Sborník prací Filosofické fakulty brněnské university A 11*, 1963, s. 21–28.
- BAUER, Jaroslav. Relativa a spojky. *Acta Universitatis Carolinae – Philologica, Slavica Pragensia IV*, 1962, s. 221–226.
- BAUER, Jaroslav. Spojky a částice. *Sborník prací Filosofické fakulty brněnské university A 12*, 1964, s. 131–139.
- BAUER, Jaroslav. Spojky a příslovce. *Sborník prací Filosofické fakulty brněnské university A 10*, 1962, s. 29–37.
- BAUER, Jaroslav. Spojky. *Český jazyk 8*, 1958, s. 220–226.
- BLAKEMORE, Diane. *Relevance and Linguistic Meaning. The Semantics of Discourse Markers*. Cambridge: Cambridge University Press, 2002.
- BLAKEMORE, Diane. *Semantic constraints on relevance*. Oxford: Blackwell, 1987.
- BLASS, Regina. *Relevance Relations in Discourse: A Study with Special Reference to Sissala*. Cambridge: Cambridge University Press, 1990. ISBN-10 0521385156. ISBN-13 978-0521385152.
- BROWN, Gillian; YULE, George. *Discourse analysis*. Cambridge: Cambridge University Press, 1983.

- DANEŠ, František. *Věta a text*. Praha: Academia, 1985.
- DANEŠ, František. Zum Status der Textebene. In: DANEŠ, F.; VIEHWEGER, D. (eds.). *Probleme der Textgrammatik II*. Berlin: Akademie-Verlag, 1977, s. 153–158.
- HAIČOVÁ, Eva; HAVELKA, Jiří; SGALL, Petr. Topic and focus, anaphoric relations and degrees of salience. In *Prague Linguistic Circle Papers / Travaux du cercle linguistique de Prague N.S.* Amsterdam: John Benjamins, 2004.
- HAIČOVÁ, Eva; SGALL, Petr; BURÁŇOVÁ, Eva. Topic-Focus Articulation and degrees of salience in the Prague Dependency Treebank. In *Formal Approaches to Function in Grammar. In honor of Eloise Jelinek Arizona*. Amsterdam/Philadelphia: John Benjamins, 2003, s. 165–177.
- HAKULINEN, A. The use of Finnish *nyt* as a discourse particle. In JUCKER, A. H.; ZIV, Y. (eds.). *Discourse markers. Description and theory*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1998, s. 83–96.
- HALLIDAY, Michael Alexander Kirkwood; HASAN, Ruqaiya. *Cohesion in English*. London: Longman, 1976.
- HARPER, Douglas. *Online Etymology Dictionary*. 2001 [cit. 2015-05-15]. Dostupné z WWW: <<http://www.etymonline.com>>.
- HARRIS, Zellig Sabbetai. Discourse Analysis. *Language* 28: 1, 1952, s. 1–30.
- HAUSENBLAS, Karel. Gramatická syntax textu. In *Referati od X. zasedanie na slovenskíe literaturni jazyci*. Skopje, 1979, s. 175–192.
- HAUSENBLAS, Karel. O studiu syntaxe běžně mluvených projevů. In *Otázky slovanské syntaxe I*. Praha, 1962, s. 313–325.
- HAUSENBLAS, Karel. On the characterization and classification of discourses. In *Travaux linguistiques de Prague 1*, 1964.
- HAUSENBLAS, Karel. Text, komunikáty a jejich komplexy. Zamyšlení pojmoslovné. *Slovo a slovesnost* 45 (1), 1984, s. 1–7.
- HAUSENBLAS, Karel. *Výstavba jazykových projevů a styl*. Praha: Univerzita Karlova: Mir, 1971.
- HAUSENBLAS, Karel; HOFFMANNOVÁ, Jana. Od syntaxe k stavbě textu. In *Syntax a jej vyučovanie*. Bratislava, 1982, s. 427–440.
- HOEY, Michael. *Signalling in discourse*. Birmingham UK: English Language Research Unit, University of East Anglia, 1979.
- HOEY, Michael. *Textual Interaction: An Introduction to Written Discourse Analysis*. London: Routledge, 2001.
- HOFFMANNOVÁ, Jana. Lingvistická pragmatika a interpretace textu. *Jazykové aktuality* 24, 1987, s. 95–96.
- HOFFMANNOVÁ, Jana. Pojem relevance v současné lingvistice, v jazykovém vyučování a v pedagogické komunikaci. *Český jazyk a literatura* 41, č. 1/2, 1990/1, s. 7–15.
- HOFFMANNOVÁ, Jana. Pragmatická perspektiva a současná lingvistika. *Slovo a slovesnost* 51, 1990, s. 131–139.
- HOFFMANNOVÁ, Jana. *Sémantické a pragmatické aspekty koherence textu*. Praha: Ústav pro jazyk český, 1983.
- HOFFMANNOVÁ, Jana. *Stylistika a... Současná situace stylistiky*. Praha: Trizonia, 1997.
- HOFFMANNOVÁ, Jana. Typen der Konektoren und deren Anteil an der Organisierung des Textes. In: KOŘENSKÝ, J.; HOFFMANNOVÁ, J. (eds.). *Text and pragmatic aspects of language*, 1984, s. 101–140.
- HOVY, Eduard H. Parsimonious or profligate approaches to the question of discourse structure relations. In *Proceedings of the 5th International Workshop on Natural Language Generation*, 1990, s. 128–34.

- HRBÁČEK, Josef. *Nárys textové syntaxe spisovné češtiny*. Praha: Trizonia, 1994. ISBN 80-85573-51-2.
- HRBÁČEK, Josef. Text, promluva a nadpromluvvové útvary. *Slovo a slovesnost* 52 (2), 1991, s. 123-129.
- KNOTT, Alistair. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Edinburgh: University of Edinburgh, disertační práce, 1996.
- MANN, William C.; THOMPSON, Sandra A. *Rhetorical Structure Theory: Description and Construction of Text Structures*. Defense Technical Information Center, 1986.
- MARTIN, James R. *English text: System and structure*. Amsterdam: Benjamins, 1992.
- MISTRÍK, Jozef. *Štruktúra textu*. Československý rozhlas, 1975.
- Mluvnice češtiny* 2. Praha: Academia, 1986.
- Mluvnice češtiny* 3. Praha: Academia, 1987.
- POLÁKOVÁ, Lucie a kol. Introducing the Prague Discourse Treebank 1.0. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*. Asian Federation of Natural Language Processing, Nagoya, Japan, 2013, s. 91–99. ISBN 978-4-9907348-0-0.
- POLÁKOVÁ, Lucie a kol. *Manual for Annotation of Discourse Relations in Prague Dependency Treebank*. Technical report no. 2012/47, Prague: ÚFAL MFF UK, 2012a.
- POLÁKOVÁ, Lucie a kol. *Prague Discourse Treebank 1.0*. Data/software, Prague: ÚFAL MFF UK, 2012b.
- PRASAD, Rashmi a kol. The Biomedical Discourse Relation Bank. *BMC Bioinformatics*, 2011, s. 1–18.
- PRASAD, Rashmi a kol. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco, 2008.
- PRASAD, Rashmi; JOSHI, Aravind; WEBBER, Bonnie. Realization of Discourse Relations by Other Means: Alternative Lexicalizations. In *Coling 2010: Posters (August 2010)*, 2010, s. 1023–1031.
- PRASAD, Rashmi; WEBBER, Bonnie; JOSHI, Aravind. Reflections on the Penn Discourse TreeBank, Comparable Corpora, and Complementary Annotation. *Computational Linguistics* 40 (4), 2014, s. 921–950.
- RYSOVÁ, Magdaléna. Alternative Lexicalizations of Discourse Connectives in Czech. In CALZOLARI, N. a kol. (eds.). *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turecko: European Language Resources Association (ELRA), 2012, s. 2800–2807. ISBN 978-2-9517408-7-7.
- RYSOVÁ, Magdaléna. Jazykové prostředky vyjadřující textové vztahy v češtině a jejich zpracování v Pražském závislostním korpusu. *Bohemistyka* 13 (1), 2013b, s. 57–71. ISSN 1642–9893.
- RYSOVÁ, Magdaléna. Verbs of Saying with a Textual Connecting Function in the Prague Discourse Treebank. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík, Island: European Language Resources Association (ELRA), 2014, s. 930–935. ISBN 978-2-9517408-8-4.
- RYSOVÁ, Magdaléna. Způsoby vyjadřování diskurzivních vztahů v češtině a jejich anotace v Pražském závislostním korpusu. In CZOK, M. a kol. (eds.). *Między regionalizmami a kosmopolityzmem. Polska, niemiecka i czeska literatura, język i kultura. Materiały IX Międzynarodowej Konferencji interFaces we Wrocławiu i w Karpaczu*. Leipzig – Dresden – Wrocław: ATUT, Neisse Verlag, 2015, s. 233–241. ISSN 1426-7241. ISBN 978-83-7977-110-3. ISBN 978-3-86276-161-6.

- RYSOVÁ, Magdaléna. Způsoby vyjadřování vztahu příčiny a důsledku v textu – alternativy pro „proto“ a „protože“. *Didaktické studie* 5 (2), 2013c, s. 39–53. ISSN 1804-1221.
- RYSOVÁ, Magdaléna; RYSOVÁ, Kateřina. The Centre and Periphery of Discourse Connectives. In AROONMANAKUN, W. a kol. (eds.). *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing (PACLIC 28)*. Bangkok, Thailand: Department of Linguistics, Faculty of Arts, Chulalongkorn University, 2014, s. 452–459. ISBN 978-616-551-887-1.
- Slovník spisovného jazyka českého*. 2., nezměněné vyd. Praha: Academia, 1989.
- STEDE, Manfred; NEUMANN, Arne. Potsdam Commentary Corpus 2.0: Annotation for discourse research. In CALZOLARI, N. a kol. (eds.). *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*.
- TÁRNYIKOVÁ, Jarmila. Discourse markers in cross-cultural perspective. In *Perspectives* 8, 1998, s. 37–44.
- TÁRNYIKOVÁ, Jarmila. *From Text to Texture: An Introduction to Processing Strategies*. Olomouc: Universita Palackého v Olomouci, 2002. ISBN 80-244-0438-9.
- TÁRNYIKOVÁ, Jarmila. Sentence adverbials or discourse markers? In *Abstracts. International Conference Bridges and Interfaces: Form, Meaning and Function*, Prague, 1998.
- ZUBATÝ, Josef. Náli, nali-ť. *Listy filologické* 37, 1910, s. 217–228.
- ZUBATÝ, Josef. O jistém způsobu užívání zájmena an a on a o jeho původě. *Listy filologické* 36 (1), 1909, s. 25–34.

## Zdroje dat

- BEJČEK, Eduard a kol. *Prague Dependency Treebank 3.0*. Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Praha, 2013.
- Český národní korpus – SYN2013PUB. Ústav Českého národního korpusu FF UK, Praha, 2013. Dostupné z WWW: <<https://ucnk.ff.cuni.cz/syn2013pub.php>>.
- Korpus DIALOG 1.1. Praha: Ústav pro jazyk český, AV ČR, 2012. Dostupné z WWW: <<http://ujc.dialogy.cz>>.