

Function Multiword Expressions Annotated with Discourse Relations in the Romanian Reference Treebank

Verginica Barbu Mititelu

Romanian Academy
Research Institute for Artificial Intelligence
Bucharest, Romania
vergi@racai.ro

Tudor Voicu

Tudor Vianu National High School
of Computer Science
Bucharest, Romania
tudor.c.voicu@gmail.com

Abstract

For the Romanian Reference Treebank, a general language corpus, covering several genres and annotated according to the principles of Universal Dependencies, we present here the annotation of some function words, namely multiword conjunctions, with discourse relations from the Penn Discourse Treebank version 3.0 inventory of such relations. The annotation process was manual, with two annotators for each occurrence of the conjunctions. Lexical-semantic relations of the types synonymy, polysemy can be established between the senses of such conjunctions. The discourse relations are added to the CoNLL-U file in which the treebank is represented.

Keywords: function MWE, discourse relation, Romanian Reference Treebank.

1 Introduction

One important characteristic of a text is its cohesion, i.e., the presence of linguistic cues to guide the reader into making connections between the ideas expressed therein (Halliday and Hasan, 1976). One such linguistic cue are the connectives, with conjunctions being one subtype thereof and the focus of this paper.

We identify the occurrences of conjunctions in a corpus that is already morpho-syntactically annotated, the Romanian Reference Treebank (Barbu Mititelu, 2018), and annotate them with discourse relations from an inventory already applied at a larger scale (Prasad et al., 2019), namely that from Penn Discourse Treebank (PDTB) version 3.0 annotation manual (Webber et al., 2019). The aim of our endeavour is to identify the possible senses Romanian conjunctions have, as well as to find the possible lexical devices to express these discourse relations in Romanian, particularly in this corpus. As a further step, we aim to use these annotated

occurrences of conjunctions as training material for understanding the structure of a text.

We start by presenting similar initiatives of annotating connectives with PDTB inventory of discourse relations (Section 2) and then outline the system of Romanian conjunctions (Section 3). The annotation method we adopted is described in Section 4 and the obtained results are presented in Section 5 and are then discussed in Section 6, before concluding the paper and envisaging further steps (Section 7).

2 Related Work

Prasad et al. (2019)’s work of creating the PDTB corpus annotated with discourse relations has proven seminal to a certain extent: the same inventory of relations was used for annotating a parallel corpus of TED-talks in 6 languages (English, Polish, German, Russian, European Portuguese, and Turkish) (Zeyrek et al., 2020), a corpus for Lithuanian (Oleškevičienė et al., 2023) and for Italian (Feltracco et al., 2017). Our work adds a new language to this landscape, i.e., Romanian. For it, version 2.0 of the PDTB annotation manual (Prasad et al., 2007) was used by Postolea (2018) for annotating adversative conjunctions in a set of 200 sentences extracted from EuroParl corpus (Koehn, 2005). However, this is not made available and no other such endeavour has been reported for Romanian yet.

3 Romanian Inventory of Conjunctions

Romanian conjunctions are devices for expressing either coordination or subordination. The former function both at the clause level and at the sentence level, connecting words and, respectively, clauses entering the same syntactic relation with their head: in ex. (1) the conjunction *and* connects two direct objects, and in ex. (2) it connects two clauses

functioning as direct objects. Subordinating conjunctions, however, are only clause linking devices (complementizers), linking a subordinate clause to the clause containing its syntactic head (ex. (3)). One conjunction is either subordinating or coordinating, never both.

(1)
Cumpăr mere și pere.
Buy.1SG apple.PL and pear.PL
'I buy apples and pears.'

(2)
Cumpăr ce găsec sau ce îmi
Buy.1SG what find.1SG or what CL.REFL.1SG.DAT
permit
afford.1SG
'I buy what I find or what I afford.'

(3)
Știu că mă iubești.
Know.1SG that me lov.2SG
'I know you love me.'

In Romanian linguistics, the class of conjunctions is made up of conjunctions and conjunctive locutions. As far as their structure is concerned, conjunctions are simple (e.g., *că* "that") or compound (e.g., *ca să* "so that", *fiindcă* (lit. 'being_that') "because"). As one can notice, the latter can be written either as distinct words (e.g., *ca să*) or as a single word (e.g., *fiindcă*). Conjunctive locutions are always made up of at least two (separate) words (e.g., *pentru că* 'for that' "because"). Under focus in this paper are only the conjunctive locutions, though further annotation will extend to simple and compound conjunctions as well (see Section 7).

A remark is necessary here with respect to the Romanian conjunction *să*. This is a complementizer that specialized as the subjunctive mood marker (Dindelegan, 2013). It can occur in main clauses, where it is only a subjunctive marker (ex. (4)), but it can also occur in subordinate clauses, where its status varies, depending on the presence of absence of another subordinating device (be it another conjunction or a relative pronoun or adverb): it is either (a) both a subjunctive marker and a complementizer when (i) no other subordinating device is present (ex. (5)), or (b) only a subjunctive marker when the clause is introduced by a relative pronoun or a relative adverb (ex. (6)). There are also cases when *să* is a component (the last one in linear order) of a conjunctive locution: e.g., *fără să* "without SĂ", *pentru ca să* "so as to".

(4)
Să vină secretara!
SĂ come.SUBJ secretary.SG.DEF
'Let the secretary come!'

(5)
Îmi dai voie să te ajut?
Me.DAT give.2SG permission SĂ you.ACC.SG help.1SG
'Do you let me help you?'

(6)
Nu m-am hotărât când
Not CL.ACC.1SG have.1SG decided when
/cu cine să vizitez parcul cel nou.
/ with who SĂ visit.1SG.SUBJ park.DEF the new
'I haven't decided when/who to visit the new park with.'

For the analysis of *să* in RRT, the relation *mark*¹ is always used to attach it to the verb in the subjunctive mood, irrespective of whether it is also a complementizer or only a subjunctive marker. Consequently, the relation *fixed*, used for fixed expressions that are function words, cannot be used for attaching *să* as the last component in a conjunctive locution, thus resulting into an inconsistent treatment of conjunctive locutions in Romanian UD treebanks: in those that do not contain *să* the components are linked the relation *fixed* to the first component in linear order (Figure 1 for example (7)), while in those containing *să* this component is not attached to the rest of the expression, but is treated like a separate word (Figure 2 for example (8)).

(7)
Echipajul său de opt oameni a pierit în timp
Crew.DEF his of eight people has vanished in time
ce se zbătea să salveze echipajul
what CL.REFL.3SG.ACC striving SĂ save.3SG crew.DEF
Santampa.
Santampa
'His crew of eight people also died while striving to save the Santampa crew.'

(8)
El continuase să meargă fără să
He had_continued SĂ walk without SĂ
se oprească
CL.REFL.3SG.ACC stop
'He had continued walking, without stopping.'

4 Work Methodology

The Corpus. We chose to annotate the conjunctions in the Romanian Reference Treebank (RRT) (Barbu Mititelu, 2018). The corpus contains 9,523 sentences and 218,511 tokens, distributed in several genres. It is released and distributed within Universal Dependencies² (de Marneffe et al., 2021), thus being tokenized, lemmatized and morpho-syntactically annotated according to the principles thereof.

¹In UD, *mark* is the relation used for linking a subordinating word to the head of the clause it introduces.

²<https://universaldependencies.org/>

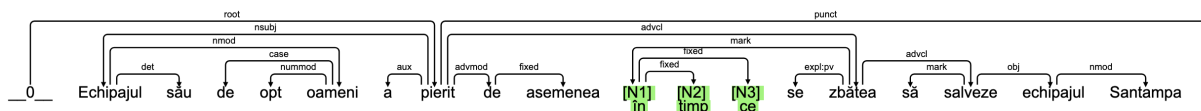


Figure 1: The annotation of a conjunctive locution with the relation *fixed*. This is the Universal Dependencies representation of ex. (7).

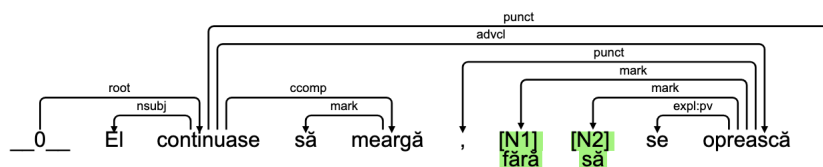


Figure 2: The annotation of conjunctive locutions containing *să*. This is the Universal Dependencies representation of ex. (8).

The Inventory of Discourse Relations. For annotation we used the discourse relations defined in the PDTB 3.0 manual. There are three levels on which relations are defined here, from more general to more specific (see Figure 3). The most refined level was always used in the annotation process.

The Inventory of Conjunctions. The aim of this work presented here is to specify the discourse relation expressed by each occurrence of conjunctive locutions in RRT. We started from a list of such conjunctions extracted from the Morphological, Orthographic and Orthoepic Romanian Dictionary (DOOM, 3rd edition)³, 71 of which were found in RRT with a total number of 479 occurrences.

Annotators. The data underwent double annotation, with a linguist and a student contributing to this task. The former was already familiar with the PDTB 3.0 manual, while for the latter this was the first experience of the kind. The degree of agreement between them is presented in Table 1. The accuracy represents the percent of annotations both annotators agreed on and it is calculated by dividing the number of cases both annotators agreed on to the total number of analyzed conjunctions. The values of the accuracy are also relevant for the possibility of automatically assigning such discourse relations, showing that the task is quite challenging for humans, thus problematic for machines.

The methodology. Each occurrence of the conjunction was annotated independently by each annotator. Only one relation was assigned, always at the lowest level possible in the PDTB hierarchy

³<https://doom.lingv.ro/>

(Figure 3). Only in a couple of cases, when the context was not considered enough to identify the sense or to distinguish between two senses, no annotation was assigned or, respectively, two senses were assigned.

5 Results

All occurrences of conjunctive locutions in RRT were annotated. In Table 2 we present the possible discourse relations they express alongside the frequency of each such relation in RRT. Although all 479 occurrences of conjunctions were annotated in the corpus, we selected only conjunctions with a frequency above 5 to show here and left out those with less occurrences.

Table 3 shows the same data, but it is more explicit in rendering the conjunctions that express relations from PDTB.

The annotation is added to the CoNLL-U format⁴ of RRT. In Figure 4 we present the current annotation of the sentence in example (9). The information about the discourse relation is added on the last column of the file: the same number is used in this last column to identify the components of the conjunctive locution, while the label of the discourse relation appears only with the first component⁵: see the highlighted lines in Figure 4, where number 1 is added in the last column of the first occurrence of components of the conjunction *pentru că*, number 2 is used for the second occurrence of the same conjunction, while the discourse

⁴<https://universaldependencies.org/format.html>

⁵It is the same system of encoding the multiword expressions used in the PARSEME treebanks (Ramisch et al., 2018).

Level-1	Level-2	Level-3
TEMPORAL	SYNCHRONOUS	–
	ASYNCHRONOUS	PRECEDENCE SUCCESSION
CONTINGENCY	CAUSE	REASON
		RESULT
	CAUSE+BELIEF	NEGRESULT
		REASON+BELIEF RESULT+BELIEF
	CAUSE+SPEECHACT	REASON+SPEECHACT RESULT+SPEECHACT
		CONDITION
	CONDITION+SPEECHACT	–
	NEGATIVE-CONDITION	ARG1-AS-NEGCOND ARG2-AS-NEGCOND
		NEGATIVE-CONDITION+SPEECHACT
	PURPOSE	ARG1-AS-GOAL ARG2-AS-GOAL
COMPARISON		CONCESSION
	CONCESSION+SPEECHACT	ARG2-AS-DENIER+SPEECHACT
	CONTRAST	–
	SIMILARITY	–
EXPANSION	CONJUNCTION	–
	DISJUNCTION	–
	EQUIVALENCE	–
	EXCEPTION	ARG1-AS-EXCPT ARG2-AS-EXCPT
		INSTANTIATION
	LEVEL-OF-DETAIL	ARG1-AS-DETAIL ARG2-AS-DETAIL
		MANNER
	SUBSTITUTION	

Figure 3: The inventory of PDTB 3.0 discourse relations.

relation expressed by each appears only for the first component, in each occurrence.

Articulațiile între șanțurile de plastic
 Joints.DEF between ditches.DEF of plastic
 se pot scurge, de asemenea - de
 CL.REFL.3SG.ACC can.3PL leak of alike - of

(9)

Conjunction	după ce “after”	pentru că “because”	în timp ce “while”	înainte de “before”	pentru ca să “in order to”	astfel încât “so that”
No.	87	54	51	35	33	27
Accuracy	89	70.4	86.3	100.0	93.9	88.9
Conjunction	pe măsură ce “as”	chiar dacă “even if”	asa încât “so that”	asa că “so that”	fără să “without”	înainte ca “before”
No.	16	13	10	9	9	9
Accuracy	75.0	100.0	100.0	100.0	77.8	100.0
Conjunction	astfel că “so that”	cu toate că “although”	de îndată ce “as soon as”	până când “until”		
No.	8	8	7	5		
Accuracy	87.5	25.0	28.6	100.0		

Table 1: The agreement between annotators for each conjunction.

obicei pentru că gunoiul sau pietrișul s-
custom for that garbage or gravel CL.REFL.3PL.ACC
au adunat între șanț și sigiliu sau pentru că
have gathered between ditch and seal or for that
însuși sigiliul s- a stricat.
itself seal.DEF CL.REFL.3SG.ACC has broken_down.
‘The joints between the plastic ditches may also leak - usually
because the garbage or the gravel have gathered between the
ditch and the seal or because the seal itself has broken down.’

6 Discussion of Results

We notice in Table 2 that the more frequent conjunctions are also more polysemous, in the sense that they are not specialized for one discourse relation; they tend to have a dominant meaning and also other meanings, more or less frequent: e.g., *pentru că* (54 total occurrences) seems specialised for Contingency.Cause.Reason (35 occurrences), but also expresses Contingency.Cause+Belief.Reason+Belief (10 occurrences), Contingency.Cause+SpeechAct.Reason+SpeechAct (4 occurrences), Contingency.Cause+SpeechAct.Result+SpeechAct (3 occurrences), Contingency.Cause.Result (2 occurrences). The most polysemous seems to be *pentru că* “because”, with 5 senses, though they are rather pragmatically distinguished than semantically: three of the relations are distinguished by the association of epistemic knowledge (belief) or a speech act. However, the more diverse polysemy is displayed by *în timp ce* “while” and *pentru ca să* “in order to”, each expressing four different discourse relations, as shown in the table.

One of the relatively frequent (35 occurrences) conjunctions which is specialized for a relation is *înainte de* “before”, which expresses Tempo-

ral.Asynchronous. Precedence.

Table 3 shows which of the relations defined in the PDTB manual are lexicalized by the analysed conjunctions. We notice again that there are prevailing ways of expressing such a relation: e.g. the relation Temporal.Synchronous tends to be expressed by *în timp ce* “while”, but there are other conjunctive locutions for it as well: *pe măsură ce* “as” *de îndată ce* “as soon as”. A rather tight competition between two conjunctions for expressing a relation is seen between *pentru ca să* “in order to” and *astfel încât* “so that” for rendering Contingency.Purpose:Arg2-as-Goal, though the former has a small advantage in our corpus.

The fact that not all PDTB relations occur in these tables does not mean they are not lexicalized in Romanian or by these conjunctions; it simply means that the analysed conjunctions in this corpus do not express them.

There are cases when one conjunction expresses two senses for the same occurrence: see the case of *după ce* which is assigned two senses for 38 occurrences: Temporal.Asynchronous.Succession and Contingency.Cause.Reason. This is expected not to be a singular case, as this is also reported for PDBT (Webber et al., 2019).

For the moment, we were not able to find any correlations between the text genre and the relation expressed by a conjunction, nor between the order of arguments and such relation.

7 Conclusions and Further Work

Our work so far ensured the familiarity of annotators with the PDTB annotation manual and the inventory of discourse relations thereof. This is

Conjunction	Total no.	No.	Sense
după ce	86	48 38	TEMPORAL:ASYNCHRONOUS:SUCCESSION TEMPORAL:ASYNCHRONOUS:SUCCESSION— CONTINGENCY:CAUSE:REASON
pentru că	54	35 10 4 3 2	CONTINGENCY:CAUSE:REASON CONTINGENCY:CAUSE+BELIEF:REASON+BELIEF CONTINGENCY:CAUSE+SPEECHACT:REASON+SPEECHACT CONTINGENCY:CAUSE+SPEECHACT:RESULT+SPEECHACT CONTINGENCY:CAUSE:RESULT
în timp ce	50	29 18 2 1	TEMPORAL:SYNCHRONOUS COMPARISON:CONTRAST COMPARISON:CONCESSION:ARG2-AS-DENIER COMPARISON:SIMILARITY
înainte de	35	35	TEMPORAL:ASYNCHRONOUS:PRECEDENCE
pentru ca să	33	30 1 1 1	CONTINGENCY:PURPOSE:ARG2-AS-GOAL CONTINGENCY:CAUSE.NEGRESULT TEMPORAL:ASYNCHRONOUS:SUCCESSION CONTINGENCY:CONDITION:ARG1-AS-COND
astfel încât	27	20 7	CONTINGENCY:PURPOSE:ARG2-AS-GOAL CONTINGENCY:CAUSE:RESULT
pe măsură ce	16	9 4 3	TEMPORAL:SYNCHRONOUS CONTINGENCY:CAUSE:REASON CONTINGENCY:CONDITION:ARG2-AS-COND
chiar dacă	13	13	COMPARISON:CONCESSION:ARG1-AS-DENIER
așa încât	9	7 2	CONTINGENCY:CAUSE:RESULT CONTINGENCY:PURPOSE:ARG2-AS-GOAL
așa că	9	7 2	CONTINGENCY:CAUSE:RESULT CONTINGENCY:PURPOSE:ARG2-AS-GOAL
fără să	9	9	EXPANSION:MANNER:ARG2-AS-MANNER
înainte ca	9	9	TEMPORAL:ASYNCHRONOUS:PRECEDENCE
astfel că	8	8	CONTINGENCY:CAUSE:RESULT
cu toate că	8	8	COMPARISON:CONCESSION:ARG1-AS-DENIER
de îndată ce	7	5 2	TEMPORAL:ASYNCHRONOUS:SUCCESSION TEMPORAL:SYNCHRONOUS
până când	5	5	TEMPORAL:ASYNCHRONOUS:PRECEDENCE
TOTAL	378		

Table 2: The PDTB discourse relations expressed by the annotated conjunctions in RRT. For the translation of the conjunctions into English see Table 1.

experience that will be further harnessed in annotating the simple and compound conjunctions in the same corpus, thus increasing the size of such data, to serve for linguistic analysis and interpretation, as well as for experiments of automatic identification of such relations in texts.

This annotation can help drawing comparisons between cross-lingually equivalent conjunctive locutions, which is of paramount importance in translation (be it manual or automatic). Corpora anno-

tated with discourse relations are used in training, tuning and testing of systems for discourse parsing, which Romanian lacks at the moment. Understanding such relations between parts of a text is vital in many NLP applications (from question-answering and summarization to automatic reasoning).

As conjunctions are not the only means of expressing discourse relations, we also envisage extending our work to the annotation of adverbs and other textual connectors and, eventually, to identify-

Sense	Total No.	No.	Connectives (count)
TEMPORAL:SYNCHRONOUS	40	29	în timp ce 9 pe măsură ce 2 de îndată ce
TEMPORAL:ASYNCHRONOUS:PRECEDENCE	49	35	înainte de 9 înainte ca 5 până când
TEMPORAL:ASYNCHRONOUS:SUCCESSION	54	48	după ce 1 pentru ca să 5 de îndată ce
TEMPORAL:ASYNCHRONOUS:SUCCESSION— CONTINGENCY:CAUSE:REASON	38	38	după ce
CONTINGENCY:CAUSE:REASON	39	35	pentru că 4 pe măsură ce
CONTINGENCY:CAUSE:RESULT	31	2	pentru că 7 astfel încât 7 așa încât 7 așa că 8 astfel că
CONTINGENCY.CAUSE.NEGRESULT	1	1	pentru ca să
CONTINGENCY:CAUSE+BELIEF:REASON+BELIEF	10	10	pentru că
CONTINGENCY:CAUSE+SPEECHACT:REASON +SPEECHACT	4	4	pentru că
CONTINGENCY:CAUSE+SPEECHACT:RESULT +SPEECHACT	3	3	pentru că
CONTINGENCY:CONDITION:ARG1-AS-COND	1	1	pentru ca să
CONTINGENCY:CONDITION:ARG2-AS-COND	3	3	pe măsură ce
CONTINGENCY:PURPOSE:ARG2-AS-GOAL	54	30	pentru ca să 20 astfel încât 2 așa încât 2 așa că
COMPARISON:CONCESSION:ARG1-AS-DENIER	21	13	chiar dacă 8 cu toate că
COMPARISON:CONCESSION:ARG2-AS-DENIER	2	2	în timp ce
COMPARISON:CONTRAST	18	18	în timp ce
COMPARISON:SIMILARITY	1	1	în timp ce
EXPANSION:MANNER:ARG2-AS-MANNER	9	9	fără să

Table 3: The conjunctions that lexicalize various PDTB discourse relations in RRT. For the translation of the conjunctions into English see Table 1.

ing cases of implicit instantiation of such relations, i.e. the relation exists in the absence of a lexicalized connector.

The Romanian inventory of such connecting devices will be added to Connective-Lex (Stede et al., 2019), a multilingual online resource of connectors and the discourse relations expressed by them.

The newly added annotation of RRT will be made freely available with the forthcoming UD

release.

References

- Verginica Barbu Mititelu. 2018. Modern syntactic analysis of Romanian. In *Clasic și modern în cercetarea filologică românească actuală*, pages 67–78, Iași, Romania.
- Gabriela Pană Dindelegan, editor. 2013. *The Grammar of Romanian*. Oxford University Press.

1	Articulațiile	articulație	NOUN	Ncfpry	Case=Acc,Nom Definite=Def Gender=Fem Number=Plur	7	nsubj	--	--
2	între	între	ADP	Spsa	AdpType=Prep Case=Acc	3	case	--	--
3	șanțurile	șanț	NOUN	Ncfpry	Case=Acc,Nom Definite=Def Gender=Fem Number=Plur	1	nmod	--	--
4	de	de	ADP	Spsa	AdpType=Prep Case=Acc	5	case	--	--
5	plastic	plastic	NOUN	Ncms-n	Definite=Ind Gender=Masc Number=Sing	3	nmod	--	--
6	se	sine	PRON	Px3-a-y-----w	Case=Acc Person=3 PronType=Prs Reflex=Yes Strength=Weak	8	expl:pv	--	--
7	pot	putea	VERB	Vmip3p	Mood=Ind Number=Plur Person=3 Tense=Pres VerbForm=Fin	0	root	--	--
8	scurge	scurge	VERB	COMMA	Tense=Pres VerbForm=Inf	7	ccomp	--	SpaceAfter=No
9	,		PUNCT	Spsa		10	punct	--	--
10	de	de	ADP	Spsa	AdpType=Prep Case=Acc	7	advmod	--	--
11	asemenea	asemenea	ADJ	Afp	Degree=Pos	10	fixed	--	--
12	-		PUNCT	DASH		22	punct	--	--
13	de	de	ADP	Spsa	AdpType=Prep Case=Acc	22	advmod	--	--
14	obicei	obicei	NOUN	Ncms-n	Definite=Ind Gender=Masc Number=Sing	13	fixed	--	--
15	pentru	pentru	ADP	Spsa	AdpType=Prep Case=Acc	22	mark	--	1:CONTINGENCY.CAUSE:REASON
16	și	și	SCONJ	Ccssp	Polarity=Pos	15	fixed	--	--
17	gunoiul	gunoi	NOUN	Ncmsry	Case=Acc,Nom Definite=Def Gender=Masc Number=Sing	22	nsubj	--	--
18	sau	sau	CCONJ	Ccssp	Polarity=Pos	19	cc	--	--
19	pietrișul	pietriș	NOUN	Ncmsry	Case=Acc,Nom Definite=Def Gender=Masc Number=Sing	17	conj	--	--
20	s-	sine	PRON	Px3-a-y-----w	Case=Acc Person=3 PronType=Prs Reflex=Yes Strength=Weak Variant=Short	22	expl:pv	--	SpaceAfter=No
21	au	avea	AUX	Va--3p	Number=Plur Person=3	22	aux	--	--
22	adunat	aduna	VERB	Vmp--sm	Gender=Masc Number=Sing VerbForm=Part	7	advcl	--	--
23	între	între	ADP	Spsa	AdpType=Prep Case=Acc	24	case	--	--
24	șanț	șanț	NOUN	Ncms-n	Definite=Ind Gender=Masc Number=Sing	22	obl	--	--
25	și	și	CCONJ	Ccssp	Polarity=Pos	26	cc	--	--
26	sigiliu	sigiliu	NOUN	Ncms-n	Definite=Ind Gender=Masc Number=Sing	24	conj	--	--
27	sau	sau	CCONJ	Ccssp	Polarity=Pos	34	cc	--	--
28	pentru	pentru	ADP	Spsa	AdpType=Prep Case=Acc	34	mark	--	2:CONTINGENCY.CAUSE:REASON
29	și	și	SCONJ	Ccssp	Polarity=Pos	38	fixed	--	--
30	insuși	insuși	DET	Dh3ms	Gender=Masc Number=Sing Person=3 PronType=Emp	31	det	--	--
31	sigiliul	sigiliu	NOUN	Ncmsry	Case=Acc,Nom Definite=Def Gender=Masc Number=Sing	34	nsubj	--	--
32	s-	sine	PRON	Px3-a-y-----w	Case=Acc Person=3 PronType=Prs Reflex=Yes Strength=Weak Variant=Short	34	expl:pv	--	SpaceAfter=No
33	a	avea	AUX	Va--3s	Number=Sing Person=3	34	aux	--	--
34	STRICAT	strica	VERB	Vmp--sm	Gender=Masc Number=Sing VerbForm=Part	22	conj	--	SpaceAfter=No
35	.		PUNCT	PERIOD		7	punct	--	--

Figure 4: Adding discourse relations in the CoNLL-U file of RRT.

- Anna Feltracco, Bernardo Magnini, and Elisabetta Jezek. 2017. Contrast-Ita Bank: A corpus for Italian Annotated with Discourse Contrast Relations. In *Proceedings of the Fourth Italian Conference on Computational Linguistics CLiC-it 2017*, pages 159–164, Rome, Italy.
- M. A. K. Halliday and R. Hasan. 1976. *Cohesion in English*. Longman.
- Philipp Koehn. 2005. *Europarl: A parallel corpus for statistical machine translation*. In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. *Universal Dependencies*. *Computational Linguistics*, 47(2):255–308.
- Giedrė Valūnaitė Oleškevičienė, Vitalija Karaciejūtė, and Dalia Gulbinskienė. 2023. *Lithuanian discourse markers and their relations in a multilingual corpus*. *Sustainable Multilingualism*, 22(1):258–272.
- Sorina Postolea. 2018. A particle filter algorithm for Bayesian wordsegmentation using annotation to identify connective meanings in a multilingual environment. romanian and english contrast markers in a parallel corpus. In *Cross-Linguistic Discourse Annotation: applications and perspectives*. *TextLink2018*, pages 107–113, Toulouse, France.
- Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo, and Bonnie Webber. 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. University of Pennsylvania.
- Rashmi Prasad, Bonnie Webber, Alan Lee, and Aravind Joshi. 2019. *Penn Discourse Treebank Version 3.0*. Linguistic Data Consortium.
- Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, and Abigail Walsh. 2018. *Edition 1.1 of the PARSEME shared task on automatic identification of verbal multiword expressions*. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 222–240, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Manfred Stede, Tatjana Scheffler, and Amalia Mendes. 2019. *Connective-lex: A web-based multilingual lexical resource for connectives*. *Discours. Revue de linguistique, psycholinguistique et informatique*, 24.
- Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The Penn Discourse Treebank 3.0 Annotation Manual. *Philadelphia, University of Pennsylvania*, 35:108.
- Deniz Zeyrek, Amália Mendes, Yulia Grishina, Murathan Kurfalı, Samuel Gibbon, and Maciej Ogródniczuk. 2020. TED multilingual discourse bank (TED-MDB): a parallel corpus annotated in the PDTB style. *Language Resources Evaluation*, 54:587–613.