NLPCC2023 Shared Task7

Chinese Essay Discourse Coherence Evaluation

GuIDeline

Hongyi Wu(1), Xinshu Shen(1), Man Lan(1), Yuanbin Wu(1), Xiaopeng Bai(2),

Shaoguang Mao(3), Tao Ge(3), Yan Xia(3)

（1. College of Computer Science and Technology, East China Normal University, Shanghai 200333;

2. Department of Chinese Language and Literature, East China Normal University, Shanghai 200333;

3. Microsoft Research Asia, Beijing 100080;）

Contact: Hongyi Wu (hongyiwu@163.com)

April 6th, 2023

Table of Contents

1. Background

In the scoring of the Chinese National College Entrance Examination (NCEE) and the Senior High School Entrance Examination, essay assessment is the most time-consuming and controversial task. While existing research has focused on language factors such as characters, words, and sentences, it has not explored the relationship between discourse coherence and text quality. The logical structure and coherence within an essay are essential for evaluation, but the lack of large-scale, high-quality discourse coherence evaluation data resources has hindered the development of AI essay grading. To address this issue, the CubeNLP laboratory of East China Normal University and Microsoft have constructed a Chinese essay coherence evaluation dataset called **LEssay**, which provIDes high-quality data resources and is significant for the development of automatic essay evaluation.

2. Task Overview

2.1 Track 1. Coherence Evaluation

2.1.1 Task Description

The concept of "coherence" is a fundamental aspect of effective language expression, particularly when it comes to organizing discourse structure. Coherence essentially refers to the seamless flow between sentences and the smooth transition of paragraphs, which plays a crucial role in ensuring that written communication is clear, concise, and easy to understand.

The chapter coherence scoring task is based on some of the data from **LEssay** and aims to measure the ability of current technology to detect coherence in the discourse structure of common essay topics. This evaluation also encourages the exploration of the role of information, such as the logical structure of a composition, in assessing coherence. In summary, coherence is not only an essential element of language expression, but also a critical factor in technological advancement.

2.1.2 Task Definition

Given a mIDdle school student essay, annotators will assess its coherence on a three-level scale of excellent, moderate, and poor. A score of 2 indicates excellent coherence, 1 indicates moderate coherence, and 0 indicates incoherence.

To evaluate the coherence of an essay, two aspects should be conSIDered:

1. The smoothness of logic: The content of the essay should have logical coherence, and the paragraphs or sentences should be closely connected and unfold in a certain logical order. Factors that can impact the smoothness of logic include improper use of related words and a lack of logical relationship between contexts.

2. The reasonableness of sentence breaks: Proper sentence breaks can help better express the author's intended meaning and make the text easier to read and understand. However, improper sentence breaks can make it difficult for the reader to understand the text, or even create ambiguity.

It is crucial to understand how coherence impacts the overall quality of an essay and how these two aspects significantly affect its coherence. Here are some examples of how coherence can be affected:

错误类型	定义	文章示例	解释
逻辑不通顺	文章逻辑不通顺通常表现在以下两个方面： 1. 关联词使用不当。关联词在作文中扮演了连接不同观点、论据和段落的作用。如果使用不当，比如使用了不适当的关联词、使用关联词的位置和语境不合理等，就会导致文章中的观点、论据和段落之间缺乏必要的联系或者合理的推理过程，从而影响文章的逻辑连贯。 2. 上下文之间缺乏逻辑关系。即使上文和下文在单独表述时可以理解其意义，但是它们彼此独立，没有意义上的联系，无法形成一个意义整体。需要注意的是，上下文之间的联系不仅仅是指它们共同探讨的主题或语义相似，也可能包括顺承、对比、因果等关系，这些联系有助于让文章逻辑更为连贯。	漫步在阳光下的我们，也被阳光覆盖着，此时我们的心情随着阳光下的景物而愉悦。“更无法理解为什么会有泪水，但当我们孤独的一个人待在阳光之外时，我们何尝能回忆到那种愉悦，又流下了泪水。 阳光是什么？它可以是宇宙中天体放出了能量，也可以是一种性格特点，更可以是人们的生存寄托。	文章中关联词使用不当的句子用橙色高亮部分标出；其中使用不当的关联词用蓝色高亮部分标出。以本篇文章为例，与“但”对应的连接词为“虽然”，文章中使用“更”这个连接词，视作关联词使用不当。
		到了小学高年级阶段，书就读得越来越多了。到了初中阶段，书就读的越来多，就越来越兴趣了。未来，到上高中、大学时期。书是要反复读且要理解含义，就更会有兴趣了，下次要少玩电子产品中的游戏，用上网搜资料，这是不必要的。做事是有时间的，大部分都不是一气呵成的。要有“少年强，则国强”的人生道路。	以本段为例，当作者写到“书是要反复读且要理解含义，就更会有兴趣了”，后面应该接和前文有解释说明关系的句子，而文中示例前后文间为转折关系。

断句不合理	断句不合理会影响文章的连贯性和清晰度，表现在以下两个方面： 1. 过度使用标点符号。一些作者可能在句子中频繁使用逗号、分号等标点符号，导致句子结构不清晰，影响读者理解。此外，标点符号的使用还需要符合语法规则，否则可能会产生误解或者歧义。 2. 句子结构混乱。如果断句不当，在该断的时候没有使用句号等符号结束句子，或者过早地使用了句号，可能导致句子结构混乱，从而影响读者对文章的整体理解。此外，句子结构的合理性还需要考虑语义和语境等因素。	有一次我上学要迟到了。闷着头硬闯红灯。就在这时，一直粗糙的大手把我拉了回来，我回头看，“胖哥”正怒视冲冲地瞪着我，我刚一开口就被“胖哥”打断了。他严肃地对我说：“同学，你不能闯红灯，你不知道有多危险啊！	以本段为例，【有一次我上学要迟到了】和【闷着头硬闯红灯】之间存在因果关系，应该用逗号。
		傅雷对儿子傅聪的指导与教诲在《傅雷家书》中体现得淋漓尽致，他以直率而真诚的语言，教导傅聪要养成坚韧的品格，不忘国家之本，他以积极向上，不畏困难，踏实做人的家风，培养出在音乐界成绩骄人傅聪，即使傅聪远在国外，却仍受质朴的家风所影响，并没有因他自己成绩的优异而骄傲放纵。由此可见，好家风是传承性的。它影响着一代代人，伴随一代代人成长，也在每代人手中的火炬下传承不息。	在这段话中，逗号的使用存在一些问题。应该将其中一些逗号（灰色高亮部分）改为分号或句号，以使句子结构更加清晰。将灰色高亮部分所在的逗号改为句号后，句子被分成了若干个更简洁的子句，每个子句使用了合适的标点符号来分隔。

2.1.3 Expected Outputs

The submission should consist of a Python-based model, a development report that includes instructions for model usage, and prediction results of the testing datasets. It is crucial to ensure that the format of the model input is consistent with that of the testing datasets. To submit your work, please write the prediction results into a JSON file using the following format: [{"ID":(str),"CoherenceGrade":(int)}, ...], with the same sample order as the testing datasets.

Below is a sample input and output for your reference.

- Input Sample:

```JSON
{
    "ID": "468",
    "Title": "夏日晚风自宜人",
    "Text": [
        "那时一个相生共荣的小院，就在校篮球场旁。",
        "历经一节班会课”开不开电风扇“的争吵，自习课上终于回归了平静。同桌轻轻拍了拍我说：”看黑板上写了什么？“我顺着看去，薄暮黄昏，散在黑板上，墨绿的，托出一行精致的白色粉笔字”夏日晚风自宜人，不妨出去走走“这是什么意思？一个高个的男生突然大嚷了一声”这是让我们把电风扇开到最大档吧！“无人回应。这时，同桌邀我出去走走，我欣然应允。",
        "我们沿着沿着雪白的跑道，向前奔去。薄暮，暖黄的光穿过林立的教学楼，透过高大的篮球架，洒在了小院里，不伤花谢，不羡柳青，花柳为木，树生盎然，青叶在树梢上摇动，光影带来了最朴素真纯的生命风度，叶影婆娑间，绿起人间四月天。日暮落在那一小丛月季上，显得浓烈而又庄重，月季的影子被一旁的栅栏轻轻牵住，不时微微晃动。",
        "”好美“同桌惊讶的指着那树那花，他张开双臂，发丝被那日光携着向身后飘去。",
        "那是什么，我问着自己，是柳暗花明，是惊人月季？不，是风；风引导着我们与自然，与世间万物交融。",
        "在多少个日暮黄昏，我们倚在栏杆上说说笑笑，却意识不到那抹清凉；多少个日子，我们漫步在操场，金灿灿的枇杷果明明如耀眼宝万般晃动，我们却不自知。",
        "”夏日晚风凉，少年亦如斯“。",
        "小时候都渴望成为一棵树，长大才明白，人不能成为树，不是因为不能像树一样高大，而是缺失树的干净、坚守、温暖的灵魂。风创造了千奇百怪的大自然，铸就了一棵棵独一无二的树。树，都能发现并体现大自然的美，人却难以做到。",
```

```
        "世间紧迫地需要一双发现美的眼睛，美，就在身边，就在大自然。",
        "倾听草木的呼唤，学着做一棵向着阳光的树。"
    ]
}
```

- Output Sample:

```
JSON
{
    "ID": "468",
    "CoherenceGrade": 1
}
```

2.1.4 Training Datasets

We offer approximately 60 Chinese essays written by mIDdle school students, of which 50 can serve as training sets and 10 as verification sets. Each data sample contains the title and content of the article. Participants are also welcome to utilize data from other sources, such as manual annotation or automatic annotation using models or tools, to enhance their training experience.

2.1.5 Testing Datasets

We offer a comprehensive collection of 5,000 Chinese essays that serve as our testing datasets. These valuable resources are made available to participants in the form of a JSON file that includes key information, such as the essay's ID, title, and text content in the format of **[{"ID":"","Title":"","Text":[]} ...]**.

To ensure the highest standards of accuracy and quality, we meticulously select a portion of the data in the test set for review. This enables us to provIDe insightful feedback to participants and further refine their method.

2.1.6 Evaluation Metrics

This task employs precision (P), recall (R), and macro F1-score (F1) to evaluate the effectiveness of coherence IDentification. Precision is calculated by divIDing the number of correctly IDentified coherence types by the total number of IDentified coherence types. Recall is calculated by divIDing the number of correctly IDentified coherence types by the total number of coherence types as labeled. F1-score is calculated by using the following formula: (2*precision*recall)/(precision+recall).

2.2 Track 2. text Topic Extraction

2.2.1 Task Description

The coherence of discourse in an article is premised on the rational layout of its content. Whether it is in Chinese, English, or any other language, writing requires a central IDea, and each paragraph is usually composed of a topic sentence that represents the central IDea of the paragraph, as well as some developing sentences used to explain, describe, or argue the topic. The topic sentence of a paragraph is crucial and plays a guIDing role in connecting all sentences in the paragraph. Writing an excellent topic sentence not only makes the article well-structured but also effectively explains the main theme of the article. Therefore, extracting the topic sentence is of vital importance in evaluating the quality of an article.

2.2.2 Task Definition

Given a mIDdle school student essay, annotators need to IDentify the topic sentence for each paragraph and one main topic sentence for the whole essay.

2.2.3 Expected Outputs

The submission should consist of a Python-based model, a development report that includes instructions for model usage, and prediction results of the testing datasets. It is crucial to ensure that the format of the model input is consistent with that of the testing datasets. To submit your work, please write the prediction results into a JSON file using the following format: [{"ID":(str), "ParagraphTopic":(list), "Full-textTopic":(str)}, ...], with the same sample order as the testing datasets.

Please note that:

1) There may be some paragraphs without a clear topic sentence.
2) In general, the main topic sentence of the entire article is one of the paragraph topic sentences. However, due to the quality differences of the texts, there may be some articles that do not have a main topic sentence for the entire article.
3) Due to improper punctuation in some articles, the topic sentence of a paragraph or the main topic sentence of the entire article may not necessarily be a complete sentence that ends with a period.

In summary, participants should be aware of the above mentioned issues while reviewing and analyzing the texts. We encourage them to utilize their critical thinking skills and language proficiency to fully comprehend the contents and extract valuable insights.

Below is a sample input and output for your reference.

- Input Sample:

```JSON
{
    "ID": "3027",
    "Title": "学会“读”",
    "Text": [
        "在生活中，人们把学习的人叫学者，把研究艺术人叫艺术爱好者，把做演讲演说多人叫演讲者。那么，一个读书人叫什么，没错，是读者。",
        "其实“读者”是对读书人的一种赞誉，正因为这一点，所以读书之人不一定就能成为一个合格的读者。要学会读书，才是一个读者。",/home/hongyi/logic/小花狮/all_datas_final.json
        "读者，要学会读背景。一个真正喜欢读书人，是不会仅仅把全书看完一遍就了事了的。读书前，先把书的背景了解，可能读时会更能理解书中想表达的意思，比如说《儒林外史》一书，如果不了解作者当时所处的社会环境是那么地腐败黑暗，你会把这本书当一本好笑小说。确定，《儒林外夫》中吴敬梓的言辞十分白话，情节内容真的再有趣不过了，但这也主是吴敬梓想要达到的，他想用这些过分荒淡、好笑到人和事，反映出当时的社会是那么的可笑而更可悲啊！了解了书的背景,才能真正体现到书中的那种讽刺与作者的无奈。",
        "读者，要学会思考。《论语》中道：“学而不思则罔”，读书也是一种学习的过程，同样也需要读者对书进行思考和探究，很多书内容有些声，可能有时你还会对作者的观点有所不理解，甚至否认。但如果你在了解书背景的同时结合书中的内容加以深究，就会有不同的感受。就像《朝花夕拾》中鲁迅在《王倡会》的描绘的父亲形象，是那么的严厉，让人觉得鲁迅是在对自己的父亲表示讨厌，对父亲十分不喜爱，但你再以鲁迅所生的环境，想一想，你就会顿开茅塞，鲁迅这里并不是在怪自己的父亲，而是想通过这件事，表现出旧中国封建的思想教育方式抹杀了为孩子的天性。正就是思考的好处。",
        "读者，更要品读。品读也可以说是复读。很多的人会对一些名著进行复读，品析内容。其实复读更有利于让对书产生理解与共鸣，古人言：“读书百遍，奇异自
```

```
现”也正如此，每一遍读你都会有新的感悟。",
        "其实这三种方式读书也同样可以运用于生活，生活中也需要这样认真的态度以面对。要学会做一个读者，也更要有才为读者后学习书中之道，做一个生活的享受者。"
    ]
}
```

- Output Sample:

```
JSON
{
    "ID":"3027",
    "ParagraphTopic":[
        "那么，一个读书人叫什么，没错，是读者。",
        "要学会读书，才是一个读者。",
        "读者，要学会读背景。",
        "读者，要学会思考。",
        "读者，更要品读。",
        "要学会做一个读者，也更要有才为读者后学习书中之道，做一个生活的享受者。"
    ],
    "Full-textTopic":"要学会做一个读者，也更要有才为读者后学习书中之道，做一个生活的享受者。"
}
```

2.2.4 Training Datasets

We offer approximately 60 Chinese essays written by mIDdle school students, of which 50 can serve as training sets and 10 as verification sets. Each data sample contains the title and content of the article. Participants are also welcome to utilize data from other sources, such as manual annotation or automatic annotation using

models or tools, to enhance their training experience.

2.2.5 Testing Datasets

We offer a comprehensive collection of 5,000 Chinese essays that serve as our testing datasets. These valuable resources are made available to participants in the form of a JSON file that includes key information, such as the essay's ID, title, and text content in the format of **[{"ID":"","Title":"","Text":[]} ...]**.

To ensure the highest standards of accuracy and quality, we meticulously select a portion of the data in the test set for review. This enables us to provIDe insightful feedback to participants and further refine their method.

2.2.6 Evaluation Metrics

This task adopts accuracy to evaluate the effectiveness of extracting paragraph and overall themes in the text. The paragraph theme sentence accuracy (ParaAcc) is defined as the number of accurately IDentified paragraph theme sentences divIDed by the total number of paragraph theme sentences. The overall theme sentence accuracy (FullAcc) is defined as the number of accurately IDentified overall theme sentences divIDed by the total number of overall theme sentences. The comprehensive evaluation accuracy is calculated as 0.3 times the paragraph theme sentence accuracy plus 0.7 times the overall theme sentence accuracy.

2.3 Track 3. Paragraph Logical Relation Recognition

2.3.1 Task Description

IDentifying the logical relations between paragraphs is an important task in natural language processing, especially in the fields of text understanding and information extraction. For example, correctly IDentifying the logical relations between paragraphs can help generate accurate and coherent summaries or answers in text summarization and question-answering systems. In student writing, IDentifying the logical relations between paragraphs can provIDe valuable insights into writing quality, coherence, and the ability to structure arguments and narratives. This task requires a deep understanding of the content and context of the text, as well as the ability to IDentify and interpret various language cues that indicate the logical relations between paragraphs. Therefore, this task aims to evaluate current technology's ability to recognize the logical relations between paragraphs in common essay topics among mIDdle school students, based on given definitions and examples.

2.3.2 Task Definition

Given two paragraphs sorted in order from a composition, the annotator needs to determine the logical relationship between the two paragraphs based on the given definitions and examples of logical relationships. The definition and examples of

logical relations are as follows:

逻辑关系	定义	示例	
		段落 1	段落 2
共现关系	同一个共现关系可以连接两个以上平等的子句或句子，如并列关系中可以包括多个并列项。共现关系包括并列、顺承、递进、对比四个类别。	从 天 安 门 往 里 走，沿着一条笔直的大道穿过端门，就到了午门的前面。	走进午门，是一个宽阔的广场。
反转关系	一篇章单位提出一个客观的或假设的事实，另一篇章单位叙述一个与之相反的情况。反转关系包括转折和让步两种类型，其中转折关系是在真实的情况上进行反转，而让步关系则是对一个假设事件的反转。	我担心父亲有一天会垮下来。	然而，父亲的精力却很旺盛。
解说关系	后一篇章单位对前一篇章单位进行补充、解说，可以针对前文整句的意思进行追补，也可以针对某个词来作一些补充。解说关系区别于共现关系的其他类型：一方面，对比并列、选择关系，解说关系中后一单位的意义是建立在前一单位的基础之上，或进行更细化的描述，或进行总结；而并列和选择关系连接的篇章单位则完全平等。另一方面，对比顺承、递进关系，顺承强调顺序，递进强调程度加深，解说强调补充。根据后一篇章单位对前一单位所起的作用，解说关系可以分为细化和泛化两个小类。	科学家们提出了许多设想。	例如，在火星或者月球上建造移民基地。
主从关系	篇章单位之间地位不平等，有主次之分，则构成主从关系。主从	弯弯的月儿小小的船，小小的船	我在小小的船里坐，看见闪闪的

	关系只可以二分，也就是说，无论主句和从句中包含几个子句，总是将主句和从句部分分别视为一个整体，先将这二者切分开来。主从关系包括背景、主观推论、客观因果、假设条件、特定条件五个类别。	儿两头尖。	星星蓝蓝的天。

2.3.3 Expected Outputs

The submission should consist of a Python-based model, a development report that includes instructions for model usage, and prediction results of the testing datasets. It is crucial to ensure that the format of the model input is consistent with that of the testing datasets. To submit your work, please write the prediction results into a JSON file using the following format: [{"ID":(str), "Relation":(str)}, ...], with the same sample order as the testing datasets.

Below is a sample input and output for your reference.

- Input Sample:

```
JSON
{
    "ID": "3027",
    "paragraph1": {
        "PID":"355",
        "Text":"在生活中，人们把学习的人叫学者，把研究艺术人叫艺术爱好者，把做演讲演说多人叫演讲者。那么，一个读书人叫什么，没错，是读者。"
        }
    "paragraph2": {
        "PID":"356",
        "Text":"其实“读者”是对读书人的一种赞誉，正因为这一点，所以读书之人不一定就能成为一个合格的读者。要学会读书，才是一个读者。"
    }
```

```
}
```

- Output Sample:

```
JSON
{
    "ID":"3027",
    "Relation":"解说关系"
}
```

2.3.4 Training Datasets

We offer approximately 120 paragraph pairs, of which 100 can be used as training sets and 20 as verification sets. Each data sample contains the text content of two paragraphs and their corresponding paragraph IDs. Please note that the dataset may contain paragraph pairs without any logical relationship.

Furthermore, participants are welcome to utilize data from other sources, such as manual annotation or automatic annotation using models or tools, to enhance their training experience. With these additional resources, participants can deepen their understanding of language composition and further hone their skills in analyzing and connecting IDeas between paragraphs.

2.3.5 Testing Datasets

We offer a comprehensive collection of 5,000 Chinese essays that serve as our testing datasets. These valuable resources are made available to participants in the form of a JSON file that includes key information, such as the essay's ID and paragraph pairs in the format of **[{"ID": "", "paragraph1": {"PID":"", "Text": ""}, "paragraph2": {"PID":"","Text":""}} ...]**.

To ensure the highest standards of accuracy and quality, we meticulously select a portion of the data in the test set for review. This enables us to provIDe insightful feedback to participants and further refine their method.

2.3.6 Evaluation Metrics

The evaluation of the recognition performance of logical relationships between paragraphs in this task will use precision (P), recall (R), and macro-F1 score (F1-score, F1). Precision is calculated as the number of correctly IDentified logical relationship types divIDed by the total number of IDentified logical relationship types. Recall is calculated as the number of correctly IDentified logical relationship types divIDed by the total number of annotated logical relationship types. F1-score is

calculated as (2*precision*recall)/(precision+recall).

2.4 Track 4. Sentence Logical Relation Recognition

2.4.1 Task Description

In the assessment of writing, the logical relationships between sentences are one of the important factors for evaluating the quality of students' compositions. The logical relationships between adjacent sentences in a composition are essential for evaluating the fluency, coherence, and overall logical structure of the composition. Correct logical relationships between sentences can make the structure of the composition clear, promote the development of IDeas, and make it easier for readers to understand the author's point of view and argument. The task of evaluating the logical relationships between sentences usually includes analyzing and classifying adjacent sentences in student compositions, evaluating the logical structure and coherence between sentences based on their relationships, such as causal relationships, comparative relationships, chronological relationships, etc. To evaluate the logical relationships between sentences in student compositions, a deep understanding of natural language processing techniques is needed, combined with an understanding and analysis of the text content, to improve the accuracy and objectivity of the evaluation.

2.4.2 Task Definition

Given two sentences from an essay that are ordered sequentially, the annotator needs to determine what type of logical relation exists between them based on given definitions and examples. The definition and examples of logical relations are as follows:

逻辑关系	定义	示例	
		句子 1	句子 2
并列关系	描述同一事件的几个方面、相关的几件事情或相对的情况，在意义上并存、共现或对立，在语序上可以调换顺序且不改变句义。	我的老爸就像一只鸡，每天都很早起，然后上班，每次他都很早起，早饭就吃两口，再上班，而且很晚睡觉。	我就像一个变色龙，总是变脸，我有时不开心的时候就板着脸，我一开心就一直微笑，我伤心的时候就外表，我的衣服也会变哦。

顺承关系	篇章单位之间存在时间、空间、步骤、逻辑事理上的先后顺序，包括顺序和逆序两种情况；但不包括同时发生的事件，同时的事件属于并列关系。由于存在先后关系，篇章单位的顺序不可随意调换。篇章单位的主体可以是同一个人或事物，也可以是不同的人或事物。	我看见女娲先杀死了一只大乌龟，用它的腿撑着天空。	接着，杀死了一只黑龙。
递进关系	后一篇章单位在数量、质量、范围、时间等方面比前一篇章单位更进一层，强调程度的增强加深；篇章单位的顺序通常不可调换。与顺承相比，顺承只表现为一种先后顺序，没有程度的加深；而递进关系则强调后者比前者在程度上更进一步。	图书馆里有各种各样的图书，种类数也数不清。	甚至连英文书都有呢。
对比关系	对比关系是指在文本中出现的两个或多个事物、概念、观点、行为、状态等之间的明显的、直接的、相对的差异或相似性。对比关系通常用于强调、比较和对照。	苹果酸甜可口，口感脆嫩，适合生食和烹饪。	而橙子则酸甜适中，多汁而且富含维生素C，适合榨汁和制作甜点。
让步关系	某一篇章单位提出一	她虽然不用功学习。	考试却及格了。

	个假设的事实，并且退让一步暂且承认这个假设的真实性，另一篇章单位叙述一个与之相反或相对的情况。语序上，假设的事实通常在前。		
转折关系	某一篇章单位提出一个客观事实，另一篇章单位叙述一个与之相反或相对的情况。语序上，转折部分一般在后，有时也会倒装变化。	月亮发出的黄色光芒，把周围的几朵灰灰的云也照黄了。	但又仔细一看，好像月亮也不是纯黄色的，有黑乎乎的东西在上面。
泛化关系	后一篇章单位是对前一篇章单位的概括、总结和泛化；篇章单位间不可调换顺序，否则转变为细化关系。与之相关的连接成分包括“总之”、“总体而言”、“综上所述”等；与细化类似，泛化关系也主要依靠意义的制约。	为了达到这个目的，他们讲究亭台轩榭的布局，讲究假山池沼的配合，讲究花草树木的映衬，讲究近景远景的层次。	总之，一切都要为构成完美的图画而存在，决不容许有欠美伤美的败笔。
细化关系	后一篇章单位是对前一篇章单位的细化描述，包括举例、解释、说明、补充等；篇章单位间不可调换顺序，否则转变为泛化关系。与之相关的连接成分包括“这”、“即”、“例如”、“也就是说”等；但与其他类型	我的老妈就像一个母老虎，我一不听话她就发脾气。	有一次，我没有写完作业她就发脾气，说："你怎么还没写完啊！。"

	相比，细化关系在多数情况下没有提示成分，通常表现为词义或句义的关联。		
客观因果关系	某一篇章单位说明原因，另一篇章单位说明由该原因导致的结果，两者均是客观事实。原因和结果的前后位置不固定，但二者有主次之分，有时原因为主，有时结果为主，视句义而定。	我查了书籍，原来农历十五、十六都为满月。	所以今天的月亮也是最大最圆的。
背景关系	篇章中经常出现事件、地点、历史等情况的介绍，此类环境信息与篇章正文构成背景关系。在背景关系中，某一篇章单位交代事情发生的历史情况、现实环境、前情概要等，如时间、地点、历史背景、政治环境等，另一篇章单位叙述事情的内容。背景关系具有特定的限制条件：如果事件背景和内容之间发生因果、转折等其他主从关系，则优先标注其他关系；只有单纯的环境描写才算作背景关系。语序上，背景部分通常在事件内容之前。	迪士尼乐园是人们向往的地方，也是周末玩耍的好去处。	今天我就给大家推荐上海迪士尼乐园。

特定条件关系	某一篇章单位提出特定的条件，另一篇章单位说明以该条件为依据推断出的结果。其中，特定条件可以包括充足条件，代表的格式是“只要……就……”；可以是必要条件，常用的连接成分有“只有……才……”、“除非……否则……”等；也可以是周遍性条件，常使用“无论……都……”、“不管……也……”等格式。	只有坚持锻炼。	才会有好身体。
假设条件关系	某一篇章单位提出虚拟性条件，另一篇章单位说明该假设条件实现后所产生的结果，或为了实现该假设条件而应采取的措施。	如果我们好好学习。	就能取得好成绩。
主观推论关系	某一篇章单位说明事实依据，另一篇章单位说明由此推断出的主观结论；与客观因果关系所不同的是，推论得到的结果是主观的。语序上，事实依据往往在前，主观结论在后；二者有主次之分，结论通常是句义的核心。	去之前一定要提前预约！	不然你可能会排两个小时的队！

2.4.3 Expected Outputs

The submission should consist of a Python-based model, a development report that includes instructions for model usage, and prediction results of the testing datasets. It is crucial to ensure that the format of the model input is consistent with that of the testing datasets. To submit your work, please write the prediction results into a JSON file using the following format: [{"ID":(str), "Relation":(str)}, ...], with the same sample order as the testing datasets.

Below is a sample input and output for your reference.

- Input Sample:

```JSON
{
    "ID": "6027",
    "sentence1": {
        "SID":"355",
        "Text":"读书前，先把书的背景了解，可能读时会更能理解书中想表达的意思。"
        }
    "sentence2": {
        "SID":"356",
        "Text":"比如说《儒林外史》一书，如果不了解作者当时所处的社会环境是那么地腐败黑暗，你会把这本书当一本好笑小说。"
    }
}
```

- Output Sample:

```JSON
{
    "ID": "6027",
    "Relation": "细化关系"
}
```

2.4.4 Training Datasets

We offer approximately 240 sentence pairs, of which 200 can be used as training sets and 40 as verification sets. Each data sample contains the text content of two paragraphs and their corresponding paragraph IDs. Please note that the dataset may contain paragraph pairs without any logical relationship.

Furthermore, participants are welcome to utilize data from other sources, such as manual annotation or automatic annotation using models or tools, to enhance their training experience. With these additional resources, participants can deepen their understanding of language composition and further hone their skills in analyzing and connecting IDeas between paragraphs.

2.4.5 Testing Datasets

We offer a comprehensive collection of 10,000 Chinese essays that serve as our testing datasets. These valuable resources are made available to participants in the form of a JSON file that includes key information, such as the essay's ID and sentence pairs in the format of **[{"ID": "", "sentence1": {"SID":"", "Text":""}, "sentence2": {"SID":"","Text":""}} ...]**.

To ensure the highest standards of accuracy and quality, we meticulously select a portion of the data in the test set for review. This enables us to provIDe insightful feedback to participants and further refine their method.

2.4.6 Evaluation Metrics

In this task, precision (P), recall (R), and Macro-F1 value (F1-score, F1) are used to evaluate the recognition performance of logical relationships between sentences.
Precision = the number of correctly IDentified logical relationships of a certain type / the total number of IDentified logical relationships of that type.
Recall = the number of correctly IDentified logical relationships of a certain type / the total number of logical relationships of that type annotated in the dataset.
F1-score = (2 * precision * recall) / (precision + recall).