

STRUCTURATION GLOBALE DU TEXTE : UNE ÉTUDE DE CORPUS

Šárka Zikánová, Jiří Mírovský, Lucie Poláková

Résumé :

La présente étude constitue une étape préparatoire à l'analyse de la structuration globale du texte (cohérence globale) dans le Corpus pragois de dépendances. Nous avons analysé les phénomènes relatifs à la structuration au niveau global, leur marquage ayant été effectué en tant que produit secondaire de l'annotation des relations discursives implicites. Ces phénomènes peuvent être rassemblés en quatre grands groupes que nous décrivons plus en détail : des phénomènes de structuration interne d'une ligne narrative, des changements de perspectives dans le texte relevant des tours de parole entre les énonciateurs, des phénomènes de structuration externe du texte et des phénomènes liés aux règles de composition inhérentes au genre de texte donné. Nous avons tenté également d'intégrer ces phénomènes dans un tableau reprenant les aspects existants ou en préparation de l'analyse textuelle dans le Corpus pragois de dépendances en guise de proposition pour l'enrichissement de l'analyse de la cohérence textuelle.

Mots-clés : structuration globale du texte, cohérence globale, Corpus pragois de dépendances, connecteurs discursifs

Le texte, en tant qu'entité complexe permettant la transmission d'informations entre locuteurs, possède une structure extraordinairement riche. Il contient des informations sur le monde réel, sur les intentions de différents acteurs, sur sa propre structure ainsi que des informations sur l'information elle-même. Il y a même lieu de s'étonner de la capacité humaine à utiliser de telles entités, à les former et à les comprendre.

Chaque plan d'organisation du texte fait l'objet de recherches dans une branche différente de la linguistique. Il existe en outre des tentatives de description de la structure du texte dans son ensemble, comme la conception des microstructures et des macrostructures de Van Dijk (Van Dijk, 1980), l'approche de Halliday et Hasan, qui décrivent la textualité comme un réseau de relations de différentes natures (Halliday et Hasan, 1976) ou encore le concept précoce de la linguistique textuelle qui définit les sept critères connus de la textualité (Dressler, 1972 ; de Beaugrande et Dressler, 1981). À notre connaissance, ces conceptions émanent de représentations théoriques que les données n'ont cependant pas (jusqu'à présent) pu refléter dans leur complexité, et pour lesquelles la pratique n'a donc pas encore pu apporter de corrections¹. C'est pourquoi ces recherches se focalisent sur des aspects précis et manifestes, en omettant certains n'ayant pas attiré l'attention des linguistes, quoique fréquents dans la langue.

Notre étude se base sur les données linguistiques et les annotations du Corpus pragois des dépendances (dit PDT, Hajič et coll., 2020). Celui-ci a été créé sur la base de la Description fonctionnelle générative de la langue (DFG) qui représente un schéma à plusieurs niveaux du système linguistique (Sgall et coll., 1986). En appliquant cette approche stratificationnelle aux données du corpus, il résulte que l'analyse d'un aspect de la langue soulevait chaque fois des interrogations à d'autres niveaux. Ainsi, petit à petit a émergé une description systématique de la phrase, qui a permis d'appréhender dans différentes

¹ À l'exception de la conception de Halliday et Hasan qui a été appliquée au corpus GeCCo (Menzel et Lapshinova-Koltunski, 2014).

directions la description des phénomènes supraphrastiques (Mírovský et coll., 2012), et en particulier des phénomènes de cohérence textuelle.

Pour la présente contribution, notre objectif est de proposer un modèle permettant de cerner l'organisation du texte, en prenant en compte différents aspects de l'organisation textuelle mais de manière à pouvoir suivre leur rôle respectif, et sans les fondre dans une classification, c'est-à-dire en saisissant leurs différentes natures. Concrètement nous nous occuperons des phénomènes de structuration globale du texte (les études anglo-américaines utilisent la notion de « cohérence globale », voir Samet et Schank, 1984)² que nous avons observés à travers l'analyse fondamentale de la cohérence. Ces phénomènes, nous les avons classés en sous-groupes, les avons incorporés à la description actuelle du texte dans le cadre théorique que nous avons défini et avons déterminé quels domaines de l'organisation du texte n'y seraient, pour diverses raisons, pas incorporés.

La classification proposée provient de données linguistiques conséquentes et est formée de la même manière que les types d'annotations précédents dans le Corpus pragois de dépendances³ : l'analyse de ce qu'on appelle cohérence locale, c'est-à-dire l'étude des relations discursives entre des segments de texte minimaux et qui est en cours depuis 2008 dans le PDT, a apporté comme acquis secondaire de nouvelles données empiriques sur la structuration globale du texte, la cohérence entre les paragraphes, etc. Ce sont précisément ces nouvelles connaissances que nous décrivons et systématisons dans ce travail.

L'analyse que nous présentons comporte plusieurs aspects, qui ont déjà été annotés, et une ébauche de ceux qu'il est encore possible d'étudier, notamment dans les corpus, afin que l'analyse de la cohérence textuelle soit la plus exhaustive possible. Nous sommes conscient que notre connaissance des phénomènes de cohérence dépend du type et de la taille du matériel linguistique que nous pouvons traiter, et que nous ne saurions dès lors considérer notre proposition comme définitive ou la seule correcte et possible. Notre objectif est, autant que possible, d'enrichir le domaine de l'étude de la cohérence textuelle par des faits empiriques issus de l'analyse linguistique minutieuse de vastes données annotées dans un corpus. Nous signalerons en outre quelques aspects de la langue que nous avons pour le moment écartés de notre recherche.

Les résultats de l'analyse proposée apportent des bénéfices à deux égards : d'une part les aspects théoriques aident à expliquer comment fonctionne concrètement la compréhension de la langue, d'autre part ils peuvent s'avérer utiles pour le traitement informatique de la langue – par exemple pour séparer les informations sur les structures de celles des contenus ou encore à distinguer les informations de contenus selon leur importance.

Contexte de la recherche

² "Global coherence is connectivity between the main events of the text (scripts, plans and goals) and the global relations hold independently of the local coherence relations between discourse segments..." (Samet et Schank, 1984).

³ Pour les analyses supraphrastiques, il n'est plus strictement possible de parler de niveaux de description de la langue, comme ils sont entendus dans le cadre de l'analyse des phrases dans la DFG. Il s'agit en effet d'une étude des relations entre des unités plus complexes (segments textuels), mais nous ne hiérarchisons pas ici les types d'analyses supraphrastiques, en d'autres termes, les perspectives d'analyse (relations discursives, références, thématiques, intentionalité, etc.)

Le Corpus pragois de dépendances est une source de données linguistiques comprenant environ 50 000 phrases en tchèque issues de textes journalistiques des années 90 du XX^e siècle. L'intégralité de ces textes est analysée pour les domaines suivants⁴ :

- Morphologie (lemmatisation et étiquetage morphologique)
- Syntaxe (structure de dépendances et types de relations)
- Sémantique phrastique (structure de dépendances et types de relations) y compris la structure informationnelle et les cadres valenciels
- Expressions polylexicales
- Relations de coréférence
- Anaphores associatives (soit les relations sémantiques du type partie – tout, ensemble – sous-ensemble)
- Relations discursives explicites (soit les relations sémantiques entre des clauses, des phrases ou entre de plus grandes portions de texte exprimées par des connecteurs simples ou composés, comme la relation causale dans *Půjdu domů, protože si potřebuju odpočinout.*⁵) ; y compris les structures énumératives (listes)
- Discours direct
- Caractéristiques propres au genre

Pour certaines des données, les annotations suivantes ont été ajoutées (PDiT–EDA 1.0 ; Zikánová et coll., 2018) :

- Relations discursives implicites (soit les relations sémantiques entre clauses, des phrases ou entre de plus grandes portions de texte qui ne sont pas exprimées par des connecteurs, comme la cause dans *Půjdu domů, potřebuju si odpočinout.*⁶)

Au cours des vingt dernières années, les études portant sur des corpus ont couvert les éléments fondamentaux de la grammaire ainsi que nombre de phénomènes textuels.

Une des dernières étapes pour l'annotation du Corpus pragois de dépendances était la description des relations discursives explicites et implicites. L'analyse de cette dernière phase d'annotation nous permet de déterminer quels phénomènes il reste à appréhender pour faire progresser la description des structures de textes dans une perspective globale.

Pour la description des relations discursives explicites et implicites dans le PDT, nous nous sommes inspiré de l'approche du Penn Discourse Treebank (Prasad et coll., 2008) qui décrit les relations entre des portions de textes « du bas vers le haut », et qui dès lors identifie les relations locales, typiquement exprimées par des connecteurs logiques, sans se proposer pour objectif de suivre si ces relations constituent une structure plus globale.

La recherche des relations discursives explicites est basée sur l'identification des connecteurs logiques dans le texte puis sur le balisage des portions reliées. En revanche, les relations implicites (c'est-à-dire ne recourant pas à des connecteurs) ne présentent pas d'éléments qui les caractériseraient formellement,

⁴ Prague Dependency Treebank – Consolidated 1.0 (Hajič et coll., 2020).

⁵ Traduisible par « *Je rentre chez moi parce que j'ai besoin de me reposer.* » (note du traducteur)

⁶ Traduisible par « *Je rentre chez moi, j'ai besoin de me reposer.* » (note du traducteur)

c'est pourquoi au cours de leur identification il a été nécessaire de distinguer une série de phénomènes cohésifs d'un autre type – en effet, des phrases qui ne sont pas reliées par des connecteurs peuvent s'enchaîner autrement que sous la forme de liens discursifs implicites, et s'intégrer au texte par des moyens très variés. C'est le cas par exemple de la relation entre une question et sa réponse, entre un titre et le texte qui suit, etc. L'aperçu suivant des phénomènes textuels constitue un résultat secondaire de l'analyse des relations discursives dans le corpus PDiT-EDA 1.0, dont certaines ont fait l'objet de nouvelles annotations (voir Zikánová 2021, Zikánová et coll., 2018)⁷ :

- Relations découlant de la nature spécifique des arguments (question, titre, auteur, description, métatexte, indication sur l'origine du message)
- Relations de type question - réponse
- Relations discursives basées sur une entité, avec un fort lien de coréférence (comme la présentation d'une entité rhématique et sa spécification dans la phrase suivante)
- Lacunes dans la cohérence (absence d'un lien quelconque, indication technique)
- Phénomènes de structuration globale du texte (sur des portions de texte plus larges ou autrement spécifiques)

Notre étude de la cohérence globale du texte dans le Corpus pragois de dépendances se base sur les annotations présentées ici et sur leur analyse. Notre objectif est de suivre comment le texte entier est segmenté ainsi que de comprendre comment tous les segments sont structurés entre eux et lesquels sont essentiels pour la compréhension du texte. Il est possible d'exploiter pour cette étape les informations que nous avons à notre disposition dans l'annotation des relations discursives implicites, concrètement les indications sur les phénomènes de structuration globale du texte.

Phénomènes de structuration globale du texte dans le corpus PDiT-EDA 1.0

Pour le passage à une perspective globale du texte et pour limiter les types de relations possibles, il est possible d'utiliser l'étiquette « macrostructure » se rapportant à la présence de phénomènes de structuration globale du texte à un endroit donné du document analysé. Cette étiquette a été utilisée de manière complémentaire dans le cadre de l'annotation des relations implicites pour le marquage des phénomènes qui, en raison d'autres caractéristiques et quoiqu'indubitablement cohésifs, n'ont pas été appréhendés sous le prisme de l'annotation discursive, que ce soit dans le cadre des relations sémantiques explicites ou implicites. L'étiquette « macrostructure » a été attribuée aux arguments ou aux relations entre des arguments qui entretenaient une relation au texte dans son ensemble ; il s'agissait d'endroits dans le texte où il n'était généralement pas possible de trouver des liens sémantiques explicites ou implicites, mais où la cohérence était assurée par d'autres moyens non décrits jusqu'à aujourd'hui. Il faut cependant préciser et garder à l'esprit durant l'analyse que ce groupe n'a été défini que provisoirement pour faciliter l'étude. Concrètement cela signifie qu'il n'y a eu ni contrôle des occurrences de ce type d'annotation, ni de son exhaustivité, que la notion n'a pas été progressivement précisée, etc. Pour ces raisons, nous ne traiterons pas statistiquement les données et nous contenterons de décrire les groupes de phénomènes de structuration globale qui s'y trouvent. Nous considérons comme un avantage le fait que les indications sur ces phénomènes issus des données initiales n'étaient pas fixées par la théorie : ces

⁷ Le mode d'identification de certains phénomènes a été récupéré de versions antérieures du Corpus pragois de dépendances.

indications ont ainsi pu appréhender tous les types de phénomènes concernés que nous avons pu retrouver dans les textes.

Nous insistons sur le fait que cette annotation ne comprend pas d'indication sur l'étendue des passages et que les annotations signalent, à des endroits donnés du texte, le pivot des phénomènes présentés.

Dans le matériel étudié, nous avons pu observer quatre groupes principaux de phénomènes liés à la structuration globale du texte :

1. Structuration interne d'une ligne narrative exprimée par un seul énonciateur
2. Changement de perspective dans le texte dû à l'alternance des énonciateurs
3. Structuration externe du texte (surtout des métacommentaires relatifs au fil du texte principal)
4. Règles de composition relatives au genre : insertion d'un type d'information déterminé à un endroit habituel d'un texte d'un genre donné (titres et sous-titres, indications sur le contact d'une entreprise à la fin d'un article, etc.)

Outre ces phénomènes, la division graphique, en particulier le découpage en paragraphes, joue également un rôle dans la structuration globale des textes écrits. Nous n'avons pas directement observé ces divisions dans notre analyse, quoiqu'elles puissent aussi se refléter à certains éléments charnières dans les textes que nous avons suivis durant l'analyse des relations implicites.

1. Structuration interne d'une ligne narrative

Des tournures typiques se rencontrent dans le cours du texte au sein d'un énoncé exprimé par un même énonciateur, comme des changements du thème propre, le passage à un récit concret après un prologue général (exemple 1), ou au contraire le passage d'un exemple concret au début d'un article à la présentation du contexte général, ou encore le passage d'un exemple étendu à un deuxième (exemple 2). Du point de vue linguistique, ces trois transitions sont signalées par exemple par un changement du lexique ou une structure spécifique de la première phrase de la nouvelle partie – dans l'exemple 1, il s'agit d'une phrase présentant un nouvel élément entrant en scène (*Máme maminku v takové čekárně / Nous avons notre maman dans une sorte de salle d'attente*) ; dans l'exemple 2, la phrase commence par un élément contrastivement et contextuellement lié, soit un élément qui n'a pas encore été abordé et dont on souligne qu'il va offrir un nouveau point de départ au récit (*les porte-monnaie*, contrastant avec *les poules* qui précèdent).

Exemple 1

Jen dobré srdce bezmocným nepomůže

Dík za to, že si konečně v článku Křehcí a neperspektivní (NLN 20. 8.) někdo povšiml našeho zahanbujícího vztahu a péče o staré lidi. Probouzí se sice zájem o postižené, vznikají různé nadace, ale ačkoliv stáří čeká na každého, nikdo - kromě církve - na žádnou pomoc nemyslí. Dokud ji člověk nepotřebuje pro své blízké, ani neví, jak to vypadá. Až při návštěvách vidí ty nemohoucí trosky lidí, kterých si vážil a měl je rád, a prchá. Raději na to nastavené zrcadlo nechce myslet.

Máme maminku po mrtvici v takové čekárně na domov důchodců, pokud se nedočká dřív smrti. Je to otravné. Návštěvní místností je široká chodba plná stařen v různém stavu chátrání, což ještě zdůrazňují vybledlé, seprané tepláky, odér moče v ovzduší a obtížné mouchy. ...

(PDIT-EDA 1.0)

Avoir bon cœur ne suffit pas à aider les personnes fragiles

Dieu merci, quelqu'un s'est enfin aperçu, dans l'article Fragiles et sans perspectives (NLN 20. 8.), de la honte que représente notre relation et nos soins envers les personnes âgées. L'intérêt pour les personnes handicapées s'éveille, différentes associations apparaissent, mais bien que la vieillesse attende chacun, personne, à part les Églises, ne pense à une forme d'aide. Tant qu'on n'en a pas besoin pour nos proches, on ne sait même pas à quoi cela ressemble. Ce n'est que lors des visites qu'on découvre ces personnes réduites à l'état d'épaves impuissantes, pour lesquelles on avait de l'estime et qu'on aimait, et qu'alors on fuit. On préfère ne pas penser à ce miroir tendu.

Nous avons notre maman qui a subi un AVC dans une sorte de salle d'attente pour une maison de retraite, à moins que la mort ne survienne avant. C'est terrible. La salle de visite est un large couloir rempli de vieillards à différents stades de déchéance, impression soulignée par les tenues délavées, l'odeur d'urine dans l'atmosphère et les mouches exaspérantes. ...

Exemple 2

U nás v Kocourkově

Poslední slovo

Jiří Hanák

Za normálních okolností se život v Kocourkově odvíjel standardně. Slepice kradl Franta Kejvalů. Všichni to věděli a všichni na něho byli krátcí, protože Franta Kejval měl svá lidská a občanská práva. Biřice už unavovalo za Frantou se honit. Po dopadení ho stejně vždycky museli po sepsání protokolu pustit. Vznikla z toho dokonce jakási gentlemanská dohoda. Čas od času se Franta sám dostavil na biřickou služebnu, jako že ho chytili. A zase šel.

Portmonky měla v referátu rodina Dunkových. Všichni to věděli a všichni na ně byli krátcí, protože Dunků bylo moc a každý z nich měl svá lidská a občanská práva. ...

(PDIT-EDA 1.0)

Chez nous à Kocourkov⁸

Dernière observation

Jiří Hanák

En temps normal la vie à Kocourkov suivait son cours habituel. Les poules, c'était Franta Kejval qui les volait. Tous le savaient et personne ne risquait la confrontation, car Franta Kejval avait ses droits humains et citoyens. Les gardiens étaient déjà fatigués de poursuivre Franta. Après l'avoir attrapé ils devaient de toute façon le libérer après avoir rédigé le procès-verbal. Une espèce d'accord entre gentlemen avait été conclu. De temps en temps, Franta se rendait lui-même au bureau des gardiens, comme s'ils l'avaient intercepté. Ensuite il repartait.

Les portefeuilles étaient le domaine de la famille Dunka. Tous le savaient et personne ne risquait la confrontation, car les Dunka étaient nombreux, et chacun d'entre eux avait ses droits humains et citoyens.

Les longues énumérations ou listes constituent des structures particulières dans une ligne narrative (voir exemple 3). Elles sont généralement introduites par une phrase qui les coiffe, pouvant contenir une expression hyperonymique ou cataphorique (dans l'exemple trois, il s'agit de l'expression *tyto důvody / ces motifs*), suivie d'une liste d'éléments qui peuvent être signalés graphiquement (puces) ou

⁸ « Kocourkov » désigne une commune imaginaire issue de la littérature tchèque, qui serait le théâtre d'histoires comiques ou loufoques relatives à ses habitants dont les actions sont généralement vaines, inutiles ou ratées. (note du traducteur)

linguistiquement (d'abord, ensuite...) L'énumération peut ensuite se conclure par une transition explicite vers une autre partie du texte, comme le renvoi métalinguistique vers l'énumération elle-même (z výčtu je vidět / il ressort de la liste) de l'exemple suivant.

Exemple 3

Občanský zákoník mu [vlastníkovi bytu, aut.] 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, zetě nebo snachu, své rodiče nebo sourozence,*

b) *jestliže nájemce přestal vykonávat práci, ke které je nájem služebního bytu vázán, a pronajímatel potřebuje byt pro jiného nájemce, který bude tuto práci vykonávat,*

...

h) *neužívá-li nájemce byt bez vážných důvodů.*

Z výčtu je vidět, *že vlastník bude mít dost možností k uplatnění svého práva k bytu.*

(PDIT-EDA 1.0)

Le Code civil lui permet [au propriétaire de l'appartement, note des auteurs], en vertu du droit de propriété § 711 alinéa 1, de résilier le bail de l'appartement avec l'autorisation du tribunal.

Un de ces motifs doit s'appliquer :

a) *le propriétaire a besoin de l'appartement pour lui-même, son conjoint, ses enfants, petits-enfants, gendre ou belle-fille, ses parents ou frères et sœurs,*

b) *si le locataire a cessé d'exécuter le travail auquel la location de l'appartement de service est liée et que le bailleur a besoin de l'appartement pour un autre locataire qui effectuera ce travail,*

...

h) *si le locataire n'utilise pas l'appartement sans motif sérieux.*

Il ressort de la liste *que le propriétaire aura suffisamment d'occasions d'exercer son droit sur l'appartement.*

Des descriptions d'événements individuels peuvent apparaître dans les textes narratifs, en tant que portions longues de texte séparées les unes des autres, et entre lesquelles il y a un passage temporel du passé vers le présent. Ce glissement temporel se trouve généralement exprimé au début de la deuxième portion par une nouvelle information chronologique (voir exemple 4).

Exemple 4

Poslední dny smutného humoristy

(NLN 28. 1. 1995)

Se zájmem jsem si přečetl pojednání Arno Paříka o Karlu Poláčkovi. Děkuji za sérii povídaní o Osvětimi. Jsem rád, že se konečně píše o tom, jak to skutečně bylo. 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, nejen pokud jde o jídlo, ale snad ještě více, pokud jde o kulturu.

A tak se zde vytvořila skupina mladých, kteří toužili po vzdělání. Jednou nebo dvakrát v týdnu jsme pořádali přednášky, na které jsme zvali různé osobnosti, aby nás seznámily s oborem, ve kterém měly úspěch. A tak jsme jednou zatoužili seznámit se blíže s českou literaturou. Koho jiného jsme mohli pozvat než tehdy již známého spisovatele Karla Poláčka. Byl jsem pověřen doprovodit ho

vždy na naši přednášku, která se většinou konala v některé větší ubikaci na půdě bývalých kasáren. ... (následuje 9 vět o Poláčkových přednáškách v Terezíně)

V prosinci 1943 jsem odešel s transportem do Osvětimi, a tak pro mě skončila doba, kdy jsem mohl poslouchat slova lidí, kteří již této zemi a světu něco dali. Měl jsem štěstí, že jsem vše přežil. Po návratu jsem se dověděl, že Karel Poláček šel za mnou také do Osvětimi. Jen neměl to štěstí.

Dnes, když v televizi vidíme zfilmovanou jeho nejkrásnější knížku *Bylo nás pět*, vzpomenu si vždy na jeho kulatou postavu s tichým hlasem.

(PDIT-EDA 1.0)

Derniers jours d'un humoriste triste

(NLN 28. 1. 1995)

C'est avec intérêt que j'ai lu l'étude d'Arno Pařík sur Karel Poláček. Un grand merci pour la série de récits sur Auschwitz. Je suis heureux qu'on écrive finalement sur ce sujet et comme ça s'est réellement passé. Moi aussi, j'ai rencontré Karel Poláček durant la guerre.

Je suis arrivé à Terezín à la fin de l'année 1943. J'avais alors vingt-deux ans. Sur ce camp de concentration planait une atmosphère étrange, dans laquelle la culture trouvait sa place. Nous jeunes gens étions affamés, pas seulement quant à la nourriture, mais peut-être encore plus de culture.

C'est ainsi que s'y constituait un groupe de jeunes personnes avides d'instruction. Une à deux fois par semaine, nous organisons des rencontres auxquelles nous invitons différentes personnes pour qu'elles nous fassent découvrir le domaine dans lequel elles excellaient. Un jour nous avons eu envie d'en savoir plus sur la littérature tchèque. Qui d'autre aurions-nous pu convier que l'écrivain Karel Poláček, déjà connu à l'époque ? J'ai été chargé de l'amener chaque fois à nos rencontres, qui prenaient généralement place dans un des plus grands cantonnements sur le terrain des anciennes casernes... (s'ensuivent 9 phrases sur les présentations de Poláček à Terezín)

En décembre 1943 je suis parti en convoi pour Auschwitz, et c'est ainsi qu'a pris fin pour moi l'époque où je pouvais écouter des personnes qui avaient apporté leur contribution à ce pays et au monde. J'ai eu la chance de survivre à tout ça. Après mon retour, j'ai appris que Karel Poláček m'avait suivi à Auschwitz. Mais il n'avait pas eu la même chance.

Aujourd'hui, quand nous voyons à la télévision l'adaptation filmée de son plus beau livre « *Nous étions cinq* », je repense chaque fois à sa silhouette ronde et à sa voix douce.

2. Changement de perspective dans le texte : alternance des énonciateurs

Tandis que des changements dans le cours du texte au sein d'une ligne narrative dépendent généralement d'un changement dans le propos, sans que la ligne soit interrompue, dans le deuxième type de ruptures il s'agit plutôt de changements au niveau de l'émetteur du texte : un changement d'énonciateur.

Des phénomènes inhérents au passage d'un énonciateur ou d'une source à une autre apparaissent souvent dans nos données (des textes journalistiques). Ces différents énonciateurs ou sources peuvent être par exemple l'auteur de l'article, d'autres personnes ou documents à travers des citations. Différents arrangements relatifs s'observent dans les textes entre le discours rapporté direct, indirect et le contenu rédigé par l'auteur, constituant des portions assez importantes et aux fonctions spécifiques. Si l'un des objectifs d'un article est par exemple de présenter des déclarations politiques, l'auteur de l'article présente également le contexte ou des réactions à ces déclarations. Différentes informations structurantes peuvent être distinguées dans le cadre du discours pris en charge par l'auteur, comme la clôture formelle de l'énoncé émanant d'une personne tierce ou d'une autre source (exemple 5), la

désignation de la personne ou de la source citée, la mise en relation des réponses d'autres personnes à la même question, etc. Du point de vue linguistique, ces structurations sont surtout caractérisées par des indications graphiques (guillemets, segmentation verticale du texte, utilisation d'une autre typographie), mais encore par le nom de la personne ou de la source ainsi que des verbes d'introduction du discours rapporté ou de l'information.

Exemple 5

Z dopisů čtenářů

*Jeden z atletů, nominovaných na paralympiádu do Barcelony, vrhač Pavel Volf z Fryčovic, se odmítl podrobit mimosoutěžnímu odběru pro antidopingovou kontrolu, do Barcelony neodletěl a nyní mu hrozí trest zastavení činnosti na dva roky. Pavel Volf nám napsal obsáhlý dopis, ze kterého vyjímáme: "Na čtvrtek 27. 8. jsem byl předvolán telegramem k mimosoutěžní kontrole, kterou měl provést MUDr. Chlumský z Ústavu sportovní medicíny. ... [následuje 16 vět – popis případu, aut.] Odmítl jsem být pasivním účastníkem komedie na svůj účet. Teď na sobě nechci nechat špínu. Proto vás zvu jako nestranné pozorovatele k případnému projednávání celého případu." **Potud Pavel Volf. ...***

(PDIT-EDA 1.0)

Extrait du courrier des lecteurs

*L'un des athlètes sélectionnés pour les jeux paralympiques de Barcelone, le lanceur de poids Pavel Volf, de Fryčovice, a refusé de se présenter à un contrôle antidopage hors de la compétition. Il ne s'est pas rendu à Barcelone et il risque à présent comme sanction une suspension de deux ans. Pavel Volf nous a écrit une longue lettre, dont voici un extrait : « Le jeudi 27.08, j'ai été convoqué par télégramme à un contrôle en dehors de la compétition, auquel devait procéder le docteur Chlumský du Département de médecine du sport. ... [suivent 16 phrases décrivant l'affaire, note des auteurs] J'ai refusé de rester acteur passif de cette comédie sur mon compte. À présent je veux effacer cette tache. C'est pourquoi je vous invite en tant qu'observateurs impartiaux des éventuelles négociations autour de cette affaire. » **Jusqu'ici Pavel Volf. ...***

Certains types de textes sont directement construits sur l'alternance des énonciateurs ou des sources. Dans nos données, c'est le cas notamment d'enquêtes dans lesquelles plusieurs intervenants répondent à une même question, mais aussi d'interviews, de réactions à des extraits de textes, pour peu que ces textes soient cités. L'alternance entre les énonciateurs peut se superposer à la division du texte en questions et réponses (par exemple dans une interview), mais cette superposition ne se trouve habituellement pas complète (dans un énoncé des questions rhétoriques peuvent survenir, au contraire dans une interview les répliques du rédacteur ne contiennent pas nécessairement de question – cependant le début des répliques d'un rédacteur se trouve à la limite entre deux plus grands segments de texte). De telles ruptures apparaissent très fréquemment au cours du texte (exemple 6).

Exemple 6

Klaus je major, a co vy?

Premiér byl slavnostně povýšen na majora v záloze. Jaké hodnosti dosáhli další politici představitelé?

... [3 delší odpovědi, aut.]

Jan Kalvoda, předseda ODA: *Musím říci, že jsem skončil svoji prezenční vojenskou službu jako četař a byl jsem na to hrdý. Pak jsem se stal podporučíkem a tím jsem zůstal až do dnešních dnů.*

Daniel Kroupa, místopředseda ODA: Žádnou vojenskou hodnost nemám, neboť na vojně jsem nebyl.

Libor Novák, výkonný místopředseda ODS: Pokud je mi znám poslední vývoj, tak jsem stále ještě četař absolvent. Od té doby se má vojenská hodnost nepohnula žádným směrem a mám pocit, že už se ani nepohne.

František Kozel (ODS), předseda kulturního výboru parlamentu: Po posledním vojenském cvičení jsem se dostal na tříhvězdičkovou hodnost, tedy na nadporučíka.

(PDIT-EDA 1.0)

Klaus est commandant, et vous ?

Le premier ministre a été solennellement élevé au grade de commandant disponible. Quels grades ont été attribués aux autres représentants politiques ?

- ... [3 longues réponses, note des auteurs]

Jan Kalvoda, président du parti ODA : Je dois dire que je suis fier d'avoir achevé mon service militaire en présentiel avec le grade de sergent. Par la suite je suis devenu sous-lieutenant, ce que je suis resté jusqu'à aujourd'hui.

Daniel Kroupa, vice-président d'ODA : Je ne dispose d'aucun grade militaire, puisque je n'ai pas fait mon service militaire.

Libor Novák, vice-président effectif d'ODS : À ma connaissance, je suis toujours promu sergent. Depuis, mon grade militaire n'a pas évolué et il me semble qu'il n'évoluera plus.

František Kozel (ODS), président de la commission culturelle du parlement : Après mon dernier exercice militaire, j'ai obtenu le grade de trois étoiles, c'est-à-dire celui de premier lieutenant.

3. Structuration externe du texte

Un autre type de rupture dans la structuration globale du texte est le changement au sein de l'énoncé pris en charge par un seul énonciateur, concrètement de l'auteur, lorsqu'il apporte un métacommentaire à son propre texte – par exemple en faisant référence à la structure du numéro du journal ou en énonçant les règles d'une rubrique journalistique (exemples 7 et 8). Nous distinguons ce type de la structuration interne (voir plus haut), qui se présente à l'intérieur d'un énoncé d'un même énonciateur. Dans le cas de la structuration externe, l'auteur s'écarte de sa ligne narrative, ne la structure pas elle-même, mais apporte des informations d'un autre type à sa marge.

Exemple 7

Není-li pohledávka zajištěná zástavním právem řádně a včas splněna, má věřitel možnost cenný papír prodat prostřednictvím obchodníka s cennými papíry a z výtěžku prodeje uspokojit svou pohledávku.

(Příště dokončení)

(PDIT-EDA 1.0)

Si la créance n'est pas assurée par une garantie du droit remplie correctement et en temps voulu, le créancier a la possibilité de vendre les valeurs immobilières par l'intermédiaire d'un négociant en valeurs mobilières et de satisfaire sa créance sur le produit de la vente.

("Suite et fin à paraître prochainement")

Exemple 8

Z dopisů čtenářů

V této rubrice citujeme i z dopisů, k nimž má redakce výhrady, i z těch, které obsahují diskutabilní tvrzení.

(PDIT-EDA 1.0)

Courrier des lecteurs

Dans cette rubrique nous citons également des extraits de lettres pour lesquelles la rédaction émet des réserves, ou qui contiennent des affirmations discutables.

4. Position habituelle de l'information dans un genre de texte donné

Le dernier groupe observé de phénomènes de structuration globale du texte est le positionnement de structures typiques à des endroits typiques d'un genre de texte donné, de manière à répondre aux attentes liées à un genre de texte. Dans les articles de presse, il s'agit par exemple des titres et intertitres, des informations simples et pratiques relatives au contenu de l'article qui se trouvent à sa fin (pour une critique d'exposition, il s'agit de son nom exact, de son lieu et de sa période d'ouverture, pour des informations relatives à une entreprise du contact de celle-ci, etc.) Ces segments se détachent du cours du texte, ils n'y sont liés par aucune expression particulière et sont pourtant la plupart du temps compréhensibles sur la base d'un traitement graphique éventuellement différent ou des attentes mentionnées liées au genre.

Nous pourrions ajouter à ce groupe la clôture caractéristique des textes d'information, dans laquelle s'ajoute aux actualités brutes présentées dans le texte des détails sans rapport direct avec l'actualité traitée. En apparence, ces derniers contextualisent plus largement l'actualité, pourtant leur fonction propre consiste probablement à conclure le texte de sorte qu'il ne s'achève pas simplement sur l'épuisement du contenu de l'information principale. L'article de l'exemple 9 traite des risques pris en s'assurant auprès d'une compagnie d'assurance étrangère ne disposant pas de l'autorisation d'intervenir en République tchèque, ainsi que du contrôle des compagnies d'assurance par l'État. La conclusion s'accompagne de l'indication relativement peu importante que des compagnies d'assurance étrangères fonctionnent de cette manière en République tchèque depuis longtemps. L'article de l'exemple 10 quant à lui explique que les clients privés doivent payer leur assurance maladie au début de l'année et apporte des informations détaillées sur les démarches concrètes. En conclusion apparaît une phrase, peu naturelle ou secondaire en regard du contenu de l'article, expliquant ce qu'est un « client privé ».

Exemple 9

Lákadla černých pojišťoven

Ivana Krčálová

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.

... [13 vět, aut.]

Pojišťovny, které povolení MF nemají, u nás své produkty podle zmíněného zákona nabízet nesmějí. Porušení zákona přitom může zprostředkovatele zahraniční pojišťovny přijít až na 10 milionů korun

pokuty. Ti, kteří se v této činnosti angažují, se navíc mohou dostat do rozporu s devizovým zákonem.

Nepovolené podnikání zahraničních pojišťoven v České republice není žádné novum. Značnou popularitu již v tomto směru získaly před uzákoněním vstupu zahraniční konkurence na náš pojišťovací trh rakouské a německé pojišťovny.

(PDIT-EDA 1.0)

L'attrait des assurances illégales

Ivana Krčálová

Les personnes signant un contrat auprès d'assurances étrangères n'ayant pas l'autorisation d'exercer chez nous s'exposent au risque que l'assurance ne respecte pas les clauses du contrat et qu'en cas d'expiration ou d'événement nécessitant une assurance, comme une blessure, elles ne reçoivent aucune compensation.

... [13 phrases, note des auteurs]

Les compagnies d'assurance n'ayant pas l'autorisation du Ministère des Finances ne peuvent, selon la loi mentionnée, proposer leurs produits chez nous. En violant la loi, le fournisseur d'assurances étrangères peut écopier d'une amende de 10 millions de couronnes. Ceux qui s'engageraient dans ces pratiques peuvent en outre se trouver en infraction par rapport à la loi sur les devises.

Le commerce non autorisé d'assurances étrangères en République tchèque n'est pas une nouveauté. Les compagnies d'assurances autrichiennes et allemandes ont déjà obtenu une popularité notable avant la législation sur l'entrée de la concurrence étrangère sur notre marché des assurances.

Exemple 10

Samoplátci musejí zaplatit pojišťovněm

Praha -

*Lidé, kteří si zdravotní pojištění platí sami, musejí nyní svou zdravotní pojišťovnu zpravit o svých příjmech, výdajích a zálohách na pojištění za loňský rok. Mluvčí VZP Zimová uvedla, že okresní pobočky posílají od začátku ledna svým pojištěncům-samoplátcům formuláře nazvané Přehled, kam je třeba tyto údaje uvést. Samoplátci jej pak musejí odevzdat do osmi dnů po daňovém přiznání, které je třeba podat do konce března. Podobné formuláře budou muset předložit i klienti ostatních pojišťoven. Ti, kteří je loni měnili, mají formuláře odevzdat každé z nich. Současně jsou samoplátci povinni donést pojišťovně i doklad o dni podání daňového přiznání. Zimová dále řekla, že Všeobecná v těchto dnech rozesílá další číslo svého časopisu Jistota. Samoplátci zde najdou podrobné informace o Přehledu. Dále jsou zde údaje o cestovním připojištění a nadstandardu při hospitalizaci. Budou však možná platit jen dočasně, protože navržená novela zákonů o pojišťovnách poskytování nadstandardních služeb omezuje. **Samoplátci jsou lidé samostatně výdělečně činní a ti, kteří mají vlastní nezdaněné příjmy.***

(PDIT-EDA 1.0)

Les clients privés doivent payer les compagnies d'assurance

Prague-

Les personnes qui paient elles-mêmes leur assurance maladie doivent désormais informer leur caisse maladie de leurs revenus, dépenses et avances d'assurance de l'année précédente. La porte-parole de VZP (Caisse maladie générale), Mme Zimová, a déclaré que depuis début janvier, les

succursales de district envoient à leurs assurés privés un formulaire appelé *Récapitulatif*, où ces informations doivent être fournies. Les clients privés doivent ensuite le soumettre dans les huit jours suivant la déclaration de revenus, qui doit être soumise avant la fin du mois de mars. Les clients d'autres compagnies d'assurance devront également soumettre des formulaires similaires. Ceux qui en ont changé l'année dernière doivent remettre les formulaires à chacune d'elles. Parallèlement, les clients privés sont tenus d'apporter à la compagnie d'assurance la preuve de la date de dépôt de la déclaration d'impôt. Mme Zimová a ajouté que la Caisse maladie générale publiait actuellement un autre numéro de son magazine *Jistota*. Les clients privés y trouveront des informations détaillées sur le *Récapitulatif*. Il y a aussi des informations sur l'assurance voyage et les hospitalisations de standards supérieurs. Cependant, elles ne seront valides que temporairement, car l'amendement proposé aux lois sur les assurances limite la fourniture de services de standards supérieurs. **Les clients privés sont les travailleurs indépendants et ceux qui ont leur propre revenu non imposé.**

Classement des phénomènes de structuration globale dans la description des textes

Comme il apparaît clairement de l'aperçu proposé, les phénomènes participant de structuration globale du texte sont relativement fréquents et hétéroclites, et sont liés à différents aspects de l'organisation textuelle. Dans cette partie, nous tenterons de les trier en veillant à ce qu'une classe ne mélange pas de phénomènes de nature différente, et de les mettre en rapport avec les types d'annotations existants dans la famille du Corpus pragoïs de dépendances et le sous-corpus PDiT-EDA 1.0.⁹ Nous proposerons certains d'entre eux comme nouveaux aspects d'analyse pouvant contribuer à ce que celle-ci soit la plus complète possible (voir tableau 1).

Type d'analyse	Aspects traités	Annotation de corpus
Morphologique	Forme des mots, lemmatisation	PDT-C 1.0, PDiT-EDA 1.0
Syntaxique	Fonctions syntaxiques, mots auxiliaires et la ponctuation	PDT-C 1.0, PDiT-EDA 1.0
Tectogrammaticale	Sémantique phrastique y compris l'articulation actuelle et la coréférence grammaticale	PDT-C 1.0, PDiT-EDA 1.0 1.0
Organisation du texte	Coréférence textuelle et anaphore associative	PDT-C 1.0, PDiT-EDA 1.0
	Relations discursives	PDT-C 1.0 – relations discursives explicites, PDiT-EDA 1.0 – relations discursives explicites et

⁹ Pour un aperçu théorique de l'analyse multiniveau du texte, voir Poláková 2014.

		implicites
	Interrogations indépendantes; questions et réponses	PDT-C 1.0, PDIT-EDA 1.0 Annoté : phrases interrogatives Non annoté : <ul style="list-style-type: none"> - Questions indirectes et autres formes éventuelles portant un sens interrogatif - Étendue des questions et des réponses - Attribution des réponses à leurs questions respectives
	Attribution (marquage du discours rapporté direct, indirect et de l'expression de l'auteur ; indication sur l'énonciateur / la source des discours et leur attribution à des discours concrets)	PDT-C 1.0, PDIT-EDA 1.0 Annoté : discours rapporté direct Non annoté : <ul style="list-style-type: none"> - Discours rapporté indirect et expression de l'auteur du texte - Indication des énonciateurs / des sources - Attribution des discours rapportés à leurs énonciateurs / sources
	Structuration interne d'un énoncé : changement de thème, liste, structure temporelle ; éventuellement d'autres phénomènes tels le résumé ou la présentation du contexte	PDT-C 1.0, PDIT-EDA 1.0 Annoté : structures énumératives Non annoté : <ul style="list-style-type: none"> - Changements de thème - Structure temporelle - Autres segments habituels (contexte, conclusion, etc.)
	Structuration externe du texte : métacommentaires	PDT-C 1.0, PDIT-EDA 1.0
	Segmentation typique du texte pour un genre donné : titres et intertitres, adresses, clôture	PDT-C 1.0, PDIT-EDA 1.0

	spécifique des informations, etc.	Annoté : <ul style="list-style-type: none"> - Titres et intertitres - Auteur de l'article - Lieu d'émission de l'information Non annoté : <ul style="list-style-type: none"> - Autres segments textuels
--	-----------------------------------	---

Tableau 1 : Proposition de description des structures textuelles

Dans ce tableau, nous nous sommes efforcé de donner un aperçu des différents aspects de la description du texte, en particulier de sa structuration globale. Nous sommes conscient du fait qu'il ne s'agit que d'une proposition provisoire qui peut, et qui devrait, être sujette à discussion. Tandis que certains aspects encadrent assez clairement des domaines limités (relations discursives, questions et réponses, attributions), d'autres seront probablement amenés à se préciser avec le traitement des données. Il s'agit en particulier de la structuration interne d'une ligne narrative¹⁰, de la structuration externe du texte et de la segmentation stéréotypées liées aux genres.

Le panorama proposé n'inclut pas certains aspects pourtant indubitablement présents dans les textes. L'annotation déjà existante dans le PDT (en particulier l'annotation tectogrammaticale) apporte aussi des indices pour le traitement de nombre de ces aspects, comme les étiquettes sémantiques (foncteurs et sous-foncteurs) du lieu, de la direction et du temps, exploitables par une analyse plus détaillée de la structuration spatiotemporelle du texte, le réseau sémantique entre les concepts se trouve dans une moindre mesure cerné dans le cadre de l'annotation des anaphores associatives, l'étiquetage de la racine syntaxique du discours rapporté direct peut servir lors de l'analyse des prises en charge énonciative (pour ces dernières mentions, voir Poláková et coll., 2015). Pour des raisons pratiques, durant notre analyse, nous avons laissé de côté ces phénomènes, qui comptent encore par exemple la fonction communicative, les structures d'intention, etc.

Conclusion

Nous avons analysé, dans les données du corpus PDIT-EDA 1.0, les phénomènes relatifs à la structuration globale du texte, qui avaient initialement été marqués comme produits secondaires lors de l'annotation des relations discursives implicites. Nous avons décrit ces phénomènes et les avons classés dans quatre groupes : phénomènes de structuration interne d'une ligne narrative, de changement de perspective dans le texte dû à l'alternance des énonciateurs, de structuration externe du texte et enfin ceux liés aux règles de composition relatives au genre. En guise de proposition pour l'enrichissement de l'analyse de la cohérence textuelle, nous avons ensuite tenté d'incorporer ces phénomènes dans les classifications existantes, ou en préparation, des aspects de l'analyse du texte dans le Corpus pragois de dépendances.

¹⁰ Outre les ruptures dans le texte que nous avons identifiées dans nos données, d'autres catégories pourraient être ici incluses, comme la progression thématique (Daneš, 1985).

Remerciements :

Ce travail de recherche a pu être effectué grâce au soutien du projet GAČR 20-09853S *Globální koherence českých textů a možnosti jejího korpusového zpracování* et à l'exploitation des données proposées par l'infrastructure de recherche LINDAT/CLARIAH-CZ soutenue par le MŠMT ČR (Ministère de l'éducation nationale tchèque), projet numéro LM2018101.

Bibliographie - Sitographie :

- Daneš František. Tematické posloupnosti a jevy příbuzné. In: Daneš František, Věta a text, 1985, 207–218.
- Dressler Wolfgang Ulrich, de Beaugrande Robert-Alain: *Introduction to Text Linguistics*. London – New York : Longman, 1981.
- Hajič Jan et coll. : *Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)*. Data/software, LINDAT/CLARIAH-CZ digital library, Czech Republic, <http://hdl.handle.net/11234/1-3185>, Dec 2020.
- Hajič Jan, Bejček Eduard, Hlaváčová Jaroslava, Mikulová Marie, Straka Milan, Štěpánek Jan, Štěpánková Barbora : Prague Dependency Treebank - Consolidated 1.0. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, European Language Resources Association, Marseille, France, ISBN 979-10-95546-34-4, pp. 5208-5218, 2020.
- Halliday Michael A. K., Hasan Ruqaiya. : *Cohesion in English*. English Language Series, London: Longman, 1976.
- Hoffmannová Jana : Hierarchizace textové informace. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*, 2017.
URL: https://www.czechency.org/slovník/HIERARCHIZACE_TEXTOVÉ_INFORMACE (cit.: 30. 11. 2021)
- Krčmová Marie : Horizontální členění textu. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*, 2017.
URL: https://www.czechency.org/slovník/HORIZONTÁLNÍ_ČLENĚNÍ_TEXTU (cit.: 30. 11. 2021)
- Křístek Michal : Vertikální členění textu. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*, 2017.
URL: https://www.czechency.org/slovník/VERTIKÁLNÍ_ČLENĚNÍ_TEXTU (cit.: 30. 11. 2021)
- Menzel Katrin, Lapshinova-Koltunski Ekaterina : Kontrastive Analyse deutscher und englischer Kohäsionsmittel in verschiedenen Diskurstypen. In: *tekst i diskurs – Text und Diskurs*. Zeitschrift der Abteilung für germanistische Sprachwissenschaft des Germanistischen Instituts Warschau. 247-266, 2014.
- Mírovský Jiří, Jínová Pavlína, Poláková Lucie : Does Tectogramatics Help the Annotation of Discourse?. In : *Proceedings of the 24th International Conference on Computational Linguistics, Coling 2012* Organizing Committee, India, pp. 853-862, 2012.
- Poláková Lucie : *K možností korpusového zpracování nadvětných jevů*. In : Naše řeč, Vol. 4-5/2014, Ústav pro jazyk český AV ČR, Praha, Česká Republika, ISSN 0027-8203, pp. 241-258. URL: <http://nase-rec.ujc.cas.cz/archiv.php?art=8342>

- Poláková Lucie, Mírovský Jiří, Zikánová Šárka, Hajičová Eva : *Discourse Relations and Connectives in Higher Text Structure*. In : *Dialogue and Discourse*, Vol. 12, No. 2, Linguistic Society of America, United States, ISSN 2152-9620, pp. 1-37, Jul 2021.
- Prasad Rashmi, Dinesh Nikhil, Lee Alan, Miltsakaki Eleni, Robaldo Livio, Joshi Aravind, Webber Bonnie : The Penn Discourse Treebank 2.0. In : *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrackech, Morocco, 2008.
- Samet Jerry, Schank Roger : *Coherence and connectivity*. In : *Linguistics and Philosophy*, Vol. 7, 57–82, 1984.
- Sgall Petr, Hajičová Eva, Panevová Jarmila : *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht : Reidel, Praha: Academia, 1986.
- Stede Manfred : Disambiguating Rhetorical Structure. *Research on Language and Computation* 6, 311–332 (2008). <https://doi.org/10.1007/s11168-008-9053-7>
- Stede, Manfred, Taboada Maite, Das Debopam : *Annotation Guidelines for Rhetorical Structure*, 2017. https://www.sfu.ca/~mtaboada/docs/research/RST_Annotation_Guidelines.pdf
- van Dijk, Teun : *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman, 1980.
- Zikánová Šárka : *Implicitní diskurzivní vztahy v češtině*. Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic, ISBN 9788088132127, 192 pp., 2021.
- Zikánová Šárka, Synková Pavlína, Mírovský Jiří : *Enriched Discourse Annotation of PDiT Subset 1.0 (PDiT-EDA 1.0)*. Data/software, Charles University, Prague, Czech Republic, <http://hdl.handle.net/11234/1-2906>, Dec 2018.