Cross-Paragraph Discourse Structure in Rhetorical Structure Theory Parsing and Treebanking for Chinese and English

Siyao (Logan) Peng

Committee members: Amir Zeldes, Nathan Schneider & Nianwen Xue
Georgetown University, Washington D.C.

December 1, 2022

Table of Contents

Rhetorical Structure Theory (RST, Mann and Thompson 1988)

- Builds a discourse tree over Elementary Discourse Units (EDUs) within a document;
- Significant at the document level;
- Important for text summarization (Xu et al., 2020; Xiao et al., 2020; Huang and Kurohashi, 2021), controlled NLG (Stevens-Guille et al., 2020; Maskharashvili et al., 2021), sentiment analysis (Kraus and Feuerriegel, 2019; Huber and Carenini, 2020a), etc.

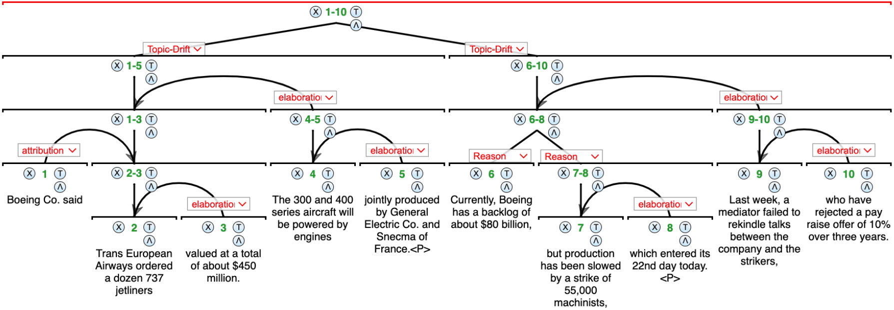


Figure 1: An example RST tree (*wsj_1153*) from RST-DT.

How can we improve RST parsing?

More data

- Complexity of RST;
- A dozen datasets in the past 20 years; still a gap for Chinese;
- Need more genres than news.

Macro structure analyses

- RST parsing is more challenging for long-distance relations – i.e., at the macro-level (Jia et al., 2018; Hou et al., 2020; Zhang et al., 2020);
- Lacks analyses associating paragraphs with the RST structure across genres, datasets, and languages.

Multilingual parsing

- No multilingual parsing experiments with Chinese;
- Lacks parsing performance analyses in multiple genres at the macro versus micro levels.

This dissertation presents:

Data: the largest and multi-genre Chinese RST corpus

- Georgetown Chinese Discourse Treebank (GCDT, Peng et al. 2022b);
- 50 documents from five different genres (total 62K+ tokens);
- Corpus statistics and annotation guidelines for Mandarin Chinese.

Analyses: paragraphs in RST

- Distribution of EDUs, sentences & paragraphs across genres/corpora;
- Paragraphs containment in RST subtrees;
- Intra- vs. inter-paragraph relations across genres/corpora.

Experiments: multigenre and multilingual RST parsing

- Benchmark monolingual and multilingual results on GCDT and GUM;
- Finetuning and automatic translation boost performance;
- Unsatisfactory results in some genres and inter-paragraph parsing.

Table of Contents

Existing discourse datasets

RST datasets

Many datasets came out in the past two decades:

- English: RST-DT (Carlson et al., 2001) and GUM (Zeldes, 2017b);
- German (Stede and Neumann, 2014), Dutch (Redeker et al., 2012), Spanish (da Cunha et al., 2011), etc.

However, there is a gap for the Chinese.

Hierarchical Chinese discourse datasets

- CDT-CDTB (Li et al., 2014a): small discourse trees (only 4.5 EDUs/tree) within paragraphs using connectives;
- MCDTB (Jiang et al., 2018): discourse trees between paragraphs;
- Sci-CDTB (Cheng and Li, 2019): Discourse Dependency for abstracts;
- Spanish-Chinese corpus (Cao et al., 2018a): translation-aligned EDUs.

However, none annotates a single-rooted tree for longer documents.

Georgetown Chinese Discourse Treebank (GCDT)

- Largest RST dataset in Mandarin Chinese (Peng et al., 2022b);
- 50 documents evenly across five genres: `academic articles`, `biographies`, `Wikipedia interviews`, `Wikinews`, and `how-to guides`;
- Uses the same relation inventory as the English GUM V8.0.0 corpus;
- Large trees for medium-to-long documents (+1K tokens/doc);
- Open-source `https://www.github.com/logan-siyao-peng/GCDT`;
- Extensive Chinese RST annotation guidelines (Peng et al., 2022a).

Genre	#Docs	#Toks	#EDUs	Source
academic	10	14,168	2,033	https://www.hanspub.org/
bio	10	13,485	2,021	https://zh.wikipedia.org/
interview	10	11,464	1,812	https://zh.wikinews.org/
news	10	11,249	1,652	https://zh.wikinews.org/
whow	10	12,539	2,199	https://zh.wikihow.com/
Total	50	62,905	9,717	

Table 1: GCDT corpus statistics.

Annotation procedures

The manual annotation process includes the following:

1. Document selection;
2. XML & metadata annotation;
3. Tokenization;
4. EDU segmentation;
5. Relation annotation.

Document selection

- Inherited four open-source genres from GUM: `biography`, `interview`, `Wikinews`, and `Wikihow`;
- Hans 汉斯 (https://www.hanspub.org/) for `academic articles`;
- A genre-balanced corpus with ten documents from each genre;
- Continuous section spans of +2K Chinese characters are extracted.

Annotation procedures

XML and metadata annotation

- Document URL, title, created/modified/accessed dates, etc;
- Gold sentence, paragraph, and section breaks.

```
<text id="gcdt_academic_dingzhen" author="Siyi Zhang, Sinuo Zhang,
Yuan Meng" dateCollected="2022-01-02" dateCreated="2021-08-12"
dateModified="XXXX-XX-XX" shortTitle="dingzhen" sourceURL="
https://www.hanspub.org/journal/PaperInformation.aspx?paperID=44511"
speakerCount="0" speakerList="none" title="
狂欢与凝视：颜值消费与田园回归——以“藏族小伙丁真爆红”事件为例" genre="academic">
狂欢与凝视：颜值消费与田园回归——以“藏族小伙丁真爆红”事件为例

<section>
1. 事件背景

2020年11月，摄影师胡波在抖音上发布了一条不到10秒的短视频，视频中的藏族小伙丁真凭借原
生态的肤色、纯真的笑容、清澈的眼神，成了网友心中的“甜野男孩”。
尤其在宣传片《丁真的世界》发布后，围绕丁真的各种话题迅速占据各大社交平台，理塘也成为当
时网友讨论度最高的旅游目的地之一。
在新媒体语境下，主流媒体积极进行流量引导，成功用一个“网红事件”赋能旅游扶贫，为贫困地区
的扶贫工作提供了新的思路。
</section>

<section>
2. “狂欢理论”概述
```

Annotation procedures

Tokenization

- Fundamental to EDU boundaries and discourse relation signals;
- Following Chinese TreeBank's (Xue, 2005) guidelines (Xia, 2000a).

EDU segmentation

- Essential to RST annotation;
- Equating EDUs with the propositional structure of clauses by using CTB's POS guidelines (Xia, 2000b);
- Notably, the first RST corpus segmented prenominal relative clauses.

```
1  狂欢 与 凝视 ： // Orgy and stare:
2  颜值 消费 与 田园 回归 // Beauty Consumption and Pastoral Return
3  —— 以 “ 藏族 小伙 丁真 爆红 ” 事件 为 例 // ——Take the incident of
   "Tibetan guy Ding Zhen's popularity" as an example
4  1. 事件 背景 // 1. Event Background
5  2020年 11月 ， 摄影师 胡波 在 抖音 上 发布 了 一 条 // In November 2020,
   photographer Hu Bo posted a post on Douyin
6  不 到 10 秒 的 // less than 10 seconds
7  短 视频 ， // short video,
8  视频 中 的 藏族 小伙 丁真 凭借 原生态 的 肤色 、 纯真 的 笑容 、 清澈 的 眼神 ，
   成 了 网友 心 中 的 “ 甜野 男孩 ” 。 // The Tibetan guy Ding Zhen in the
   video has become the "sweet boy" in the hearts of netizens with his
   original skin color, innocent smile and clear eyes.
```

EDU segmentation examples I

GCDT is the first RST corpus that segments prenominal relative clauses.

Relative clause

(1) 2 月 ， 上议院 通过 了 || **毁坏 机器 的** || **工人 必须 判处**
February , house-of-lords pass LE || destroy machine DE || worker must sentence
死刑 的 || **法案** 。
death-penalty DE || bill .

'In February, the House of Lords passed the bill announcing that workers who destroy machines must be sentenced to death.' *gcdt_bio_byron*

Moreover, I segment predicative adjectives/nouns without an overt copula.

Predicative noun

(2) (通口一叶) **原名 通口奈津 或 通口夏子** ， || 是 日本
Higuchi-Ichiyo original-name Higuchi-Najin or Higuchi-Natsuko , || COP Japan
明治 初期 主要 的 女性 小说家 。
Meiji early-period leading DE female novelist .

'Higuchi Ichiyo's original name is Higuchi Najin or Higuchi Natsuko, and he was Japan's leading female novelist in the early Meiji period.' *gcdt_bio_higuchi*

EDU segmentation examples II

Like in other RST corpora, I segment adverbial clauses from their main clauses.

Adverbial means clause

(3)

往往	由	公安	机关	||	**以**	**寻衅滋事**	**为**	**由**	||	处以	行政	拘留	。
often	by	security	department	||	taking	trouble-provoking	as	reason	||	sentence	administrative	detention	.

'(They) were often sentenced to administrative detention because of making trouble by the security departments.'

source: gcdt_academic_supervision

Adverbial conditional clause

(4)

一旦	**发生**	**疾病**	**，**	||	死亡率	近	100%	。
once	disease	occur	,	||	mortality	nearly	100%	.

'Once a disease occurs, the mortality is nearly 100%.'

source: gcdt_academic_rabies

EDU segmentation examples III

I also segment coordinated and attributional clauses following the English RST-DT and GUM guidelines.

Coordinated clause

(5)

他	**是**	欧拉	的	同	时代	人	，	||	也	**是**	密友	。
3SG.M	COP	Euler	DE	same	era	people	,	||	also	COP	close-friend	.

'He was Euler's contemporary and a close friend.'
source: gcdt_bio_bernoulli

Attributional clause

(6)

他	自己	**说**	：	||	“	在	应用文	方面	，	英文	、	德文	、	法文
3SG.M	self	say	:	||	“	in	formal-writing	aspect	,	English	,	German	,	French

没有	问题	。	”
NEG-have	problem	.	”

'He said: "as for formal writing, there is no problem with English, German, and French."'
source: gcdt_bio_chao

Annotation procedures

Relation annotation

- Two-level relation labels from GUM V8.0.0 (Zeldes, 2017a) with 15 coarse and 32 fine-grained relations.

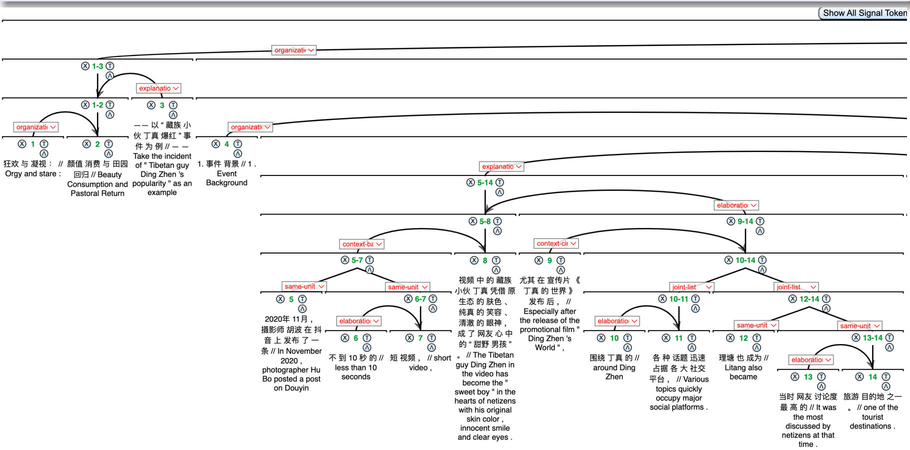


Figure 2: The beginning of the RST tree for *gcdt_academic_dingzhen*.

Inter-annotator agreement (IAA)

Procedure

- Assessed IAA on five test documents, one per genre;
- A second Chinese native-speaker linguist and RST specialist read the guidelines and conducted independent EDU segmentation;
- Measured segmentation agreement and adjudicated segmentation;
- Separately annotated discourse trees on gold EDUs.

Genre	Segmentation		Relation		
	Accuracy	Cohen's κ	Span	Nuclearity	Relation
`academic`	97.8%	0.89	80.3	59.3	49.7
`bio`	98.0%	0.91	81.5	63.9	55.6
`interview`	97.0%	0.88	83.5	62.5	54.6
`news`	98.0%	0.91	80.9	61.9	54.3
`whow`	97.3%	0.90	91.9	77.7	69.3
Overall	97.4%	0.89	84.2	66.1	57.7

Results

- Similar IAA to RST-DT's 78.7, 66.8, and 57.1 (Morey et al., 2017);
- Nuc/Rel agreements on `whow` are ~20 points higher than `academic`.

Discourse markers in GCDT

Backgrounds

- Discourse markers (DMs, e.g., connectives) are important to document structuring (Li et al., 2014b; Prasad et al., 2008; Das and Taboada, 2018);
- Das and Taboada (2018) introduced discourse signals into the RST framework, and Liu and Zeldes (2019) experimented with anchoring discourse signals in the RST-DT corpus;
- Current GCDT release does not include discourse signal annotation.

Procedure

- Extracted explicit DMs from the Chinese Discourse TreeBank (CDTB, Zhou and Xue 2015) and compared their frequencies in GCDT;
- Unfortunately, unattested DMs in CDTB are not included.

Discourse markers – results 1

Connective	Translation	CDTB all	GCDT all	GCDT aca.	bio	int.	news	whow
其中	among them	111	37	71	69	9	27	0
此外	besides that	47	23	29	15	27	27	15
二是	secondly	7	0	0	0	0	0	0
的同时	the same time of	22	5	7	0	0	9	8
与此同时	at the same time	15	0	0	0	0	0	0
如果	if	23	182	14	31	127	73	654
但是	but	18	95	86	108	91	55	131
因此	because of that	16	89	143	69	200	0	31
虽然	even though	11	68	21	92	127	73	38
因为	because of	11	118	36	85	273	127	100
所以	therefore	7	63	29	85	45	82	77

Table 2: Frequent discourse markers in CDTB and their frequencies in GCDT across genres (all counts are per 100K tokens within the specific corpus or genre).

Connectives differ vastly between CDTB and GCDT

- Enumeration signals, e.g., 其中 (among them), 此外 (besides that), 二是 (secondly), and temporal markers, e.g., 的同时 (the same time of), and 与此同时 (at the same time) are more frequent in CDTB;
- GCDT incorporates deeper logical reasoning and more argumentative connectives, e.g. 如果 (if), 但是 (but), 因此 (because of that), 虽然 (even though), 因为 (because of), 所以 (therefore), etc.

Discourse markers – results 2

Connective	Translation	CDTB all	GCDT all	GCDT aca.	bio	int.	news	whow
随着	along with	49	26	79	8	27	0	8
通过	by means of	21	66	164	62	36	27	23
从而	accordingly	8	10	36	8	0	0	0
以来	since then	32	21	21	0	27	64	0
首先	at first	7	8	14	0	0	18	8
如果	if	23	182	14	31	127	73	654
那么	in that case	7	39	7	8	36	0	138

Table 3: Frequent discourse markers in CDTB and their frequencies in GCDT across genres (all counts are per 100K tokens within the specific corpus or genre).

Genre differences in GCDT

- Event companionship signals are frequent in `academic`, e.g., 随着 (along with), 通过 (through), and 从而 (accordingly);
- Temporal markers in `news`, e.g., 以来 (since then) and 首先 (at first);
- Conditional markers in `whow`, e.g., 如果 (if) and 那么 (in that case).

Discourse relations

I examined discourse relations in three datasets:

- **GCDT**: 50 Chinese documents from 5 genres;
- **GUM-5**: 99 GUM documents from the same five genres in GCDT;
- **GUM-12**: 193 English documents from all 12 genres of GUM.

Relation Name	GCDT	GUM-5	GUM-12	Relation Name	GCDT	GUM-5	GUM-12
joint-list	22.16%	11.87%	12.90%	attribution-positive	4.35%	4.27%	4.56%
same-unit	18.74%	9.53%	8.40%	explanation-evidence	4.12%	3.88%	2.70%
elaboration-attribute	7.83%	8.66%	7.81%	adversative-contrast	3.25%	3.00%	3.35%
joint-sequence	5.10%	7.62%	6.82%	context-background	2.76%	3.72%	3.80%
joint-other	4.69%	7.91%	7.25%	context-circumstance	2.76%	3.22%	3.26%
elaboration-additional	4.35%	10.65%	9.01%	organization-preparation	2.27%	2.37%	2.28%

Table 4: Top frequent relations in GCDT, GUM-5 and GUM-12 as of V8.0.0.

Results

- Significant variations in *joint-list*, *same-unit* & *elaboration-additional*;
- The high percentage of *same-unit* in GCDT is caused by Rel-N order;
- Genre differences: e.g., more *explanation-evidence* in GUM-5 due to the more argumentative and formal nature of these five genres.

Data – interim summary

- Georgetown Chinese Discourse Treebank (GCDT) – a large open-license RST corpus for Mandarin Chinese (Peng et al., 2022b);

Data – interim summary

- Georgetown Chinese Discourse Treebank (GCDT) – a large open-license RST corpus for Mandarin Chinese (Peng et al., 2022b);
- Extensive annotation guidelines for Chinese EDU segmentation and RST relation annotation (Peng et al., 2022a);

Data – interim summary

- Georgetown Chinese Discourse Treebank (GCDT) – a large open-license RST corpus for Mandarin Chinese (Peng et al., 2022b);
- Extensive annotation guidelines for Chinese EDU segmentation and RST relation annotation (Peng et al., 2022a);
- More annotated data for longer documents and diverse genres;

Data – interim summary

- Georgetown Chinese Discourse Treebank (GCDT) – a large open-license RST corpus for Mandarin Chinese (Peng et al., 2022b);
- Extensive annotation guidelines for Chinese EDU segmentation and RST relation annotation (Peng et al., 2022a);
- More annotated data for longer documents and diverse genres;
- Comparisons of discourse markers and relations with benchmark datasets.

Table of Contents

Backgrounds on paragraphs

What are paragraphs?

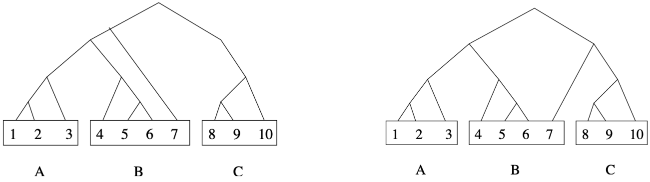
- Explicit textual breaks in a document to facilitate reading;
- Paragraph breaks signify thematic discontinuities (Ji, 2008);
- For example, proper nouns (rather than pronouns) and themes (rather than agentive subjects) are preferred at the beginning of a paragraph (Filippova and Strube, 2006; Stark, 1988).

Intra- vs. inter-paragraph relations in RST

- Inter-paragraph relations are different from intra-paragraph ones (Sporleder and Lascarides, 2004a; Wang et al., 2017a; Feng and Hirst, 2014);
- For example, ELABORATION, TOPIC-COMMENT, and TEXTUAL-ORGANIZATION relations are more common across paragraphs in RST-DT.

In RST annotation

- Assumption: discourse sub-trees are formed within the same textual chunk before crossing over to form a larger structure (Nishida and Nakayama, 2020; Cao et al., 2018b);
- Sporleder and Lascarides (2004b) categorize misalignments into two types: unambiguous (Figure 3a) and ambiguous (Figure 3b) .



(a) Misaligned but unambiguous (b) Misaligned and ambiguous

Figure 3: Two misaligned paragraph types (Sporleder and Lascarides, 2004b).

Paragraph boundaries

What is RST parsing?

RST parsing merges a sequence of gold or predicted EDUs and forms a labeled tree structure for the entire document.

Paragraph boundaries in RST parsing

- Sporleder and Lascarides (2004b): word co-occurrence is the most crucial factor for inter-paragraph relation modeling;

Paragraph boundaries

What is RST parsing?

RST parsing merges a sequence of gold or predicted EDUs and forms a labeled tree structure for the entire document.

Paragraph boundaries in RST parsing

- Sporleder and Lascarides (2004b): word co-occurrence is the most crucial factor for inter-paragraph relation modeling;
- Joty et al. (2013); Feng and Hirst (2014) are the first to parse intra-sentential versus inter-sentential relations separately;

Paragraph boundaries

What is RST parsing?

RST parsing merges a sequence of gold or predicted EDUs and forms a labeled tree structure for the entire document.

Paragraph boundaries in RST parsing

- Sporleder and Lascarides (2004b): word co-occurrence is the most crucial factor for inter-paragraph relation modeling;
- Joty et al. (2013); Feng and Hirst (2014) are the first to parse intra-sentential versus inter-sentential relations separately;
- The two-way structural distinction becomes three-way: intra-sentential, inter-sentential, and inter-paragraph in multi-stage parsing; or treating paragraph boundaries as features (Wang et al., 2017b, 2019; Nishida and Nakayama, 2020; Kobayashi et al., 2020; Koto et al., 2021; Shen et al., 2022; Yu et al., 2022);

Paragraph boundaries

What is RST parsing?

RST parsing merges a sequence of gold or predicted EDUs and forms a labeled tree structure for the entire document.

Paragraph boundaries in RST parsing

- Sporleder and Lascarides (2004b): word co-occurrence is the most crucial factor for inter-paragraph relation modeling;
- Joty et al. (2013); Feng and Hirst (2014) are the first to parse intra-sentential versus inter-sentential relations separately;
- The two-way structural distinction becomes three-way: intra-sentential, inter-sentential, and inter-paragraph in multi-stage parsing; or treating paragraph boundaries as features (Wang et al., 2017b, 2019; Nishida and Nakayama, 2020; Kobayashi et al., 2020; Koto et al., 2021; Shen et al., 2022; Yu et al., 2022);
- The SOTA RST parser (Liu et al., 2021a) is not paragraph-informed.

This paragraph analysis examines:

Three corpora

- English RST-DT (Carlson et al., 2003) and GUM (Zeldes, 2017a);
- Chinese GCDT (Peng et al., 2022b).

Three aspects

- Basic statistics for discourse structures;
- Paragraph containment analysis;
- Relation distributions within versus across paragraphs.

Corpus	#tokens	#docs	#double	Genres
RST-DT	205K	385	53	only `news`
GUM	180K	193	0	12 genres: `interviews, news, travel guides, how-to guides, academic writing, biographies, fiction, forums, conversations, political speeches, textbooks, and vlogs`
GCDT	62K	50	5	A subset of 5 genres from GUM: `interview, news, how-to guides, academic writing, and biographies`

Table 5: Three corpora: RST-DT, GUM, and GCDT.

Preprocessing: paragraph and RST alignment

Procedure

I used empty columns in the Discourse Dependency Structure (DDS, Morey et al. 2018) 10-column format to include sentence-level and paragraph-level segmentation information.

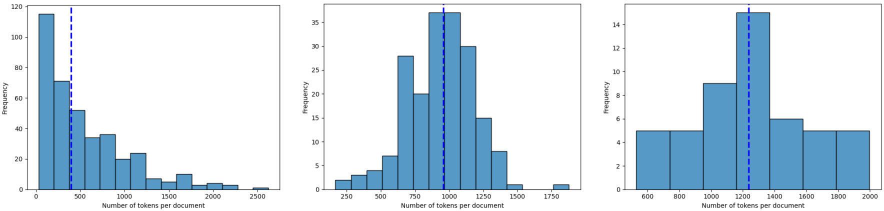
EDU	Tokens	Order	Token Span	i-th S(entence) in j-th P(aragraph)	k-th S in doc	Head	Relation		
1	Early life	4	STARTTOK=1 ENDTOK=2	HEAD-1-S-1	S-1	3	organization-heading_r	_	_
2	Born in Tianjin with ancestry in Changzhou , Jiangsu province ,	1	STARTTOK=3 ENDTOK=13	P-2-S-1	S-2	3	context-background_r	_	_
3	Chao went to the United States with a Boxer Indemnity Scholarship in 1910	0	STARTTOK=14 ENDTOK=26	P-2-S-1	S-2	0	ROOT	_	_
4	to study mathematics and physics at Cornell University ,	0	STARTTOK=27 ENDTOK=35	P-2-S-1	S-2	3	purpose-goal_r	_	_
5	where he was a classmate and lifelong friend of Hu Shih , the leader of the New Culture Movement .	0	STARTTOK=36 ENDTOK=55	P-2-S-1	S-2	4	elaboration-attribute_r	_	_

Table 6: An example of sentence and paragraph aligned Discourse Dependency Structure (DDS) for *gum_academic_chao*.

Statistics of structural segments

Observations

- GUM and GCDT annotate longer documents, whereas RST-DT documents lean to the shorter end with a median of ∼400 tokens.



(a) RST-DT (b) GUM (c) GCDT

Figure 4: #Tokens/doc in three corpora; blue dashed lines indicate the medians.

- GUM's `conversation` and `vlog` do not have paragraph boundaries;
- `Whow` has the shortest EDUs, sentences, and paragraphs in GCDT.

Paragraph containment

Definition

- *A paragraph is properly contained in an RST tree if the paragraph contains at most one outgoing edge.*
- Essentially: EDUs in a paragraph form a proper, single-rooted subtree before merging with discourse units from other paragraphs.

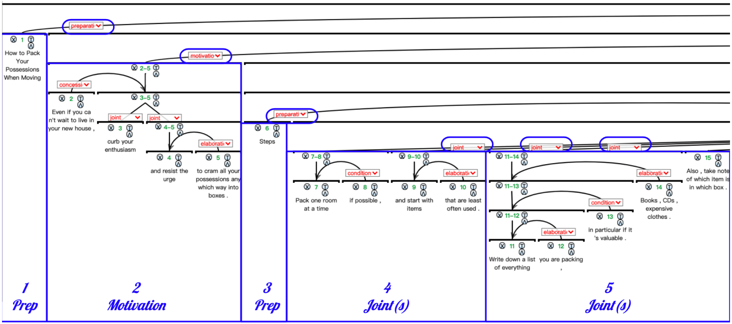
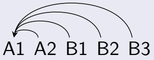


Figure 5: An example of properly-contained paragraphs in *gum_whow_packing*.

Containment violation categories by attachment positions

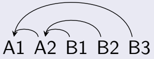
same-parent-EDU

All outgoing edges of the paragraph attach to the same external parent EDU.



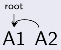
same-parent-para

All outgoing edges of the paragraph attach to the same external paragraph.



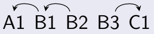
has-root

Any outgoing edge of the paragraph is the root of the document.



two-way

Some outgoing edges of the paragraph are attached to the external left parent(s); others are attached to the right.



Containment violation results

Containment violation types	RST-DT	GUM	GCDT
same-parent-EDU	**220 (5.60%)**	**109 (3.74%)**	**22 (2.03%)**
same-parent-para & diff-parent-EDU	34 (0.87%)	21 (0.72%)	6 (0.55%)
has-root	3 (0.08%)	12 (0.41%)	3 (0.28%)
two-way	55 (1.40%)	68 (2.33%)	14 (1.29%)
other violations	53 (1.35%)	58 (1.99%)	4 (0.37%)
Total number of paragraphs with outgoing violation	365 (9.29%)	268 (9.20%)	49 (4.52%)
Total number of paragraphs in the corpus	3930	2914	1083

Table 7: Types of violated paragraphs in RST-DT, GUM, and GCDT.

Results

- Over 90% of paragraphs are properly contained in all three corpora;
- *same-parent-EDU* account for +40% of the violations in all corpora.

Same-parent-EDU violation example

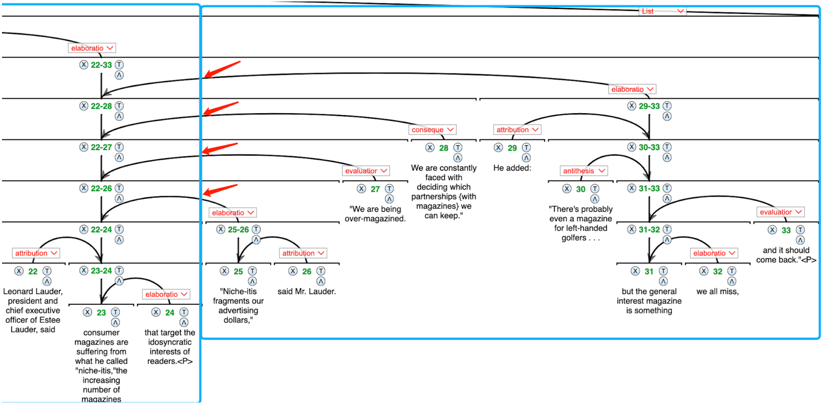


Figure 6: An example with four *same-parent-EDU* edges in *wsj_1123*.

Direct modification versus paragraph integrity

- The four outgoing edges serve different functions: *elaboration* (25-26), *evaluation* (27), *consequence* (28), and *elaboration* (29-33).

Containment violation – GUM genres

genre	same-parent-EDU	same-para & diff-parent-EDU	has-root	two-way	others	all violations	#paras
academic	3.61%	0%	1.20%	4.22%	2.41%	11.45%	166
bio	2.87%	1.08%	0%	2.87%	1.08%	7.89%	279
conversation	0%	0%	0%	0%	0%	0%	9
fiction	5.56%	1.01%	0.76%	4.29%	4.29%	15.9%	396
interview	1.50%	0.30%	0.30%	2.99%	2.40%	7.49%	334
news	5.33%	0.94%	0%	0.31%	1.57%	8.15%	319
reddit	3.57%	0.60%	0.89%	4.46%	2.38%	11.90%	336
speech	7.45%	1.24%	0%	3.11%	3.11%	14.91%	161
textbook	5.84%	1.95%	0.65%	0.00%	0.65%	9.09%	154
vlog	0%	0%	0%	0%	0%	0%	10
voyage	1.47%	0.88%	0%	0.29%	0.29%	2.94%	340
whow	3.03%	0%	0.47%	0.93%	1.40%	5.83%	429
all genres	3.72%	0.72%	0.41%	2.32%	1.98%	9.14%	2933

Table 8: Percentages of violation types per genre in GUM.

Paragraph violation analysis is incompatible with conversational genres:

- conversation and vlog do not exhibit paragraph boundaries;
- fiction, reddit, and speech have the highest paragraph containment violations due to imposed short paragraphs.

Containment violation example – GUM fiction

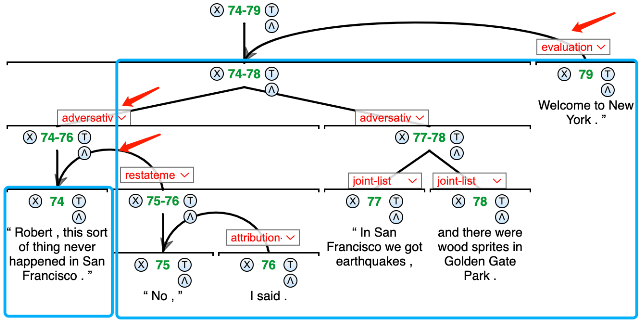


Figure 7: A violation example with three outgoing edges in *gum_fiction_pixies*.

Robert's response includes three discourse intentions toward Jenna:

1. *Restating* that this does not happen in San Francisco;
2. *Contrasting* that earthquakes and wood sprites exist in SF;
3. *Evaluating* Jenna's arrival in New York.

Containment violation – GCDT genres

- By design, GCDT focuses on written genres and longer documents and has the lowest violation rate among the three corpora;
- `Whow` creates the most violation in GCDT due to itemization.

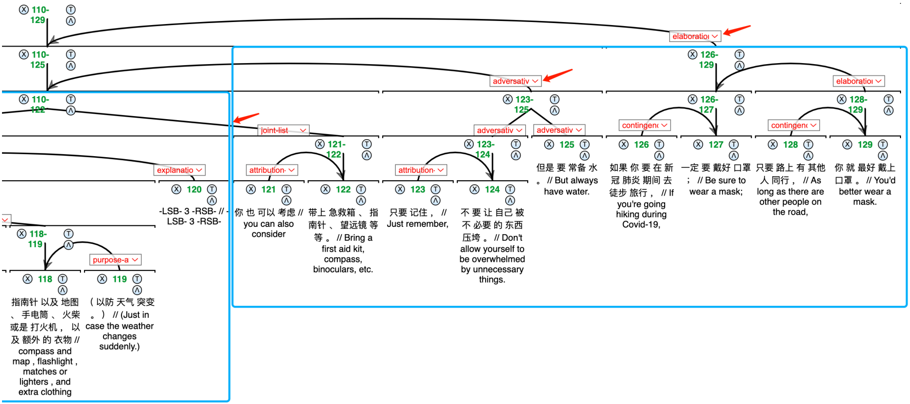


Figure 8: A violation example with three outgoing edges in *gcdt_whow_hiking*.

Intra- vs. inter-paragraph relations in RST-DT

- Many relations are limited to one scenario or another:
  - The source and content of *Attribution* occur within a paragraph;
  - *Same-Unit* links discontinuous fragments within a sentence;
  - *Topic-Change* links larger inter-paragraph units of topic shifts/drifts.
- Top relations are frequent both between and across paragraphs: *Elaboration*, *Joint*, *Contrast*, *Explanation*, *Background*, and *Cause*.

relation	nuclearity	intra-paragraph		inter-paragraph	
Attribution	mono	3068	17.6%	2	0.0%
Same-Unit	multi	1403	8.1%	1	0.0%
Topic-Change	both	6	0.0%	199	4.5%
Elaboration	mono	6213	35.7%	1689	38.4%
Joint	multi	1344	7.7%	643	14.6%
Contrast	both	893	5.1%	237	5.4%
Explanation	mono	751	4.3%	235	5.3%
Background	mono	730	4.2%	207	4.7%
Cause	both	588	3.4%	105	2.4%

Table 9: More frequent RST-DT classes in intra- & inter-paragraph scenarios.

Intra- vs. inter-paragraph relations in RST-DT

- *Enablement*, *Condition*, *Comparison* and *Manner-Means* – have much higher frequency occurring within a paragraph;
- Document-level structuring relations – *Evaluation*, *Summary*, *Topic-Comment*, *Textual-Orgainization* – appear more frequently in the inter-paragraph scenario.

relation	nuclearity	intra-paragraph		inter-paragraph	
Enablement	mono	556	3.2%	12	0.3%
Condition	both	308	1.8%	20	0.5%
Comparison	both	263	1.5%	40	0.9%
Manner-Means	mono	213	1.2%	13	0.3%
Evaluation	both	343	2.0%	256	5.8%
Summary	mono	141	0.8%	82	1.9%
Topic-Comment	both	78	0.4%	78	1.8%
Textual-Organization	multi	38	0.2%	119	2.7%

Table 10: Less frequent RST-DT classes in intra- & inter-paragraph scenarios.

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;
- *Contingency* is high in `whow` and low in `bio` for hypothetical events;

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;
- *Contingency* is high in `whow` and low in `bio` for hypothetical events;
- *Elaboration* is more frequently used in formal `academic`, `news` and `textbook` whereas *evaluation* is more prone to `fiction` and `reddit`;

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;
- *Contingency* is high in `whow` and low in `bio` for hypothetical events;
- *Elaboration* is more frequently used in formal `academic`, `news` and `textbook` whereas *evaluation* is more prone to `fiction` and `reddit`;
- *Explanation* occurs often within paragraph in argumentative `academic` but rare in factual `news`;

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;
- *Contingency* is high in `whow` and low in `bio` for hypothetical events;
- *Elaboration* is more frequently used in formal `academic`, `news` and `textbook` whereas *evaluation* is more prone to `fiction` and `reddit`;
- *Explanation* occurs often within paragraph in argumentative `academic` but rare in factual `news`;
- *Purpose* is high in intra-paragraph `whow` as steps of instructions;

Intra- vs. inter-paragraph relations in GUM

Overall

- Local relations, e.g., *contingency* & *mode* are mostly intra-paragraph;
- Higher-order organizational relations, e.g., *joint* and *organization*, occur much (∼35%) more frequently in inter-paragraph scenarios.

Genres

- *Attribution* dominates intra-paragraph `news` and inter-para `academic`;
- *Contingency* is high in `whow` and low in `bio` for hypothetical events;
- *Elaboration* is more frequently used in formal `academic`, `news` and `textbook` whereas *evaluation* is more prone to `fiction` and `reddit`;
- *Explanation* occurs often within paragraph in argumentative `academic` but rare in factual `news`;
- *Purpose* is high in intra-paragraph `whow` as steps of instructions;
- Inter-paragraph *topic* is high in `interview`'s questions-answer pairs.

Intra- vs. inter-paragraph relations in GCDT

Overall

- GCDT uses the same relation inventory and includes five written genres from GUM's twelve genres;
- *Joint* and *organization* are noticeably higher inter-paragraph;
- *Contingency*, *mode*, *purpose*, *restatement*, and *same-unit* are completely or almost absent in the inter-paragraph scenario.

Genres

- Genre distributions of GCDT echo GUM;
- For example, *attribution* is high in intra-paragraph `news` and *topic* is high in inter-paragraph `interview`.

Analyses – interim summary

- The association between paragraphs and RST in three corpora;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;
- Though $+90\%$ of paragraphs are properly contained, *same-parent-EDU* violation is common in all corpora;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;
- Though +90% of paragraphs are properly contained, *same-parent-EDU* violation is common in all corpora;
- Paragraph containment is not compatible with conversational genres in GUM;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;
- Though $+90\%$ of paragraphs are properly contained, *same-parent-EDU* violation is common in all corpora;
- Paragraph containment is not compatible with conversational genres in GUM;
- *Elaboration* and *joint* are the most frequent relations across corpora;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;
- Though $+90\%$ of paragraphs are properly contained, *same-parent-EDU* violation is common in all corpora;
- Paragraph containment is not compatible with conversational genres in GUM;
- *Elaboration* and *joint* are the most frequent relations across corpora;
- Relation distributions show genre-wise patterns, e.g., *attribution* is high in intra-paragraph `news`;

Analyses – interim summary

- The association between paragraphs and RST in three corpora;
- RST-DT documents are shorter, and GUM differ largely across genres;
- Though $+90\%$ of paragraphs are properly contained, *same-parent-EDU* violation is common in all corpora;
- Paragraph containment is not compatible with conversational genres in GUM;
- *Elaboration* and *joint* are the most frequent relations across corpora;
- Relation distributions show genre-wise patterns, e.g., *attribution* is high in intra-paragraph `news`;
- Paragraph analyses substantiate the wide applicability of the RST framework to different datasets, genres, and languages.

Table of Contents

Multilingual parsing

Backgrounds

- 10+ RST datasets have become available in many languages;

Multilingual parsing

Backgrounds

- 10+ RST datasets have become available in many languages;
- Combining multilingual training data increases training size and domain generalizability and achieves SOTA results in RST parsing;

Multilingual parsing

Backgrounds

- 10+ RST datasets have become available in many languages;
- Combining multilingual training data increases training size and domain generalizability and achieves SOTA results in RST parsing;
- Two common approaches to encode EDUs across languages:
  - Translating EDUs across languages (Cheng and Li, 2019; Liu et al., 2020, 2021b);
  - Mapping word embeddings into the same space (Braud et al., 2017; Iruskieta and Braud, 2019; Liu et al., 2020, 2021b);

Multilingual parsing

Backgrounds

- 10+ RST datasets have become available in many languages;
- Combining multilingual training data increases training size and domain generalizability and achieves SOTA results in RST parsing;
- Two common approaches to encode EDUs across languages:
  - Translating EDUs across languages (Cheng and Li, 2019; Liu et al., 2020, 2021b);
  - Mapping word embeddings into the same space (Braud et al., 2017; Iruskieta and Braud, 2019; Liu et al., 2020, 2021b);
- DMRST (Liu et al., 2021b): a SOTA multilingual RST parser with a pointer-network decoder for top-down depth-first span splitting;

Multilingual parsing

Backgrounds

- 10+ RST datasets have become available in many languages;
- Combining multilingual training data increases training size and domain generalizability and achieves SOTA results in RST parsing;
- Two common approaches to encode EDUs across languages:
  - Translating EDUs across languages (Cheng and Li, 2019; Liu et al., 2020, 2021b);
  - Mapping word embeddings into the same space (Braud et al., 2017; Iruskieta and Braud, 2019; Liu et al., 2020, 2021b);
- DMRST (Liu et al., 2021b): a SOTA multilingual RST parser with a pointer-network decoder for top-down depth-first span splitting;
- Yet, there is a lack of multilingual parsing results in different genres, intra- versus inter-paragraph structures, and in Chinese.

Experiments

Contributions

- Benchmark parsing results on the new GCDT corpus (Peng et al., 2022b) and the GUM V8.0.0 (Zeldes, 2017a);
- Multilingual training on GCDT and GUM with finetuning and automatic translation;
- Performance evaluations on different genres and intra-paragraph versus inter-paragraph discourse units.

Experiment setups

- **Parser:** Multilingual DMRST parser (Liu et al., 2021b);
- **Datasets:**
  - **GCDT**: 50 Chinese documents from 5 genres;
  - **GUM-12**: 193 English documents from 12 genres;
  - **GUM-5**: 99 GUM documents from the same five genres in GCDT;
- **Metrics:** Micro-averaged original Parseval for Span, Nuclearity (Nuc), and Relation (Rel) (Morey et al., 2017) on 15 relation classes;
- **Language Models:** different Chinese, English, and multi-lingual BERT and RoBERTa models (Devlin et al., 2019; Cui et al., 2021; Liu et al., 2019; Conneau et al., 2020).

Monolingual results

corpus	monolingual embedding	Span	Nuc	Rel
GCDT	*bert-base-chinese*	73.15±0.53	55.71±0.66	50.81±0.65
	bert-base-multilingual-cased	67.34±1.32	47.66±0.73	43.97±0.93
	hfl/chinese-roberta-wwm-ext	**75.51±0.68**	**57.08±0.81**	**51.76±0.97**
	xlm-roberta-base	74.35±0.54	54.17±1.20	50.45±1.09
GUM-12	*bert-base-cased*	60.93±0.63	47.92±0.62	40.20±0.40
	bert-base-multilingual-cased	64.47±0.50	50.69±0.32	43.25±0.35
	roberta-base	**68.59±0.58**	**55.32±0.27**	**46.29±0.46**
	xlm-roberta-base	66.12±0.59	52.58±0.52	45.06±0.45

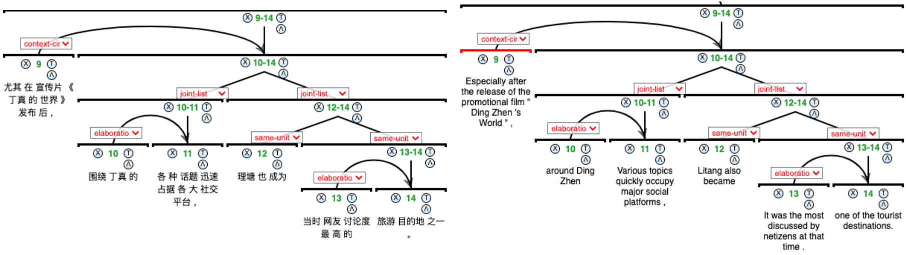
Table 11: Monolingual results on GCDT and GUM-12 with Chinese, English, and multilingual BERT and RoBERTa embeddings (averaged over five runs).

- RoBERTa outperforms BERT in both English and Chinese;
- Monolingual RoBERTa embeddings achieve SOTA.

More multilingual setups

- Using best-performing language embeddings from monolingual experiments;
  - Chinese (ZH): *hfl/chinese-roberta-wwm-ext* (Cui et al., 2021);
  - English (EN): *roberta-base* (Liu et al., 2019);
  - Multilingual (XLM): *xlm-roberta-base* (Conneau et al., 2020);
- **Finetuning**: first train on GUM+GCDT and then continue to train only on the training set of the target corpus;
- **Translation**: conduct EDU-wise automatic translations using Google Translate and merge the translated RST trees with the target dataset; thus, replacing the multilingual embeddings with monolingual ones.

An automatically translated example



(a) The original annotation.

(b) The zh→en translation.

Figure 9: The original and automatic zh→en EDU-wise translated subtrees of *gcdt_academic_dingzhen*.

Observations

- Admittedly, automatic translation is error-prone;
- Nevertheless, combining automatically translated RST trees enhances parsing performance.

Multilingual results 1

Experiment	Span	Nuc	Rel
	Test on GCDT		
Train on GCDT	**74.35±0.54**	54.17±1.20	50.45±1.09
Train on GCDT+GUM-5	74.24±0.48	56.68±0.86	52.21±0.83
Train on GCDT+GUM-12	74.33±0.49	**57.24±0.99**	**52.61±1.13**
Human Agreement	**84.27**	**66.15**	**57.77**
	Test on GUM-5		
Train on GUM-5	72.45±0.97	56.78±0.80	47.69±0.88
Train on GUM-5+GCDT	**72.56±0.71**	**60.63±0.43**	**52.57±0.77**
	Test on GUM-12		
Train on GUM-12	66.12±0.59	52.58±0.52	45.06±0.45
Train on GUM-12+GCDT	**70.32±0.37**	**57.49±0.73**	**49.14±0.34**

Table 12: Multilingual parsing results with GCDT+GUM using *xlm-roberta-base* (averaged over five runs).

- Firstly, joint training outperformed monolingual results on GCDT, GUM-5 and GUM-12.

Multilingual results 2

Experiment	Span	Nuc	Rel
Train on GCDT+GUM-5	74.24±0.48	56.68±0.86	52.21±0.83
Train on GCDT+GUM-12	**74.33±0.49**	**57.24±0.99**	**52.61±1.13**
Human Agreement	**84.27**	**66.15**	**57.77**

Table 13: Multilingual parsing results with GCDT+GUM-5 versus GCDT+GUM-12 when tested on GCDT using *xlm-roberta-base* (averaged over five runs).

- Secondly, more genres from GUM (GCDT+GUM-12) achieved slightly better performance than training only using the same genres (GCDT+GUM-5) when tested on GCDT.

Multilingual results 3

Experiment	Span	Nuc	Rel
Test on GCDT			
Train on GCDT+GUM-5	74.24±0.48	56.68±0.86	52.21±0.83
Train on GCDT+GUM-12	74.33±0.49	57.24±0.99	52.61±1.13
Train on GCDT+GUM-5 + finetune on GCDT	**76.97±0.32**	57.94±0.82	53.38±0.51
Train on GCDT+GUM-12 + finetune on GCDT	76.95±0.65	**59.40±0.64**	**55.28±0.23**
Human Agreement	**84.27**	**66.15**	**57.77**
Test on GUM-5			
Train on GUM-5+GCDT	72.56±0.71	**60.63±0.43**	**52.57±0.77**
Train on GUM-5+GCDT + finetune on GUM-5	**73.44±0.36**	59.40±0.56	50.57±0.97
Test on GUM-12			
Train on GUM-12+GCDT	**70.32±0.37**	**57.49±0.73**	**49.14±0.34**
Train on GUM-12+GCDT + finetune on GUM-12	66.00±0.24	53.13±0.22	45.47±0.42

Table 14: Multilingual parsing results with finetuning on GCDT+GUM combinations using *xlm-roberta-base* (averaged over five runs).

- Thirdly, training on the GCDT+GUM-combined training sets and continuing finetuning on the training set of the target corpus improves performance on Chinese GCDT but deteriorates on the English GUM.

Multilingual results 4

Experiment	Span	Nuc	Rel	Experiment	Span	Nuc	Rel
Train on GCDT+GUM-5 and Dev/Test on GCDT				**Train on GUM-5+GCDT and Dev/Test on GUM-5**			
multitrain w/ XLM RoBERTa	74.24	56.68	52.21	multitrain w/ XLM RoBERTa	72.56	60.63	52.57
train+finetune w/ XLM RoBERTa	76.97	57.94	53.38	train+finetune w/ XLM RoBERTa	73.44	59.40	50.57
+en→zh trans. w/ XLM RoBERTa	74.80	56.58	51.18	+zh→en trans. w/ XLM RoBERTa	72.21	60.07	52.32
+en→zh trans. w/ ZH RoBERTa	**77.66**	**59.29**	**54.66**	**+zh→en trans. w/ EN RoBERTa**	**74.73**	**62.65**	**54.32**
Train on GCDT+GUM-12 and Dev/Test on GCDT				**Train on GUM-12+GCDT and Dev/Test on GUM-12**			
multitrain w/ XLM RoBERTa	74.33	57.24	52.61	multitrain w/ XLM RoBERTa	70.32	57.49	49.14
train+finetune w/ XLM RoBERTa	77.25	**59.43**	**55.41**	train+finetune w/ XLM RoBERTa	66.00	53.13	45.47
+en→zh trans. w/ XLM RoBERTa	73.99	56.31	51.51	+zh→en trans. w/ XLM RoBERTa	70.28	57.63	49.26
+en→zh trans. w/ ZH RoBERTa	**78.11**	59.42	54.41	**+zh→en trans. w/ EN RoBERTa**	**71.41**	**59.17**	**50.63**
Human Agreement	**84.27**	**66.15**	**57.77**				

Table 15: Multilingual parsing results with finetuning and automatic translation on GCDT+GUM combinations (averaged over five runs).

- Lastly, augmenting with automatic translation and using monolingual embeddings achieved the best performance in most scenarios.

Genres in RST parsing

Backgrounds

- Domain adaption is difficult in discourse parsing (Ji et al., 2015; Atwell et al., 2021; Nishida and Matsumoto, 2022);
- Many benchmark RST parsers are trained only on English Wall Street Journal news.

Genre-wise analysis

- Three models using the Chinese RoBERTa embedding (Cui et al., 2021) are selected to assess per-genre performances:
  - Trained on GCDT;
  - Trained on GCDT + en$\rightarrow$zh translated GUM-5;
  - Trained on GCDT + en$\rightarrow$zh translated GUM-12.

Genre-wise results

Genre	Trained on GCDT			Trained w/ trans on GCDT+GUM-5			Trained w/ trans on GCDT+GUM-12			Human Agreement		
	Span	Nuc	Rel	Span	Nuc	Rel	Span	Nuc	Rel	Span	Nuc	Rel
academic	74.64	54.07	48.33	72.25	47.37	43.54	75.12	51.20	44.98	80.38	59.33	49.76
bio	72.87	54.26	52.71	74.81	57.75	53.49	77.52	59.69	55.43	81.57	63.92	55.69
interview	74.68	56.33	52.53	80.38	61.39	55.70	77.85	56.96	48.73	83.55	62.50	54.61
news	76.63	56.52	50.54	83.15	64.13	57.07	78.80	60.33	54.35	80.98	61.96	54.35
whow	77.89	57.76	54.79	80.20	66.34	62.71	80.20	65.68	61.06	91.99	77.70	69.34
Overall	75.45	55.85	52.07	77.97	59.71	55.04	78.06	59.44	53.87	84.27	66.15	57.77

Table 16: Genre-wise performances of sample models trained on GCDT, as well as translation-augmented GCDT+GUM-5 and GCDT+GUM-12 combinations using *hfl/chinese-roberta-wwm-ext*.

Results

- How-to guides (whow) performs much better than academic for both models and humans – a good human-model alignment;
- Model results on whow are the farthest from the human performance.

Paragraph structures in RST parsing

Backgrounds

- Inter-paragraph relations differ from intra-paragraph ones (Sporleder and Lascarides, 2004b; Wang et al., 2017b; Feng and Hirst, 2014);
- Previous RST parsers use sentence and paragraph boundaries either as features or in a multi-stage pipeline (Liu and Lapata, 2017; Wang et al., 2017b, 2019; Kobayashi et al., 2020);
- Yet, there is no parsing analysis on paragraph structures.

Paragraph analyses

- I examine the DMRST parser's performance on intra-paragraph versus inter-paragraph units compared to human agreements;
- GCDT double-annotated the five test documents, which enables a fair comparison between model and human performances.

Intra- versus inter-paragraph results

Structure	Precision			Recall			F1		
	Span	Nuc	Rel	Span	Nuc	Rel	Span	Nuc	Rel
	Trained on GCDT								
Intra-	66.15	50.41	45.63	78.32	59.69	54.03	71.72	54.66	49.48
Inter-	8.37	6.35	5.98	53.85	40.83	38.46	14.49	10.99	10.35
	Trained on GCDT+GUM-12 (*trans*)								
Intra-	67.53	52.25	46.46	79.96	61.87	55.01	73.22	56.66	50.37
Inter-	9.75	8.10	7.73	62.72	52.07	49.70	16.88	14.01	13.38
	Human Performance								
Intra-	71.00	53.73	45.72	83.01	63.62	54.14	76.01	58.25	49.58
Inter-	14.17	12.42	12.05	91.12	79.88	77.51	24.52	21.50	20.86

Table 17: Intra- versus inter-paragraph precision, recall and F1 on Span, Nuclearity and Relation for models trained on GCDT and GCDT+GUM-12 (*trans*) with *hfl/chinese-roberta-wwm-ext*.

Results

- The test set has 918 intra- and 169 inter-paragraph instances;
- Intra-paragraph recalls are the highest; inter-precisions the lowest;
- Models' intra-paragraph performances are close to humans;
- Substantial performance gaps in inter-paragraph precision and recall.

Experiments – interim summary

- Benchmark parsing results on monolingual GCDT and the similar English GUM V8.0.0;

Experiments – interim summary

- Benchmark parsing results on monolingual GCDT and the similar English GUM V8.0.0;
- Joint training on GCDT and GUM demonstrates that multilingual training, finetuning, and automatic EDU translation boost parser performance;

Experiments – interim summary

- Benchmark parsing results on monolingual GCDT and the similar English GUM V8.0.0;
- Joint training on GCDT and GUM demonstrates that multilingual training, finetuning, and automatic EDU translation boost parser performance;
- Monolingual RoBERTa embeddings outperform multilingual embeddings in applicable settings;

Experiments – interim summary

- Benchmark parsing results on monolingual GCDT and the similar English GUM V8.0.0;
- Joint training on GCDT and GUM demonstrates that multilingual training, finetuning, and automatic EDU translation boost parser performance;
- Monolingual RoBERTa embeddings outperform multilingual embeddings in applicable settings;
- Parsing performance varies widely between genres and deteriorates largely for inter-paragraph relations.

Table of Contents

Conclusion

This dissertation examines

How paragraph structures are presented in RST and whether SOTA RST parsers capture them properly in Chinese and English.

Conclusion

This dissertation examines

How paragraph structures are presented in RST and whether SOTA RST parsers capture them properly in Chinese and English.

Three areas of contribution

1. Created the largest and multigenre Chinese RST Corpus – GCDT, and highlighted annotation decisions for Mandarin Chinese;

Conclusion

This dissertation examines

How paragraph structures are presented in RST and whether SOTA RST parsers capture them properly in Chinese and English.

Three areas of contribution

1. Created the largest and multigenre Chinese RST Corpus – GCDT, and highlighted annotation decisions for Mandarin Chinese;
2. Examined paragraph containment and intra- versus inter-paragraph relations distributions across corpora and genres;

Conclusion

This dissertation examines

How paragraph structures are presented in RST and whether SOTA RST parsers capture them properly in Chinese and English.

Three areas of contribution

1. Created the largest and multigenre Chinese RST Corpus – GCDT, and highlighted annotation decisions for Mandarin Chinese;
2. Examined paragraph containment and intra- versus inter-paragraph relations distributions across corpora and genres;
3. Presented benchmark monolingual and multilingual parsing scores on GCDT and GUM; However, in some genres and the inter-paragraph scenario, humans' and parsers' performances are unsatisfactory.

Future work

The overall hypothesis

Higher-level discourse structures can remedy the smaller number of macro-level relations in RST training data and help RST parsing.

Future work

The overall hypothesis

Higher-level discourse structures can remedy the smaller number of macro-level relations in RST training data and help RST parsing.

Long term goals

1. To further understand the functionality of macro-level structures in medium-to-long documents;
2. To strengthen RST parsers' capability in capturing document-level structures;
3. To provide more effective representations for downstream document-level natural language understanding tasks.

Future directions I

Macro-level boundaries in human annotations

- Would human annotations differ when given the macro-level boundary information versus when such information is hidden from them?
- Analyses on annotation revisions provide more linguistic insights into macro-level RST parsing.

Topic segmentation

- Topic segmentation (Hearst, 1997) divides a document into multi-paragraph segments where each segment discusses a specific topic;
- I hope to include topic segmentation into GUM and GCDT and re-evaluate RST trees at the higher topic level.

Future directions II

Weighted evaluations

- The RST tree structure implies an imbalance sampling between macro-level and micro-level relations;
- Should we weigh macro-level discourse structures unequally from micro-level relations for IAA and parsing evaluations?
- Should we adjust weights for different downstream applications?

Silver data

- The expense of human labor motivates the growth of silver-quality datasets (Gessler et al., 2020; Huber and Carenini, 2020b);
- How can these silver data improve RST parsing besides pretraining?

Thank you for coming.
Questions? Comments?

Extra: sample parsed trees

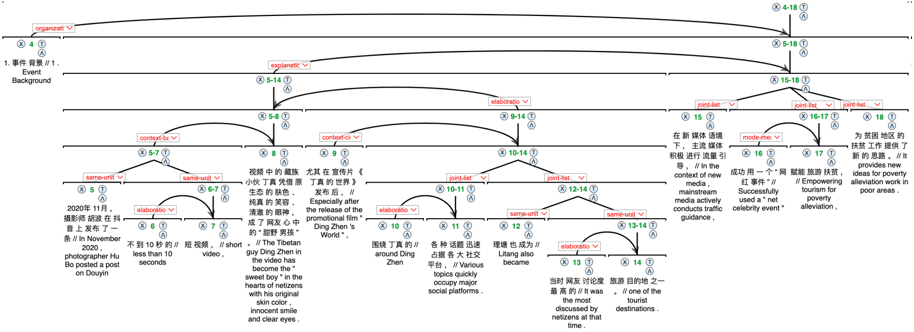


Figure 10: Gold annotation of *gcdt_academic_dingzhen*.

Extra: sample parsed trees

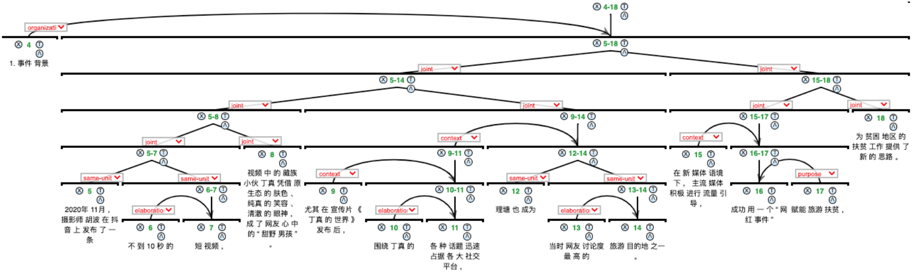


Figure 11: GCDT-trained parse on *gcdt_academic_dingzhen*.

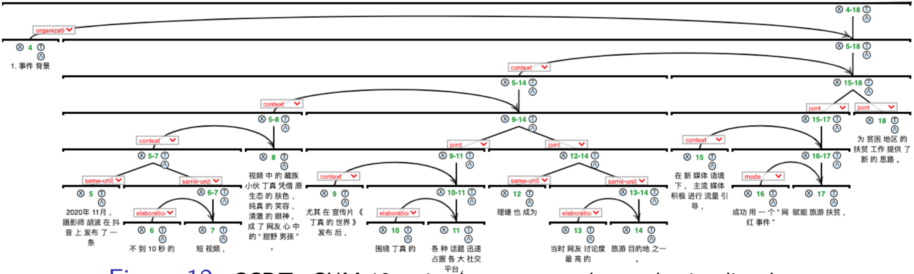


Figure 12: GCDT+GUM-12-trained parse on *gcdt_academic_dingzhen*.

References I

Katherine Atwell, Junyi Jessy Li, and Malihe Alikhani. 2021. Where Are We in Discourse Relation Recognition?. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Singapore and Online, 314–325. https://aclanthology.org/2021.sigdial-1.34

Chloé Braud, Maximin Coavoux, and Anders Søgaard. 2017. Cross-lingual RST Discourse Parsing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, Valencia, Spain, 292–304. https://aclanthology.org/E17-1028

Shuyuan Cao, Iria da Cunha, and Mikel Iruskieta. 2018a. The RST Spanish-Chinese Treebank. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 156–166. https://aclanthology.org/W18-4917

Shuyuan Cao, Iria da Cunha, and Mikel Iruskieta. 2018b. The RST Spanish-Chinese Treebank. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 156–166. https://www.aclweb.org/anthology/W18-4917

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2001. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*. https://aclanthology.org/W01-1605

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*. Springer, 85–112. https://www.aclweb.org/anthology/W01-1605

References II

Yi Cheng and Sujian Li. 2019. Zero-shot Chinese Discourse Dependency Parsing via Cross-lingual Mapping. In *Proceedings of the 1st Workshop on Discourse Structure in Neural NLG*. Association for Computational Linguistics, Tokyo, Japan, 24–29. https://doi.org/10.18653/v1/W19-8104

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 3504–3514. https://doi.org/10.1109/TASLP.2021.3124365

Iria da Cunha, Juan-Manuel Torres-Moreno, and Gerardo Sierra. 2011. On the Development of the RST Spanish Treebank. In *Proceedings of the 5th Linguistic Annotation Workshop*. Association for Computational Linguistics, Portland, Oregon, USA, 1–10. https://aclanthology.org/W11-0401

Debopam Das and Maite Taboada. 2018. RST Signalling Corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation* 52, 1 (March 2018), 149–184. https://doi.org/10.1007/s10579-017-9383-x

References III

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Vanessa Wei Feng and Graeme Hirst. 2014. A Linear-Time Bottom-Up Discourse Parser with Constraints and Post-Editing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 511–521. https://doi.org/10.3115/v1/P14-1048

Katja Filippova and Michael Strube. 2006. Using linguistically motivated features for paragraph boundary identification. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Sydney, Australia, 267–274. https://www.aclweb.org/anthology/W06-1632

Luke Gessler, Siyao Peng, Yang Liu, Yilun Zhu, Shabnam Behzad, and Amir Zeldes. 2020. AMALGUM –A Free, Balanced, Multilayer English Web Corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 5267–5275. https://aclanthology.org/2020.lrec-1.648

Marti A. Hearst. 1997. Text Tiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics* 23, 1 (1997), 33–64. https://www.aclweb.org/anthology/J97-1003

References IV

Shengluan Hou, Shuhan Zhang, and Chaoqun Fei. 2020. Rhetorical structure theory: A comprehensive review of theory, parsing methods and applications. *Expert Systems with Applications* 157 (2020), 113421. https://doi.org/10.1016/j.eswa.2020.113421

Yin Jou Huang and Sadao Kurohashi. 2021. Extractive Summarization Considering Discourse and Coreference Relations based on Heterogeneous Graph. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, Online, 3046–3052. https://doi.org/10.18653/v1/2021.eacl-main.265

Patrick Huber and Giuseppe Carenini. 2020a. From Sentiment Annotations to Sentiment Prediction through Discourse Augmentation. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), 185–197. https://doi.org/10.18653/v1/2020.coling-main.16

Patrick Huber and Giuseppe Carenini. 2020b. MEGA RST Discourse Treebanks with Structure and Nuclearity from Scalable Distant Sentiment Supervision. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 7442–7457. https://www.aclweb.org/anthology/2020.emnlp-main.603

Mikel Iruskieta and Chloé Braud. 2019. EusDisParser: improving an under-resourced discourse parser with cross-lingual data. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*. Association for Computational Linguistics, Minneapolis, MN, 62–71. https://doi.org/10.18653/v1/W19-2709

References V

Shaojun Ji. 2008. What do paragraph divisions indicate in narrative texts? *Journal of Pragmatics* 40, 10 (Oct. 2008), 1719–1730. https://doi.org/10.1016/j.pragma.2007.11.010

Yangfeng Ji, Gongbo Zhang, and Jacob Eisenstein. 2015. Closing the Gap: Domain Adaptation from Explicit to Implicit Discourse Relations. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 2219–2224. https://doi.org/10.18653/v1/D15-1264

Yanyan Jia, Yuan Ye, Yansong Feng, Yuxuan Lai, Rui Yan, and Dongyan Zhao. 2018. Modeling discourse cohesion for discourse parsing via memory network. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Melbourne, Australia, 438–443. https://doi.org/10.18653/v1/P18-2070

Feng Jiang, Sheng Xu, Xiaomin Chu, Peifeng Li, Qiaoming Zhu, and Guodong Zhou. 2018. MCDTB: A Macro-level Chinese Discourse TreeBank. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 3493–3504. https://aclanthology.org/C18-1296

Shafiq Joty, Giuseppe Carenini, Raymond Ng, and Yashar Mehdad. 2013. Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, 486–496. https://www.aclweb.org/anthology/P13-1048

Naoki Kobayashi, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. 2020. Top-Down RST Parsing Utilizing Granularity Levels in Documents. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 05 (April 2020), 8099–8106. `https://doi.org/10.1609/aaai.v34i05.6321` Number: 05.

Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2021. Top-down Discourse Parsing via Sequence Labelling. *arXiv:2102.02080 [cs]* (Feb. 2021). `http://arxiv.org/abs/2102.02080` arXiv: 2102.02080.

Mathias Kraus and Stefan Feuerriegel. 2019. Sentiment Analysis Based on Rhetorical Structure Theory: Learning Deep Neural Networks from Discourse Trees. *Expert Syst. Appl.* 118, C (mar 2019), 65–79. `https://doi.org/10.1016/j.eswa.2018.10.002`

Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014a. Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 2105–2114. `https://doi.org/10.3115/v1/D14-1224`

Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014b. Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 2105–2114. `https://doi.org/10.3115/v1/D14-1224`

References VII

Yang Liu and Mirella Lapata. 2017. Learning Contextually Informed Representations for Linear-Time Discourse Parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 1289–1298. https://doi.org/10.18653/v1/D17-1133

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [cs]* (July 2019). http://arxiv.org/abs/1907.11692 arXiv: 1907.11692.

Yang Liu and Amir Zeldes. 2019. Discourse Relations and Signaling Information: Anchoring Discourse Signals in RST-DT. (2019). https://doi.org/10.7275/VH3W-4240 Publisher: University of Massachusetts Amherst.

Zhengyuan Liu, Ke Shi, and Nancy Chen. 2020. Multilingual Neural RST Discourse Parsing. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), 6730–6738. https://doi.org/10.18653/v1/2020.coling-main.591

Zhengyuan Liu, Ke Shi, and Nancy Chen. 2021a. DMRST: A Joint Framework for Document-Level Multilingual RST Discourse Segmentation and Parsing. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*. Association for Computational Linguistics, Punta Cana, Dominican Republic and Online, 154–164. https://aclanthology.org/2021.codi-main.15

References VIII

Zhengyuan Liu, Ke Shi, and Nancy Chen. 2021b. DMRST: A Joint Framework for Document-Level Multilingual RST Discourse Segmentation and Parsing. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*. Association for Computational Linguistics, Punta Cana, Dominican Republic and Online, 154–164. `https://doi.org/10.18653/v1/2021.codi-main.15`

William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse* 8, 3 (1988), 243–281.

Aleksandre Maskharashvili, Xintong Li, Symon Jory Stevens-Guille, and Michael White. 2021. Neural Methodius Revisited: Do Discourse Relations Help with Pre-Trained Models Too? (2021), 13.

Mathieu Morey, Philippe Muller, and Nicholas Asher. 2017. How much progress have we made on RST discourse parsing? A replication study of recent results on the RST-DT. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 1319–1324. `https://doi.org/10.18653/v1/D17-1136`

Mathieu Morey, Philippe Muller, and Nicholas Asher. 2018. A Dependency Perspective on RST Discourse Parsing and Evaluation. *Computational Linguistics* 44, 2 (June 2018), 197–235. `https://doi.org/10.1162/COLI_a_00314`

Noriki Nishida and Yuji Matsumoto. 2022. Out-of-Domain Discourse Dependency Parsing via Bootstrapping: An Empirical Analysis on Its Effectiveness and Limitation. *Transactions of the Association for Computational Linguistics* 10 (02 2022), 127–144. https://doi.org/10.1162/tacl_a_00451 arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00451/1987031/tacl_a_00451.pdf

Noriki Nishida and Hideki Nakayama. 2020. Unsupervised Discourse Constituency Parsing Using Viterbi EM. *Transactions of the Association for Computational Linguistics* 8 (2020), 215–230. https://doi.org/10.1162/tacl_a_00312

Siyao Peng, Yang Janet Liu, and Amir Zeldes. 2022a. *Chinese Discourse Annotation Reference Manual*. Research Report. Georgetown University (Washington, D.C.). https://hal.archives-ouvertes.fr/hal-03821884

Siyao Peng, Yang Janet Liu, and Amir Zeldes. 2022b. GCDT: A Chinese RST Treebank for Multigenre and Multilingual Discourse Parsing. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Online only, 382–391. https://aclanthology.org/2022.aacl-short.47

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. (2008), 8.

References X

Gisela Redeker, Ildikó Berzlánovich, Nynke van der Vliet, Gosse Bouma, and Markus Egg. 2012. Multi-Layer Discourse Annotation of a Dutch Text Corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. European Language Resources Association (ELRA), Istanbul, Turkey, 2820–2825. `http://www.lrec-conf.org/proceedings/lrec2012/pdf/887_Paper.pdf`

Andrew Shen, Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2022. Easy-First Bottom-Up Discourse Parsing via Sequence Labelling. In *Proceedings of the 3rd Workshop on Computational Approaches to Discourse*. International Conference on Computational Linguistics, Gyeongju, Republic of Korea and Online, 35–41. `https://aclanthology.org/2022.codi-1.5`

Caroline Sporleder and Alex Lascarides. 2004a. Combining Hierarchical Clustering and Machine Learning to Predict High-Level Discourse Structure. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. COLING, Geneva, Switzerland, 43–49. `https://www.aclweb.org/anthology/C04-1007`

Caroline Sporleder and Alex Lascarides. 2004b. Combining Hierarchical Clustering and Machine Learning to Predict High-Level Discourse Structure. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. COLING, Geneva, Switzerland, 43–49. `https://aclanthology.org/C04-1007`

Heather A. Stark. 1988. What do paragraph markings do? *Discourse Processes* 11, 3 (July 1988), 275–303. `https://doi.org/10.1080/01638538809544704` Publisher: Routledge _eprint: https://doi.org/10.1080/01638538809544704.

References XI

Manfred Stede and Arne Neumann. 2014. Potsdam Commentary Corpus 2.0: Annotation for Discourse Research. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA), Reykjavik, Iceland, 925–929. http://www.lrec-conf.org/proceedings/lrec2014/pdf/579_Paper.pdf

Symon Stevens-Guille, Aleksandre Maskharashvili, Amy Isard, Xintong Li, and Michael White. 2020. Neural NLG for Methodius: From RST Meaning Representations to Texts. In *Proceedings of the 13th International Conference on Natural Language Generation*. Association for Computational Linguistics, Dublin, Ireland, 306–315. https://aclanthology.org/2020.inlg-1.37

Tishuang Wang, Peifeng Li, and Qiaoming Zhu. 2019. A Multi-stage Strategy for Chinese Discourse Tree Construction. In *2019 International Conference on Asian Language Processing (IALP)*. 302–307. https://doi.org/10.1109/IALP48816.2019.9037684

Yizhong Wang, Sujian Li, and Houfeng Wang. 2017a. A Two-Stage Parsing Method for Text-Level Discourse Analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Vancouver, Canada, 184–188. https://doi.org/10.18653/v1/P17-2029

Yizhong Wang, Sujian Li, and Houfeng Wang. 2017b. A Two-Stage Parsing Method for Text-Level Discourse Analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Vancouver, Canada, 184–188. https://doi.org/10.18653/v1/P17-2029

Fei Xia. 2000a. The Segmentation Guidelines for the Penn Chinese Treebank (3.0). (2000), 33.

Fei Xia. 2000b. The Part-of-Speech Guidelines for the Penn Chinese Treebank (3.0). (2000).

Wen Xiao, Patrick Huber, and Giuseppe Carenini. 2020. Do We Really Need That Many Parameters In Transformer For Extractive Summarization? Discourse Can Help !. In *Proceedings of the First Workshop on Computational Approaches to Discourse*. Association for Computational Linguistics, Online, 124–134. https://doi.org/10.18653/v1/2020.codi-1.13

Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Discourse-Aware Neural Extractive Text Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 5021–5031. https://doi.org/10.18653/v1/2020.acl-main.451

Nianwen Xue. 2005. Annotating Discourse Connectives in the Chinese Treebank. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*. Association for Computational Linguistics, Ann Arbor, Michigan, 84–91. https://www.aclweb.org/anthology/W05-0312

Nan Yu, Meishan Zhang, Guohong Fu, and Min Zhang. 2022. RST Discourse Parsing with Second-Stage EDU-Level Pre-training. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 4269–4280. https://doi.org/10.18653/v1/2022.acl-long.294

Amir Zeldes. 2017a. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation* 51, 3 (Sept. 2017), 581–612. https://doi.org/10.1007/s10579-016-9343-x

References XIII

Amir Zeldes. 2017b. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation* 51, 3 (2017), 581–612.

Longyin Zhang, Yuqing Xing, Fang Kong, Peifeng Li, and Guodong Zhou. 2020. A Top-down Neural Architecture towards Text-level Parsing of Discourse Rhetorical Structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 6386–6395. https://doi.org/10.18653/v1/2020.acl-main.569

Yuping Zhou and Nianwen Xue. 2015. The Chinese Discourse TreeBank: a Chinese corpus annotated with discourse relations. *Language Resources and Evaluation; Dordrect* 49, 2 (June 2015), 397–431. https://doi.org/10.1007/s10579-014-9290-3 Num Pages: 397-431 Place: Dordrect, Netherlands, Dordrect Publisher: Springer Nature B.V..