

<b>0 Preprocessing</b>	<b>2</b>
0.1 Gold metadata information	2
0.2 Gold section, paragraph and sentence splits	2
0.3 Gold tokenization	2
0.4 Automatic dependency parsing	2
<b>1 EDU Segmentation</b>	<b>4</b>
1.1 Segmented Units	4
1.1.1 Purpose clauses	4
1.1.2 Relative clauses	4
1.1.3 Manners and Means	7
1.1.4 Reported speeches and cognitive predicates	7
1.1.5 Coordinations	11
1.1.6 Predicative adjectives	11
1.1.7 Parentheticals and references	12
1.1.8 Dashes, hyphens, and colons	15
1.1.9 Strong discourse cues	17
1.1.10 Translanguaging	19
1.2 Not Segmented Unit	22
1.2.1 Complement clauses	22
1.2.2 Prepositional phrases	24
1.2.3 Dislocated NPs	25
1.2.4 MSP	25
1.2.5 Coordinations	26
1.2.6 Existential clauses	27
1.3 Compare & Contrast:	28
1.3.2 Validation Checklist	29
1.3.3 Preposition or Verb Table	30
<b>2 Relation Annotation</b>	<b>32</b>
2.0 Relation marking for relative clauses.	32
2.1 Nucleus-Satellite Relations	33
2.1.1 adversative-antithesis	33
2.1.2 adversative-concession	33
2.1.3 attribution-negative	33
2.1.4 attribution-positive	33
2.1.5 causal-cause	33
2.1.6 causal-result	33
2.1.7 context-background	34

2.1.8 context-circumstance	34
2.1.9 contingency-condition	34
2.1.10 elaboration-additional	34
2.1.11 elaboration-attribute	34
2.1.12 evaluation-comment	34
2.1.13 explanation-evidence	35
2.1.14 explanation-justify	35
2.1.15 explanation-motivation	36
2.1.16 mode-manner	36
2.1.17 mode-means	36
2.1.18 organization-heading	37
2.1.19 organization-phatic	37
2.1.20 organization-preparation	37
2.1.21 purpose-attribute	37
2.1.22 purpose-goal	38
2.1.23 restatement-partial	38
2.1.24 topic-question	38
2.1.25 topic-solutionhood	38
2.2 MultiNuclear relations	38
2.2.1 adversative-contrast	38
2.2.2 joint-disjunction	39
2.2.3 joint-list	39
2.2.4 joint-sequence	39
2.2.5 joint-other	39
2.2.6 restatement-repetition	39
2.3 same-unit	40
2.4 Other relation issues	41
2.4.1 Attachment order of partial modification	41
3.0 Questions for Janet	44
3.1 Questions for Amir	47
3.2 others	77

# 0 Preprocessing

## 0.1 Gold metadata information

Gold metadata annotation assembles [the Georgetown University Multilayer \(GUM\) Corpus](#). We include the following metadata at the beginning of *raw/\*.txt* documents:

- ***title***: the Chinese title of each document displayed in the source data. e.g. 老龄化对言语感知的影响
- ***shortTitle***: a one-word English short title for each document, e.g. *aging*
- ***type***: the genre of each document, e.g. *academic*
- ***text id***: a unique text id per document, which is the combination of the corpus name, genre and short title of the document, e.g. *gum\_zh\_academic\_aging*
- ***author***: the author of the document, e.g. *Meijuan Ning* for academic articles or *Wikipedia, The Free Encyclopedia*
- ***dateCreated, dateModified & dateCollected***: dates when the document is first created, last modified in the source and collected by this corpus. The dates follow YYYY-MM-DD format and XXXX-XX-XX if such information is unknown
- ***sourceURL***: the source URL where the document is retrieved, e.g. <https://www.hanspub.org/journal/PaperInformation.aspx?paperID=28037>
- ***speakerCount***: the number of speakers in the document, e.g. *0*
- ***speakerList***: the list of speakers in the document, e.g. *none*

## 0.2 Gold section, paragraph and sentence splits

Gold section, paragraph and sentence splits are annotated in the *raw/\*.txt* documents for future macro- versus micro- level RST analyses:

- Section and subsection breaks are marked by: `<section/>`, `<subsection/>`, `<subsub...>`
- paragraph breaks are marked by two new line feeds: `\n\n`
- sentence breaks are marked by one new line feed: `\n`

## 0.3 Gold tokenization

This corpus follows the tokenization guidelines for [The Segmentation Guidelines for the Penn Chinese Treebank \(3.0\)](#) and mirrors decisions in the [Chinese Treebank \(CTB\) 9.0 corpus](#).

## 0.4 Automatic dependency parsing

We use the Python stanza package for automatic dependency parsing. Instead of the default *gsdsimp* model trained on [UD\\_Chinese\\_GSDSimp](#), we convert the [Chinese Treebank 9.0](#) constituent trees to dependencies using [Stanford CoreNLP](#) and use them to train stanza. Though not natively annotated in dependencies, the CTB corpus is more consistently annotated for tokenization, POS tags and syntactic relations.

The java conversion command is the following:

```
java -mx1024m -cp "*;"  
edu.stanford.nlp.trees.international.pennchinese.ChineseGrammaticalStructure -treeFile <FILE>  
-basic -conllx
```

# 1 EDU Segmentation

Chinese examples are shown in italics with gold tokenization and double pipelines (||) indicating EDU boundaries, **EDUs under discussion are highlighted in bold**.

|| is used to indicate segmented EDU boundaries.

## 1.1 Segmented Units

### 1.1.1 Purpose clauses

Purpose clauses are segmented. Most of these purpose clauses are examples of Serial Verb Constructions (SVC) in Chinese.

Examples:

1. 他 于 **1724** 年 前往 圣彼得堡 || 出任 数学 教授 , || 但 不 喜欢 那里 。  
"He in 1724 go-to St-Petersburg || take-office math professor || but NEG like there"  
source: gum\_zh\_bio\_bernoulli
  
2. ... || 于是 和 母亲 移居 到 諾丁汉郡 的 世袭 领地 || 生活 。 || [ 6 ]  
"... || so and mother move to Nottinghamshire DE hereditary territory || live || [6] "  
source: gum\_zh\_bio\_byron

### 1.1.2 Relative clauses

Relative clauses marked by **de** (的) are segmented forming a SAME-UNIT + ELAB-ATTR relation.

Examples:

3. ... || 约翰 因为 不能 承受 || 和 他的 后代 做 比较 的 || “ 羞耻 ” , || 把 丹尼尔 逐出 他 的 家族 。  
"John because NEG can endure || with he DE descendants make comparison DE || shame || BA Daniel expel he DE family "  
source: gum\_zh\_bio\_bernoulli
  
4. ... || 2月 , 上议院 通过 了 || 毁坏 机器 的 || 工人 必须 判处 死刑 的 || 法案 || ...  
"... || February Upper-House pass LE || destroy machine DE || worker must condemn death-penalty DE || bill || ... "  
source: gum\_zh\_bio\_byron

Circumstances are frequently realized as relative clauses with constructions such as:

- 在 || VP 的 || 时候 || (at the time of VP)
- 在 || VP 的 || 情况 下 || (at the circumstance of VP)

where SAME-UNIT+ELAB-ATTRIB is also applied.

5. 在 || 他 还 只 有 十 几 岁 的 || 时 候 , || 他 就 发 现 了 || *n* 次 多 项 式 可 以 用 根 式 解 的 充 要 条 件 , || ...

"at || he still only teens year-old DE || time, || he already discover LE || *n*-degree polynomial can use radical solve DE necessary-sufficient condition || ..."

source: gum\_zh\_bio\_galois

Relative clauses can also be let by 之, which is a more formal alternative to 的.

6. 而 它 们 的 ||

And they DE ||

逝 去 、 ||

passing-away ||

消 亡 之 ||

dying-out DE ||

处 正 是 ||

place exactly is ||

生 出 它 们 的 ||

give-birth they DE ||

本 根 之 道 。

root DE way

source: gum\_zh\_academic\_taoist

In rare cases, relative clauses can omit the overt DE. In these situations, we ensure the pre-nominal modifier is clausal (by running LE-insertion, ZHE-insertion or adverb insertion tests), and segment these DE-less relative clauses.

7. 进 入 仙 桃 ||

enter Xiantao ||

人 员 必 须 进 行 体 温 检 测

people must conduct temperature check

source: gum\_zh\_news\_hubei

8. 24 日 22 时 , (潜江市) 关 闭 潜 江 站

24th 22:00, (Qianjiang-city) close Qianjiang-station

||

||

离 开 潜 江 ||

leave Qianjiang ||

通 道 。

pathway ||

"they will close Qianjiang-station's pathway to leave Qianjiang. "

source: gum\_zh\_news\_hubei

9. 尽管 有一些 对于 ||  
*Although EXIST some toward ||*  
含有 致命 病毒 ||  
**contain deadly virus ||**  
疫苗 安全性 的 抨击 , ||  
**vaccine safety DE criticism ||**  
*"Although there are some criticisms towards vaccines that contain deadly virus"*  
source: gum\_zh\_academic\_rabies

We also segment relativized prepositional phrases since they function as clausal modifiers to nominals.

10. 因为 ||  
because ||  
"because"  
关于 他的 行为 的 ||  
about he POS behavior DE ||  
"that are about his behaviors"  
所有 纪录 都 是 一些 奇闻逸事 。  
all record all COP some anecdotes  
"all records are anecdotes."  
source: gum\_zh\_bio\_emperor

However, 的(DE)-marked clauses can directly function as complement clauses. In these situations, they are not segmented. In the following example, "非 标记性 主位 传达 的" is the subject of "是 旧 信息" thus they belong to the same EDU.

11. 与 主语 重合 的 ||  
with subject coincide DE ||  
"that coincides with the subject" ||  
非 标记性 主位 传达 的 是 旧 信息 , ||  
non- marking theme convey DE COP old information , ||  
"those conveyed by non-marking theme are old information," ||  
而 述位 传达 的 是 新 信息 。 ||  
but rheme convey DE COP new information ||  
"but those conveyed by rheme are new information." ||  
source: gum\_zh\_academic\_iconicity

12. 对方 在 这 段 对话 中 得到 的 只 有 乐 趣 和 充 满 趣 味 回 忆 ,  
the-other-side at this CL conversation middle get DE only have fun and full joy memory  
"those the other side gets in this conversation are only fun and joyful memories, "  
source: gum\_zh\_whow\_flirt

### 1.1.3 Manners and Means

Manner clauses usually occur in the middle of a sentence. Here are some examples:

13. 而且 他 试图 || 用 这 一 方 式 || 解 释 波 义 耳 定 律 , || ...  
"and he try || use this one method || explain Boyle's law || ..."  
source: gum\_zh\_bio\_bernoulli

14. 有 一 种 业 余 玩 法 是 将 边 上 的 球 ||  
*EXIST one CL amateur method COP BA edge DE ball ||*  
按 照 半 色 、 实 色 、 半 色 、 实 色 的 顺 序 ||  
*according-to half-color , solid-color, half-color , solid-color DE order ||*  
摆 放 。  
*place*  
source: gum\_zh\_whow\_pool

15. 然 后 再 ||  
*after-that then ||*  
看 情 况 ||  
*see situation ||*  
采 取 进 一 步 行 动 。  
*take further action*  
source: gum\_zh\_news\_tiktok

16. 读 起 来 ||  
*read start ||*  
挺 拗 口 的  
*very mouthful DE*  
source: gum\_zh\_whow\_glowstick

### 1.1.4 Reported speeches and cognitive predicates

**Reported speeches and cognitive predicates suggest segmentation of the complement when two conditions are simultaneously met**, where the writer information is an attribution to the content in the second EDU.

- (a) The main predicate belongs to one of the following categories:
  - saying verbs;
  - cognitive verbs (feelings, thoughts, hopes);
  - perception verbs (see, feel, hear, sense).
- (b) The complement is by itself a full clause

The following verbs are attested to introduce reported speeches in Chinese:

- 说 say
- 声称 claim
- 要求 request
- 宣称 claim
- 宣布 announce
- 建议 suggest
- 询问 inquire
- 提到 mention
- 称 claim
- 描述 describe
- 否认 deny
- 显示 indicate
- 主张 assert
- 建议 suggest
- 规定 prescribe
- 讨论 discuss
- 综述 sum up
- 提出 put forward
- 标志 signify
- 表示 express
- 明确 clarify
- 强调 emphasize
- 指责 accuse
- 确定 affirm
- 说明 explain
- 认为 think
- 以为 think
- 想像 imagine
- 相信 believe
- 知道 know
- 懂得 understand
- 明白 understand
- 推荐 recommend
- 写着 (that) writes
- 感觉 feel
- 想着 think
- 约定 agree
- 觉得 feel
- 希望 hope
- 表明 indicate
- 意味 imply
- 提示 suggest
- 倡导 advocate
- 期望 expect

- 赞成 approve
- 计划 plan
- 打算 plan
- 决定 decide
- 盼 hope
- 鼓励 encourage
- 透露 reveal
- 探讨 discuss
- 看到 see
- 见到 see
- 想到 think of
- 思考 think
- 发现 discover
- 考虑 consider
- 记住 remember
- 记得 remember
- 谨记 remember

Negative attribution verbs:

- 害怕 be afraid of

For reference, the following verbs are included as reported speech and cognitive verbs in English RST-DT (Carlson & Marcu 2003):

- say, tell, state, announce, declare, suggest, advise, report, indicate, point out, explain, ask
- think, believe, know, imagine, suppose, conjecture, wish, hope, predict, fear, estimate, calculate, anticipate, expect, dream
- see, feel, hear, sense

**17. 他 自己 说 : || “在 应用文 方面 , 英文 、 德文 、 法文 没有 问题 。**

*"he self say : || " in formal-writing aspect, English, German, French NEG-have problem ."*

source: gum\_zh\_bio\_chao

18. ... || 即 援引 他 棺材 上 的 银盘 刻印 , ||  
... || *that-is* cite his coffin on-top-of DE silver-place engraving ||  
认为 ||  
*think* ||  
他 是 “年 约 65 岁。”  
*he is age about 65 years-old*  
source: gum\_zh\_bio\_emperor

Moreover, subject of the reported speech can be implied or inherited from previous context.

19. 至少 让 其他 人 知道 ||  
*at-least let other people know ||*  
你 要 去 哪儿 ,  
*you want go-to where*  
source: gum\_zh\_whow\_hiking

Such attribution of reported speech can as well be negative.

20. 但是 , 如果 不 知道 ||  
But if NEG know  
怎么 正确 地 驱逐 老鼠 ||  
*How-to correct-ly expel mice,*  
source: gum\_zh\_whow\_mice

One sentence can contain multiple occurrences of combinations of speech verb + content. In the following example, 提到 (mention), 称 (state) and 希望 (hope) introduce new EDUs. However, the EDU separation after 说 (say) is due to the relative clause after instead of a complete clausal complement of saying.

21. 他 提到 ||  
*he mentioned ||*  
较早前 接受 电视台 节目 采访 时 ||  
*earlier receiving TV show interview at-time-of ||*  
说 ||  
*say ||*  
“二元 优惠 计划 可能 要 调高 金额 至 三元 ” 的 ||  
*"2-yuan discount plan may have-to increase amount to 3-yuan " DE*  
讲法 , ||  
*statement ||*  
称 ||  
*state ||*  
这 不 是 政府 的 立场 , ||  
*This NEG COP government DE position,*  
希望 ||  
*hope ||*  
不 要 引起 一些 不 必要 的 误会 。 ||  
*NEG will cause some NEG necessary DE misunderstandings.*  
source: gum\_zh\_news\_unemployment

Here is a counter-example where 宣称 (claim) does not introduce a new EDU because the following portion is not a clause, but a fixed expression 宣称 A 为 B (claim A to be B).

22. 卡美哈梅哈 五世 宣称 诺顿一世 为 “全 美国 唯一 的 统治者 ”。  
*Kamehameha V declared Norton I to-be "all America only DE ruler".*

source: gum\_zh\_bio\_emperor

### 1.1.5 Coordinations

**Coordinated copula clauses are separated.**

Examples:

23. 他 是 欧拉 的 同 代 人 , || 也 是 密 友 。

"he COP Euler DE same era people , || also COP close-friend"

source: gum\_zh\_bio\_bernoulli

**Subordinated coordinating conjunctions are also separated.**

Examples:

24. ... || 每 天 忙 着 为 希 腊 军 队 筹 集 物 资 , || 购 买 先 进 武 器 , || 调 节 内 部 纠 纷 , || ...

" ... || everyday be-busy for Greek army raise supply , || purchase modern weapon , || resolve internal conflict , || ... "

source: gum\_zh\_bio\_byron

### 1.1.6 Predicative adjectives

**Predicative adjectives in Chinese do not require overt copula and can be segmented from other clauses.**

25. 拜 伦 先 天 性 的 跛 足 , || 而 他 的 母 亲 性 情 乖 弑 、 喜 怒 无 常 , || ...

"Byron congenital DE lame, || but he DE mother (COP) temper grumpy ,happy-sad unstable || ..."

source: gum\_zh\_bio\_byron

Similar to English, when multiple predicative adjectives are conjoined with the same subject, they jointly form one EDU. **We DO NOT segment these coordinated predicative adjectives.**

26. 我 知 道 ||

I know ||

这 周 你 很 辛 苦 、 很 不 容 易

this week you very hard-working 、 very NEG easy

source: gum\_zh\_whow\_procrastinating

However, exceptions apply when a predicative adjective is conjoined by a strong discourse marker, for example **而且** (but also) in the following example.

27. ... || 由 于 证 据 含 糊 不 清 、

矛 盾 || 而 且 寥 寥 无 几 , || ...

"... || since evidence (COP) ambiguous , contradictory ,|| but-also rare , "  
source: gum\_zh\_bio\_byron

28. 又 不 想 显 得 太 主 动 , 甚 至 猥 琐 ? ||

again NEG want appear too active , even wretched ?

"Don't you want to appear too active, or even wretched? "

source: gum\_zh\_whow\_flirt

29. 虽 然 安 全 , 但 不 方 便 , ||

although safe but NEG convenient

"Although safe, ?|| but inconvenient,"

source: gum\_zh\_interview\_wimax

Moreover, when some of the conjoined predicative adjectives take PP or NP complements, they are separated from other adjectives and form their individual EDU.

30. 突 发 疫 情 是 指 突 发 的 、 群 发 的 、 ||

sudden epidemic COP refer-to suddenly-happened DE grouply-happened DE ||

对 公 共 健 康 或 经 济 、 政 治 、 社 会 等 影 响 大 的 ||

toward public health or economy politics society etc. influence big DE ||

( 已 造 成 危 害 ||

( already cause harm ||

或 可 能 造 成 危 害 ), ||

or may cause harm ), ||

需 要 紧 急 采 取 控 制 措 施 , ||

need urgently take control measure ||

与 传 染 病 有 关 的 ||

to infectious-disease related DE ||

公 共 卫 生 事 件 ||

public health event ||

[ 1 ]。

[ 1 ].

source: gum\_zh\_academic\_governance

### 1.1.7 Parentheticals and references

**Parentheses are separated**, including round "()" , square "[]" and curly "{}" brackets.  
However, "《 》" mark book titles in Chinese and does not create EDU boundaries.

31. 约 翰 还 曾 试 图 盗 窃 丹 尼 尔 的 著 作 《 *Hydrodynamica* 》 || ( 流 体 力 学 ) ||

"John also once try steal Daniel DE piece Hydrodynamica || ( fluid mechanics ) ||  
并 把 它 重 新 命 名 为《 **Hydraulica** 》。  
also BA it anew name to-be **Hydraulica** ."  
source: gum\_zh\_bio\_bernoulli

In accordance with RST-DT and GUM guidelines, **supporting references are separated from the contents.**

32. 希腊 政府 为 拜伦 举 行 了 隆 重 的 国 葬 仪 式 。||  
"Greek government for Byron take-place LE grand DE state-funeral ceremony . ||  
[1][2]||  
[1][2]||"

source: gum\_zh\_bio\_byron

However, **exceptions apply when square brackets denote International Phonetic Alphabet (IPA), or when the brackets denote mathematical equations.**

33. 参 加 了 ||  
attend LE ||  
需 要 识 别 ||  
need-to identify ||  
音 节 [ ba ] 或 [ pa ] 和 [ ba ] 、 [ da ] 或 [ ga ] 合 成 的 ||  
syllable [ ba ] or [ pa ] and [ ba ] , [ da ] or [ ga ] synthesize DE ||  
连 续 统 的 ||  
continuum DE ||  
实 验 。  
experiment  
source: gum\_zh\_academic\_aging

34. 当 且 仅 当  $p=2^{\{k\}}+1$  ||  
if and only if  $p=2^{\{k\}}+1$  ||  
source: gum\_zh\_bio\_galois

35. 你 将 获 得  $2^{(X-1)}$  元 。 ||  
you will win  $2^{(X-1)}$  yuan 。 ||  
source: gum\_zh\_bio\_bernoulli

In addition, **inserted core arguments are not separated whereas optional modifiers are in the following examples.**

36. 在 此 之 前 , ( 我 们 ) 都 密 集 地 和 秘 书 长 与  
prior-to-this , ( we ) all intensive -ly with secretary-general and  
副 秘 书 长 进 行 联 系 , ||

deputy-secretary-general conduct connect , ||  
"Prior to this , ( we ) were in intensive contact with the secretary general and the deputy  
secretary general ,"  
source: gum\_zh\_interview\_cycle

37. 基本上 是 必须 要 与 ||  
basically COP must must-be with ||  
"basically we must be with"  
( 自由车 ) ||  
( bicycle ) ||  
"the ( bicycle )"||  
协会 进行 多 次 协调 的 , ||  
association conduct multiple CL coordination DE , ||  
"association to conduct multiple coordinations ,"  
source: gum\_zh\_interview\_cycle

38. 我们 ||  
we ||  
"we"  
( 德懋 国际 ) ||  
( Demao International ) ||  
"( Demao International )"  
非常 荣幸 , ||  
very honored , ||  
"are very honored,"  
能 赞助 这 次 的 环台赛 。 ||  
can sponsor thi CL DE Ring-Taiwan-Tour. ||  
" that we can sponsor this Ring Taiwan Tour ."  
source: gum\_zh\_interview\_cycle

Moreover, Note that **parenthetical dates in article citations are not EDUs, but parenthetical dates describing dated events, birth years, etc. are EDUs:**

In English, we see the following:  
[We read Smith (2000)]  
[Jane Smith] [(1901-1974)] [was a paleontologist]

Similarly in Chinese:

39. 另外 樋口 的一些作品 , 林文月 翻译 并 发表 至  
in-addition Higuchi de some works , Lin-Wenyue translate and publish in  
《联合文学》杂志 中 , 如 《比肩》( 1998年 1月 )、  
" United Literature" magazines in , for-example " Bijian " ( January 1998 ) ,

《浊江》(1998年9月)。||  
"Zhuojiang" (September 1998). ||

"In addition, some of Higuchi's works were translated and published by Lin Wenyue in "United Literature" magazines, such as 'Bijian' (January 1998) and 'Zhuojiang' (September 1998)."

source: gum\_zh\_bio\_higuchi

### 1.1.8 Dashes, hyphens, and colons

Similar to RST-DT, when dashes and multi-hyphens introduce parenthetical information or subtitles, we break the sentence and include the dashes and hyphens in the embedded EDU.

40. 德沃夏克 在 纽约 遇到了 他 后来 的 学生 哈里·布雷 ||  
*Dvorak at New-York met PERF 3SG.M future DE student Harry-Bray ||*  
—— 最早 的 美国 黑人 作曲家 之一。||  
-- *most early DE American black composers one-of . ||*  
"Dvorak met his future student Harry Bray in New York – one of the first African-American composers."

source: gum\_zh\_bio\_dvorak

Single hyphens commonly denote a combined meaning between words and thus do not create new EDUs.

41. 菱形 球框 里 球 的 摆放 方式 是 按照  
*rhombus-shaped bracket inside ball DE placement way COP according to*  
1 - 2 - 3 - 2 - 1 的 顺序 来 的 。||  
1 - 2 - 3 - 2 - 1 DE order place DE . ||  
"The balls are placed in the rhombus in the order 1 - 2 - 3 - 2 - 1."  
source: gum\_zh\_whow\_pool

42. 1933年 至 1936年 年间 , ||  
1933 to 1936 between-years , ||  
横跨 旧金山湾 的 ||  
*span-across SF-bay DE ||*  
旧金山 - 奥克兰 海湾 大桥 ||  
*SF - Oakland bay bridge ||*  
( 又 称 海湾 大桥 ) ||  
( *also name bay bridge* ) ||  
建成。  
*build*  
source: gum\_zh\_bio\_emperor

**Exception:** when a multi-hyphen or dash functions as a nominal combinator (similar to a single hyphen), we do not segment it.

TODO: Example hidden

Similar to dashes, colons introduce new EDU segments even the fragment occurs after the colon is a word or phrase, "as long as the text that follows the colon provides further elaboration on the topic introduced by the colon" (Carlson et al. 2003). In other word, when it is not adnominal, we segment them.

44. ( 又 翻译 作 : ||  
(also translated as : ||  
雅可比 )  
Jacobi )  
source: gum\_zh\_bio\_galois

45. 英语 中 主要 分为 三个 “态” : ||  
*English in mainly divide-into three CL "states": ||*  
主动态 , 中动态 和 被动态 。  
*active-voice , middle-voice and passive-voice .*  
source: gum\_zh\_academic\_iconicity

Exceptions apply when the nominal phrase after the colon is adnominal:

46. 也 有些 学者 认为 是 骨骼 的 发育不良 ||  
also some scholar think COP skeleton DE dysplasia ||  
[ 19 ] : pp. 3-4 。  
[ 19 ] : pp. 3-4 。  
source: gum\_zh\_bio\_byron

However, phrases separated by semicolons are not separate EDUs. The following example is a single long EDU.

47. 1月 25日 14时 , 封闭 沪渝 高速 黄石 ;  
January 25th 14:00 , close Shanghai-Chongqing Expressway Huangshi;

大广 高速 黄石西、 大冶 金湖、 阳新 龙港；  
*Daguang Expressway Huangshi West, Dayu Jinhu, Yangxin Longgang;*  
 杭瑞 高速 阳新 枫林、 木港、 排市；  
*Hangrui Expressway Yangxin Fenglin, Mugang, Paishi;*  
 黄咸 高速 大冶 陈贵、 灵乡、 金牛 共 10 个出口 ||  
*Huangxian Expressway Daye Chengui, Lingxiang, Jinniu total 10 CL exits ||*  
 source: gum\_zh\_news\_hubai

In the following example, semicolons with enumerations also do not create new EDUs as long as they graphically reside in the same sentence.

48. 英语 中动态 具有 如下 特点： ||  
*English middle-voice have following characteristics : ||*  
 1) 非 事件性； 2) 泛指性； 3) 施动性； 4) 情态 概念 ||  
 1) non eventuality ; 2) generality; 3) agency; 4) modal concept ||  
 [2]。 ||  
 [2]。 ||  
 source: gum\_zh\_academic\_iconicity

### 1.1.9 Strong discourse cues

As stated in the RST-DT manual, "phrasal expressions that occur with strong discourse cues are marked as EDU". In this Chinese corpus, we categorize and exemplify the following Chinese tokens or phrases as strong discourse cues. When making decisions regarding whether specific tokens are discourse cues or not, we make reference to Explicit Connectives annotated in the the PDTB-styled Chinese Discourse Treebank (CDTB) and TED Chinese Discourse Treebank (TED-CDB).

#### Adversarial Discourse Markers

- 尽管 although
- 虽然 although
- 不管是 no matter what/how
- 除了 apart from
- 除 apart from
- 但 but
- 但是 but
- 可是 instead
- 此外 besides
- 然而 however

#### Attributional Discourse Markers

- 根据 according to
- 据 according to
- 按照 according to
- 按 according to

- 依照 according to
- 基于 based on

### Causal Discourse Markers

- 因为 because (of)
- 所以 so
- 因 because (of)
- 由于 due to
- 基于 because of
- 经过 as a result of

### Circumstantial Discourse Markers

- 如果 if
- ... 的话 in the case of ...
- 随着 along with
- 通过 by means of
- 透过 through
- 经过 through
- ... 时 at time of ...
- ... 情况下 under the circumstance of ...
- ... 状态下 under the situation of ...
- ... 过程中 in the process of ...
- ... 的时候 at time of ...
- ... 的同时 at the same time of ...

### Coordinating Discourse Markers

- 不论 regardless of
- 无论 regardless of
- 不但 not only
- 不仅 not only
- 而且 but also
- 还是 instead
- 并且 in addition
- 并 at the same time
- 越...越... the more... the more...

### Elaborating Discourse Markers

- 针对 regarding

### Topic Discourse Markers

- 对(于) ... 来说 as far as...concerned (when taking an complement)
- 对(于) ... 而言 as far as...concerned (when taking an complement)
- 从 ... 来看 from the view of ... (when taking an complement)

49. 而且越 多 ||

*and the-more more ||*

越 适得其反 。

*the-more backfire ||*

source: gum\_zh\_whow\_flirt

50. 薄荷油 是 天然 的 驱逐剂 , ||

*peppermint COP nature DE repellant ||*

对 啮齿类 动物 来说 ||

*to rodent animal regard ||*

太 刺激 , ||

*too irritating ||*

source: gum\_zh\_whow\_mice

51. 过 一 会 再 想 一 个 好 玩笑 ||

*after a-while again think-of one CL good joke ||*

"after a while, think of another good joke" ||

发 过去 , ||

*send to-there , ||*

"and send it over" ||

总 比 你 弄 得 对 方 不 自 在 , ||

*at-least than you make the-other-one NEG comfortable , ||*

"at least that would be better than making the other one uncomfortable" ||

然后 又 试 图 给 自 己 解 释 ||

*then again try GEI self explain ||*

"and then trying to explain yourself" ||

要 容 易 得 多 。

*COP easy DER much . ||*

"much easier . "

source: gum\_zh\_whow\_flirt

For reference, the followings are strong discourse markers in English RST-DT:  
because, in spite of, despite, regardless, irrespective, without, according to, as a result of, not  
only ... but also.

### 1.1.10 Translanguaging

When translanguaging happens, especially between English and Chinese, we take the English portions as a fixed block and merge them into Chinese syntax. As a result, we DO NOT segment the English phrases.

**52. Max hit Harry and Harry hit Max 表示 //**

indicate

**Max hit Harry** 在先，//

at-first

**Harry hit Max** 在后。//

at-last

source: gum\_zh\_academic\_iconicity

**53. Sammy 's mad and I 'm glad 和 He comes , I will stay 。//**

and

source: gum\_zh\_academic\_iconicity

Similarly, foreign book titles should be blocks, thus not segmented.

**54. “珍娜·玛柏” 这 频道 成名 于《 How to trick people**

" Jenna Marber " this channel become-famous at

**into thinking you 're good looking 》和《 How To Avoid Talking To People and You Do n't Want To Talk To 》这 2 部 视频 ， //**

this 2 CL video , //

其中《 How to trick people into thinking you're good looking 》在 //

among at //

上传 后 的 //

upload after DE //

第一 周 便 录得 超级 530 万 次 的 观看数 //

first week already accept super 5.3M CL DE view //

[ 13 ] //

[ 14 ] . //

而《 How To Avoid Talking To People You Do n't Want To Talk To 》

and

则 于 2011 年 8 月 分别 被 《 纽约 时报 》 //

then at 2011 Aug separately BEI New York Times //

[ 15 ] //

和《 ABC 新闻 》 //

and ABC News //

[ 16 ] //

报导 。

report . //

source: gum\_zh\_bio\_marble

### 1.1.11 Stranded left-side tokens

Due to pre-verbal modification in Mandarin Chinese, we often see some small segments stranded on the left side of a sentence. We segment all these stranded spans and they form a *same-unit* relation with the discontinuous right-side span.

These examples include (but not limited to):

- **Adverb || adjunct clause || main clause**
- e.g. "However, || because he likes CS, || John went to CMU."
- **Subject || adjunct clause || main clause**
- John, || because he likes CS, || went to CMU.
- **PP || adjunct (PP or clause) || main clause**
- In the summer, || because he likes CS, || John will go to CMU.

## 1.2 Not Segmented Unit

|| is used to indicate segmented EDU boundaries.

### 1.2.1 Complement clauses

**Complement clauses are not segmented**, for example, clausal subjects and objects.

55. ... 甚至 更 让 考官 恼怒 的 是 , 他 将 ||  
"... even more make examiner angry DE COP , he BA ||  
擦 黑板 的 ||  
erase black-board DE ||  
抹布 扔在 了 考官 的 脑袋 上 || ...  
rag throw-at LE examiner DE head on-top-of || ..."||  
source: gum\_zh\_bio\_galois

56. ... 求解 复合 运动 经常 需要 把 运动 分解  
"... solve compound movement usually need BA movement decompose  
为 平移 和 转动 。 ||  
to-be translation and rotation ."  
source: gum\_zh\_bio\_bernoulli

57. ... || 这 两 方面 的 原因 使得 他 形成 了  
" .. || this two aspect DE reason make he form LE  
孤僻 和 忧郁 的 性格 。 || [ 1 ] ||  
solitary and melancholy DE personality . || [ 1 ] " ||  
source: gum\_zh\_bio\_byron

**Subject clauses are not segmented in Chinese, not even for coordinated subject clauses**

58. 因此 , 研究 老年人 言语 感知 特点  
therefore , study elderly speech perception characteristics  
"Therefore , studying the characteristics of speech perception in the elderly"  
和 探索 老年人 言语 感知 策略 , 能 ||  
and explore elderly speech perception strategy , can ||  
"and exploring speech perception strategies in older adults, can" ||  
为 提高 老年人 言语 感知 能力 ||  
for improve elderly speech perception ability ||  
"for improving the speech perception ability of the elderly ||"  
提供 参考 , ||  
provide reference , ||  
"provide a reference, ||"

也能 对老年人 言语感知 障碍 的 临床诊断 治疗、  
also can for elderly speech perception disorder DE clinical diagnosis treatment ,  
老年人 助听器 的 研发 提供 新思路， ||  
elderly hearing-aids DE research provide new idea , ||  
"It can also provide new ideas for the clinical diagnosis and treatment of speech  
perception disorders in the elderly and the research of hearing aids for the elderly," ||  
对于 促进 老年人 与 他人 之间的 交流 沟通 ||  
for promote elderly with others in-between DE communication communication ||  
"for the promotion of communication between the elderly and others" ||  
有 着 重要 作用 。 ||  
have PART important role . ||  
"has an important role." ||  
source: gum\_zh\_academic\_aging

#### 59. Halliday & Matthiessen ||

[ 1 ] ||  
认为 ||  
believe ||  
主位 表达 旧 信息 ，  
theme express old information ,  
"themes expressing old information,"  
述位 表达 新 信息 是 非 标记性 信息 匹配 结构 ; ||  
rheme express new information COP non- marking information match structure ||  
"rheme expressing new information is a non-marking information matching structure;" ||  
source: gum\_zh\_academic\_iconicity

On the other hand, coordinated object clauses are segmented under 2 conditions: the verb  
is an attribution verb and the objective clauses do not share the same subordinated  
object.

### 1.2.2 Prepositional phrases

**Prepositional phrases are not segmented.**

60. ... || 在 流体 力学 和 空气 动力学 中 有 关键性 的 作用 。  
"... || in fluid mechanics and aero dynamics within have critical DE effect ."||  
source: gum\_zh\_bio\_bernoulli

However, when a preposition heads a clausal complement, the phrase is separated from  
others.

61. 他 ||  
*he*  
对 修改 版权法 , ||  
*towards amend copyright-law ||*  
使 文件 共享 合法化 ||  
*make document sharing legalize ||*  
持 开放 态度  
*hold open mind*  
source: gum\_zh\_interview\_falkvinge

62. 这样 能 避免 你 ||  
*this can avoid you ||*  
被 忙碌 的 日程 压得 ||  
*BEI busy DE schedule squeeze DER ||*  
喘 不 过 气。||  
*breath NEG PAST breath ||*  
[ 16 ]  
[ 16 ]  
source: gum\_zh\_whow\_procrastinating

### 1.2.3 Dislocated NPs

Topicalization happens quite commonly in Chinese. However, dislocated NPs are not adpositions.

63. 资讯 安全 , 有 无意 与 恶意 的 攻击者 ,  
information security , has unintentional and malicious DE attackers ,  
要 怎么 去 阻止 , ||  
need how go stop ||  
"How can information security stop unintentional and malicious attackers,"  
source: gum\_zh\_interview\_wimax

### 1.2.4 MSP

MSP is a special part-of-speech label in Chinese treebank, reserved for a small set of "other particles." In our segmentation task, when nominals outside the MSP phrase are arguments of the predicate with the MSP phrase, we treat them as a single segment. Among such MSP particles, the most common ones are 所 (suo), 而 (er), and 来 (lai).

64. 而 没有 明确 意识 到 ||

but no clear realize ASP ||

人 的 本 质 所 应 有 的 ||

human DEC nature MSP should have DEC ||

丰 富 内 涵 , ||

rich connotation ||

"without clearly realizing the rich connotation that human nature should have"

source: gum\_zh\_academic\_socialized

65. 许 多 乌 克 兰 人 都 能 讲 述 ||

many Ukrainians all can tell ||

自 己 祖 辈 在 大 饥 荒 中 所 经 历 的 ||

selves' ancestor at great famine in MSP experience DEC ||

血 泪 历 史 。 ||

blood-and-tear history . ||

"Many Ukrainians can tell the Blood and Tears History that their ancestors experienced in the great famine."

source: gum\_zh\_news\_famine

66. 但 很 可 能 所 有 他 的 声 明 和 行 为 都 是 ||

but very possible all 3SG DEC statement and behavior all COP ||

对 于 贫 穷 的 压 力 而 产 生 的

toward poverty DEC pressure MSP arise DEC ||

富 于 创 意 的 ||

be-full-of creation DEC ||

反 应 。 ||

reaction . ||

"But it's likely that all his statements and actions are reaction that arose from the pressure of poverty and that are creative."

source: gum\_zh\_bio\_emperor

67. 我 们 如 何 从 象 似 性 角 度 来 分 析 它 呢 ?

1PL how from iconographic angle MSP analyze 3SG.INANIMATE PART ?

"How can we analyze it from an iconographic point of view ?"

source: gum\_zh\_academic\_iconicity

On the other hand, two other MSP tokens – 以(yi) and 去(qu) – connect two clauses where the latter expresses the purpose of the former. In these cases, we segment them and draw a backward purpose-goal relation.

68. 获 得 更 多 的 选 票 ,

get more many DEC vote , ||

以 确 保 我 们 不 会 在 选 举 日 用 完 。 ||

**MSP** ensure 1PL NEG will at election-day run-out . ||  
" to get more votes to make sure we don't run out on election day."  
source: gum\_zh\_interview\_falkvinge

69. 伽罗瓦 使用 群论 的 想法 ||  
Galois use group DEC idea ||  
去 讨论 方程式 的 可解性 , ||  
**MSP** discuss equation DEC solvability , ||  
"Galois uses the idea of group to discuss the solvability of equations ,"  
source: gum\_zh\_bio\_galois

### 1.2.5 Coordinations

**Multiple verbs with the same explicit object or prepositional complement are not segmented.**

70. ... 他们 同时 参加 并 试图 获得 巴黎 大学 的  
"... they simultaneously participate-in and attempt-to win Paris University DE  
科学 竞赛 的 第一名 || ...  
science competition DE first place || ..."  
source: gum\_zh\_bio\_bernoulli

71. 永远 不 要 试图 接近 或者 跟 野生 动物 进行 互动 ,  
never NEG should attempt approach or with wild animal conduct interact  
"never try to approach or interact with wild animals,"  
source: gum\_zh\_whow\_hiking

These also include cases with 把(BA) or 被(BEI).

72. 那就 把 手机 放在 其它 房间 或者 直接 关机 。  
then just BA phone put-in other room or just turn-off .  
"Then put your phone in another room or just turn it off."  
source: gum\_zh\_whow\_procrastinating

**Copula-less coordinated nominal or adjectival phrases are not segmented when conjoined with other copula-ed propositions.**

Examples:

73. 生 于 荷兰 格罗宁根 , 著名 数学家 , 约翰·伯努利 之 子 , ||  
*born in Netherlands Groningen , famous mathematician, John-Bernoulli 's child, ||*

为 伯努利 家族 代表 人物 之一 。  
being Bernoulli family representative person one-of ." |||  
source: gum\_zh\_bio\_bernoulli

74. 字 宜仲 , || 生 于 天津 , 江苏 阳湖人 , 语言学家 , |||  
*style-named Yizhong, || born in Tianjin, Jiangsu Yanghu-nese, Linguist, |||*  
精研 北方话 与 吴语 方言 的 音系 。 |||  
*intensively-study Northern-dialect and Wu dialect DE phonology . " |||*  
source: gum\_zh\_bio\_chao

### 1.2.6 Existential clauses

Simple clause 有(you)-constructions are formed by Locative NP + Existential Verb + Object NP. Moreover, the complement of 有 can also be a clause. In this case, the locative NP, existential you and object clause together forms one EDU.

75. 只要 路上 有 其他 人 同行 , |||  
as-long-as on-the-road EXIST other people travel-together , |||  
"As long as there are other people on the road,"  
source: gum\_zh\_whow\_hiking

76. 抖音 中 , 有 42.1% 的 视频 是 关于 普通 人  
douyin within , EXIST DE video COP about normal people  
在 疫情 期间 的 抗疫 生活 。 |||  
at epidemic time DE anti-epidemic life . |||  
"In Douyin , 42.1% of the videos are about ordinary people's anti-epidemic life during the epidemic."  
source: gum\_zh\_academic\_peoples

If you are interested in Chinese existential, feel free to check out this work: "Topic-comment and the Chinese Existential Construction"  
([https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-23\\_2\\_04.pdf](https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-23_2_04.pdf))

### 1.3 Compare & Contrast:

### 1.3.1 Tokenization matters

when 还有 is one token, it is CC between nominals, not an existential verb, so it does not create a new EDU.

77. 要 随身 携带 急救箱，还有 手机。||  
*should with-you carry first-aid-kit , and cell-phone . ||*  
"You should carry a first aid kit with you and your cell phone ."  
source: gum\_zh\_whow\_hiking

Faithfulness to main-subordinating clause distinction in syntax and nuclearity-satellite distinction in RST are more important than creating extra same-units.

78. 前往 帕劳 的 ||  
go-to Palau DE ||  
"that traveled to Palau" ||  
旅客， ||  
traveler , ||  
"travelers"  
在 结束 5 天 或 7 天 的 行程 后 ||  
at finish 5 day or 7 day DE itinerary after ||  
"after finishing their 5-day or 7-day itinerary" ||  
返回 台湾， ||  
go-back Taiwan , ||  
"go back to Taiwan" ||  
source: gum\_zh\_news\_bubble

79. 中国 作为 世界 第二 大 的 电影 市场 , ?||好莱坞 的 制片厂 老板 希望 ||  
China as world second large DE movie market , ?|| Hollywood DE studio boss hope ||  
"China as the second largest film market in the world , bosses of Hollywood studios  
hope"||  
巩固 电影 在 中国 市场 的 前景 , ||  
consolidate movie at China market DE prospect ||  
"consolidate the prospects of films in the Chinese market ,"  
source: gum\_zh\_news\_five

### 1.3.2 The part-of-speech of some tricky tokens

Prepositions	为(wèi) for	作为 when expressing
由 by	截至 till	identity or property
以 as/by	靠 by	比起 compared to

沿着 along with  
随着 along with  
借由 by  
相对于 relative to

**Localizers**  
以来 up until

**Adverbs**  
特别是 especially  
尤其是 especially

例如 for example  
一起 together  
看起来 seems

**Verbs**  
利用 make use of  
为(wéi) COP  
相比 compared to  
像是 seems like  
伴随 go along with

作为 when expressing  
regarding sth/sb as (i.e.,  
can be replaced by 当作)  
一样 the same as (e.g., in  
像...一样)

## 2 Relation Annotation

### 2.0 Some annotation principles

#### 2.0.1 Relation marking for relative clauses

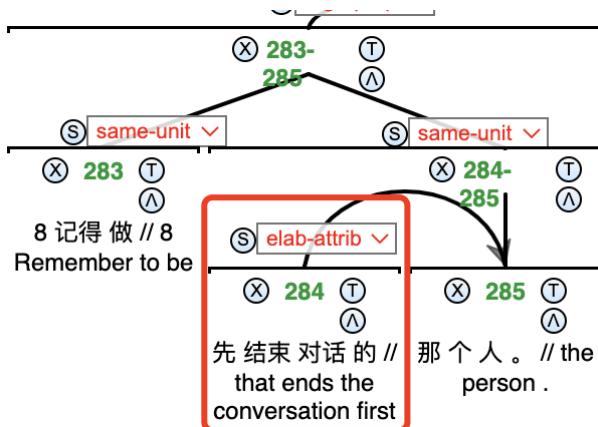
One of the largest differences between English and Chinese regarding the structure of an RST tree is the excess amount of the **combination of same-unit + elaboration-attribute** that are used to structure relative clauses in Chinese.

In the RST English GUM corpus, the only two attributional relations that modify part of a clause, usually a noun phrase, are elaboration-attribute and purpose-attribute. Most commonly, they are *elaboration-attribute*.

Another difference from English is that these elaboration-attribute relations are most frequently prenominal. This is due to the fact that **relative clauses in Chinese are prenominal**. There is not much previous research address prenominal relative clauses in RST. Our decision agrees with Shinmori et al. 2003 (see Figure 5-6 on page 7) who uses Elaboration for prenominal relative clauses in Japanese.

In the following example, we observe EDU\_284 "that ends the conversation first" breaks the main clause "remember to be ... the person." Following previous RST-DT guidelines, this creates a same-unit + elaboration-attribute structure for EDU\_283-285.

80. EDU\_283 8 记得 做 // 8 Remember to be  
EDU\_284 先 结束 对话 的 // **that ends the conversation first**  
EDU\_285 那个人。 // **the person**.  
source: gum\_zh\_whow\_flirt



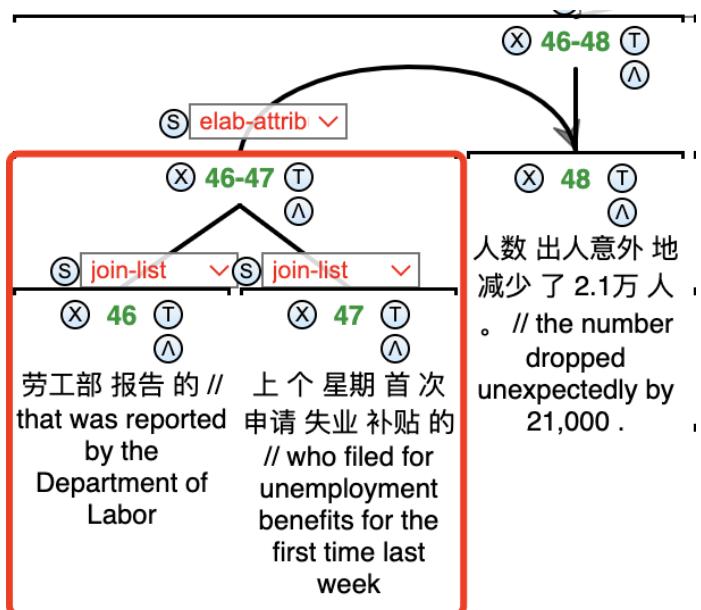
Moreover, multiple relative clauses can be coordinated to modify the same noun phrase, such as the EDU\_46 and EDU\_47 that modify EDU\_48 below.

81. EDU\_46 劳工部 报告 的 // **that was reported by the Department of Labor**

**EDU\_47** 上个星期首次申请失业补贴的 // **who filed for unemployment benefits for the first time last week**

**EDU\_49** 人数出人意外地减少了2.1万人。// **the number dropped unexpectedly by 21,000.**

source: gum\_zh\_news\_estate



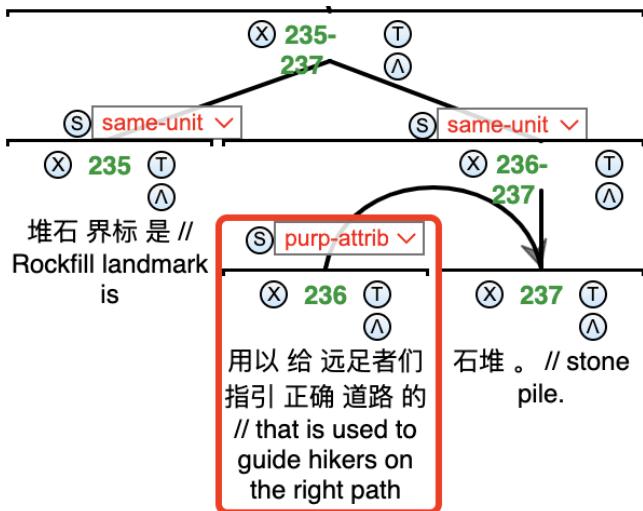
Purpose-attribute is the other attributive relation. In the following example, the stone piles are for the purpose of guiding the correct paths for hikers. In this case, we choose the label **purpose-attribute**.

82. EDU\_235 堆石界标是 // Rockfill landmark is

EDU\_236 用以给远足者们指引正确道路的 // **that is used to guide hikers on the right path**

EDU\_237 石堆。// **stone pile.**

source: gum\_zh\_whow\_hiking

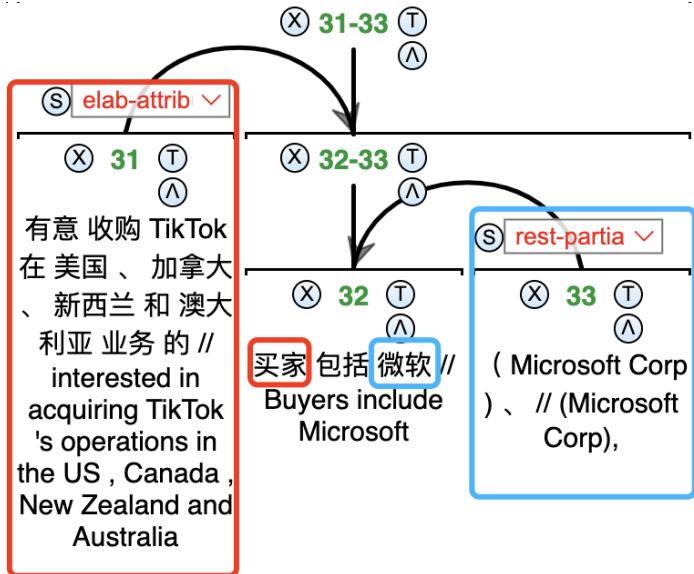


## 2.0.2 Attachment order of partial modification

In cases where the subject and object of an EDU are modified separated by two other EDUs, we attach the subject modifier higher than the object one, based on the syntactic hierarchy that subject governs object.

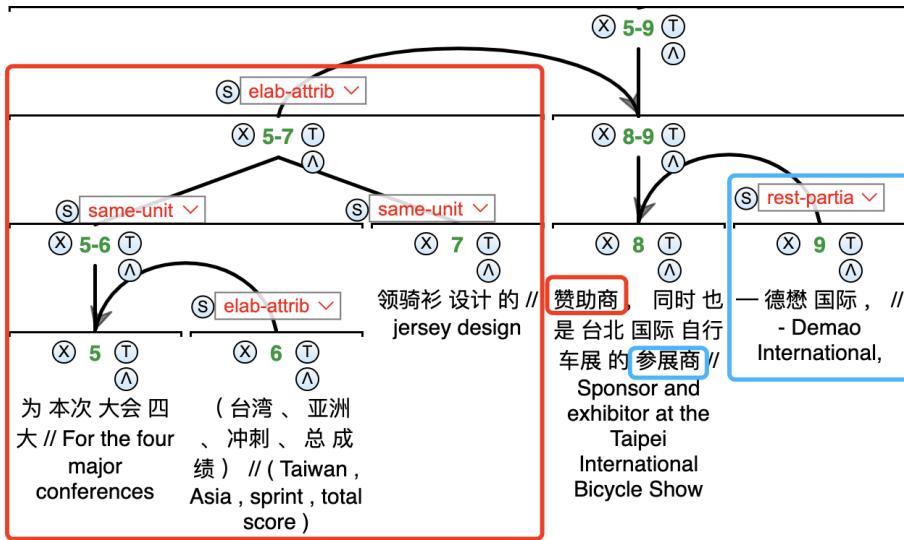
In the following example, the relative clause modifying the subject "买家 // Buyers" is attached higher than the parenthetical "( Microsoft Corp )" which modifies the object "微软 // Microsoft."

83. EDU\_31 有意 收购 TikTok 在 美国 、 加拿大 、 新西兰 和 澳大利亚 业务 的 //  
 interested in acquiring TikTok's operations in the US , Canada , New Zealand and  
 Australia ||
- EDU\_32 买家 包括 微软 // Buyers include Microsoft ||
- EDU\_33 ( Microsoft Corp ) 、 // (Microsoft Corp), ||
- source: gum\_zh\_news\_tiktok



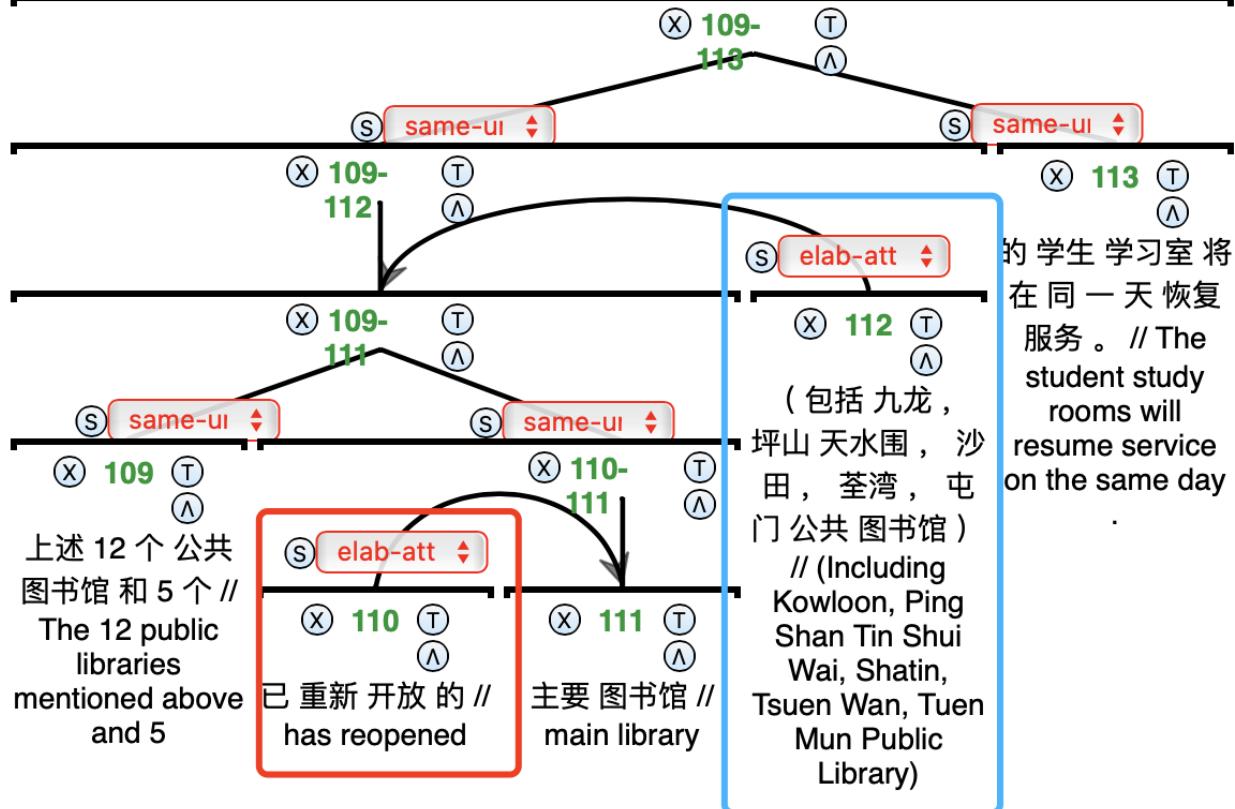
Similarly, modifier of "赞助商 // Sponsor" is higher than "参展商 // exhibitor" in the following example.

84. EDU\_5 为本次大会四大 // **For the four major conferences** ||  
     EDU\_6 (台湾、亚洲、冲刺、总成绩) // (**Taiwan, Asia, sprint, total score**) ||  
     EDU\_7 领骑衫设计的 // **jersey design** ||  
     EDU\_8 赞助商，同时也是台北国际自行车展的 参展商 // **Sponsor and exhibitor at the Taipei International Bicycle Show** ||  
     EDU\_9 –德懋国际，// - **Demao International**, ||  
     source: gum\_zh\_interview\_cycle



In the following example, we see two levels of *same-unit* + *elaboration-attribute* relations. At the lower level, EDU\_110 "has reopened" modifies "main library"; whereas the combination of 12 public and 5 main libraries "include Kowloon, Ping Shan Tin Shui Wai, etc." as in EDU\_112.

85. EDU\_109 上述 12 个 公共 图书馆 和 5 个 // The 12 public libraries mentioned above and 5
- EDU\_110** 已 重新 开放 的 // has reopened
- EDU\_111 主要 图书馆 // main library
- EDU\_112** (包括 九龙 , 坪山 天水围 , 沙田 , 荃湾 , 屯门 公共 图书馆 ) //  
(Including Kowloon, Ping Shan Tin Shui Wai, Shatin, Tsuen Wan, Tuen Mun Public Library)
- EDU\_113 的 学生 学习室 将 在 同 一 天 恢复 服务 。 // The student study rooms will resume service on the same day .
- source: gum\_zh\_news\_kangle

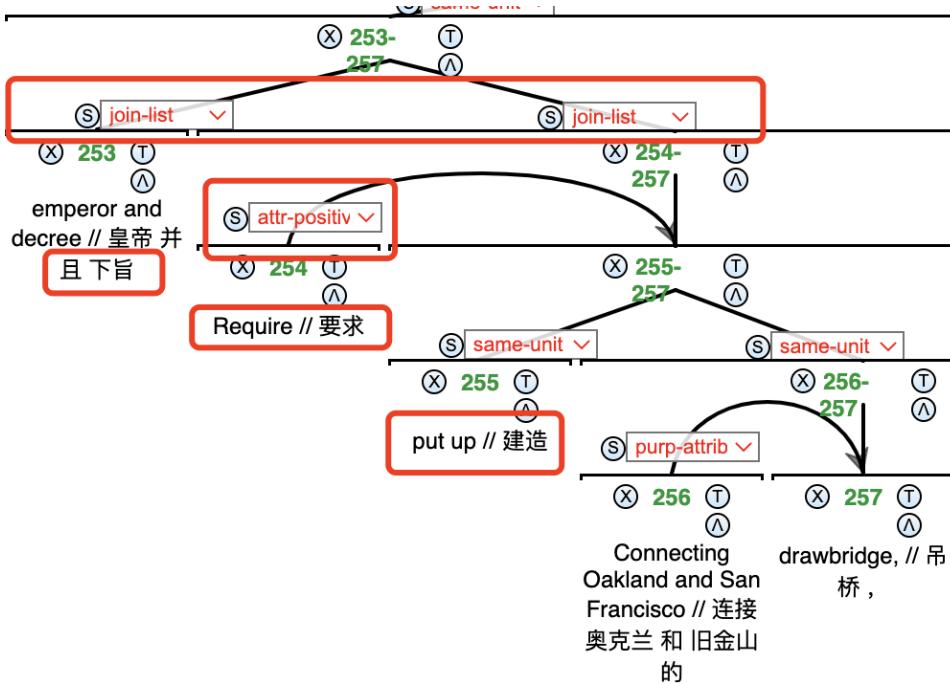


### 2.0.3 Implicit coordination

In Chinese, implicit verb phrase coordination is common, and they are annotated as conjunctions syntactically in the Chinese Treebank.

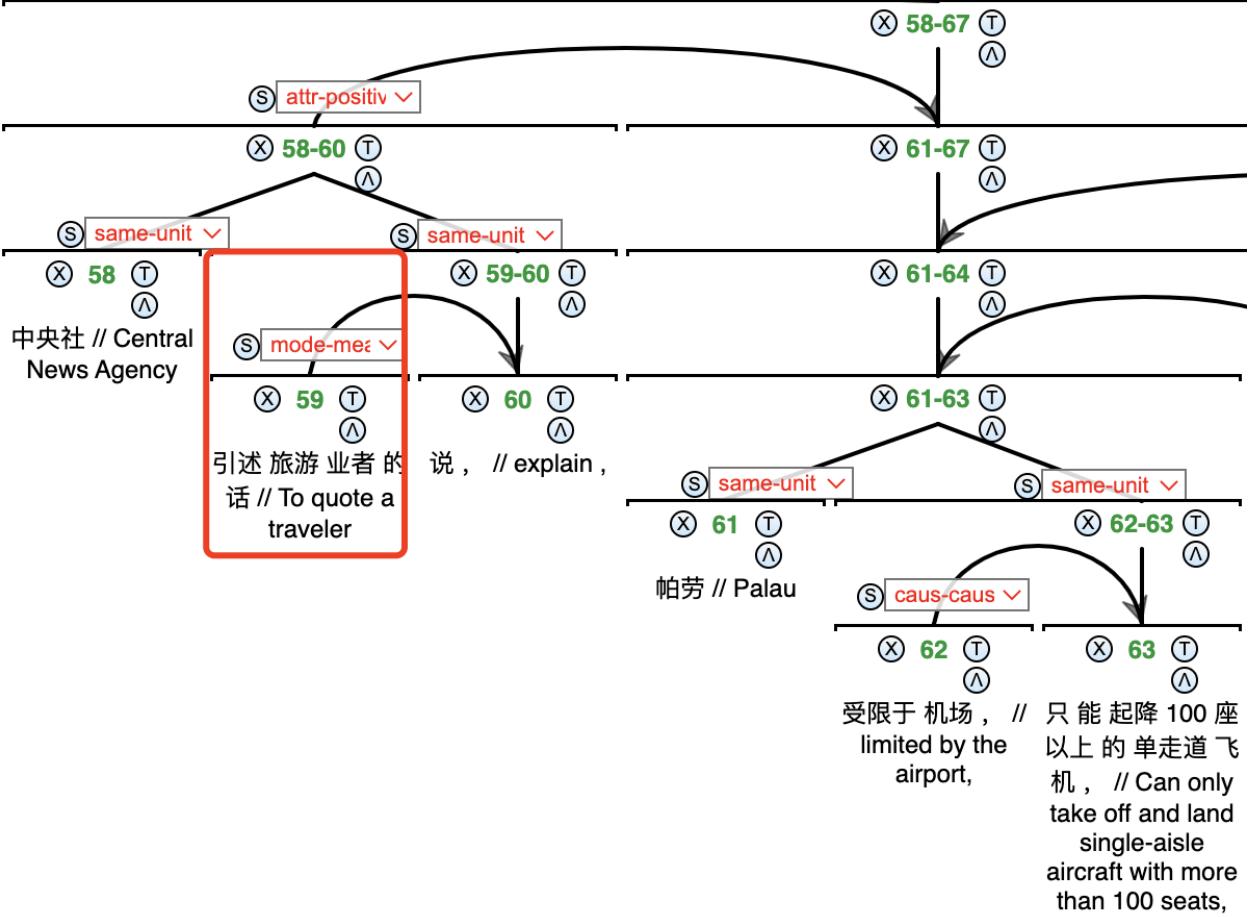
In this corpus, unless there is a significant nucleus-satellite imbalance between these implicitly coordinated verb phrases, they form a joint-list multinuclear relation as in the following example.

86. EDU\_253   皇帝 并且 下旨 // the emperor decrees  
               EDU\_254   要求 // requires  
               EDU\_255   建造 // put up  
               EDU\_256   连接 奥克兰 和 旧金山 的 // that connects Oakland and San Francisco  
               EDU\_257   吊桥 , // drawbridge,  
               source: gum\_zh\_bio\_emperor



In contrast, in the following example, the content of the saying event is more important than how the saying is addressed (by quoting a traveler). Thus, the verb phrases before 说 "say" are considered satellites.

87. EDU\_58 中央社 // Central News Agency  
 EDU\_59 引述 旅游业者 的话 // **To quote a traveler**  
 EDU\_60 说 , // **explain** ,  
 EDU\_61 帕劳 // Palau  
 EDU\_62 受限于 机场 , // limited by the airport,  
 EDU\_63 只能 起降 100 座 以上的 单走道 飞机 , // Can only take off and land single-aisle aircraft with more than 100 seats,  
 source: gum\_zh\_news\_bubble



## 2.1 Nucleus-Satellite Relations

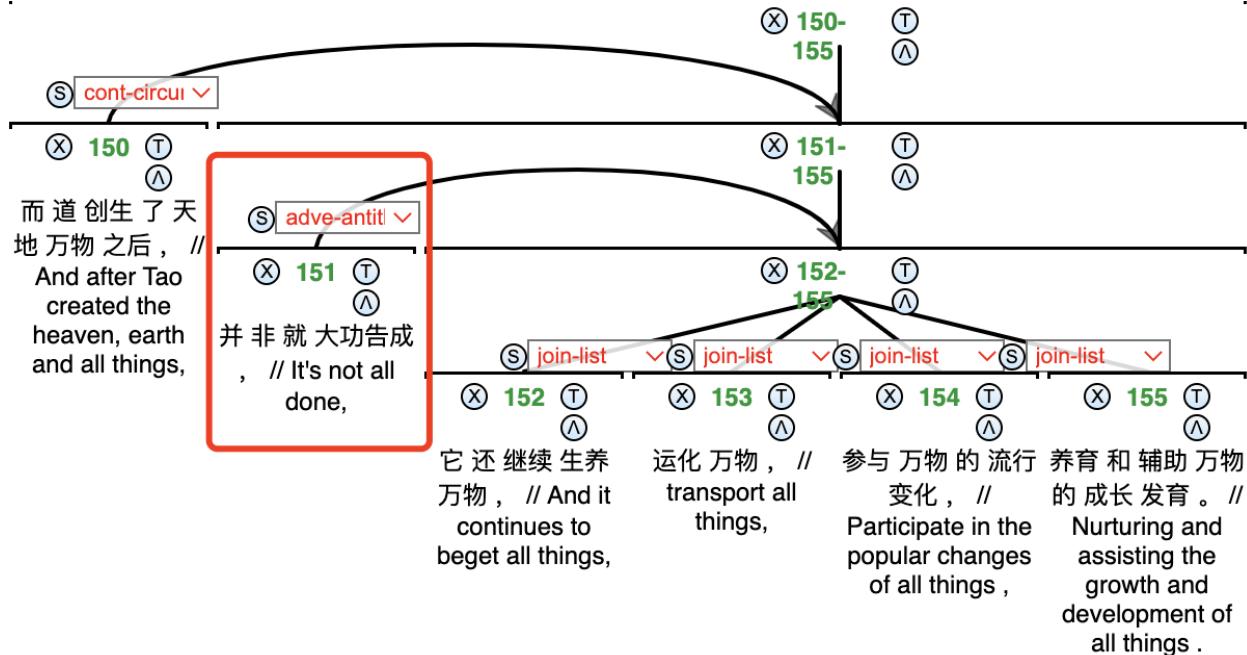
### 2.1.1 adversative-antithesis

**adversative-antithesis:** the Reader finds the Nucleus more credible than the Satellite.

For example, in the following example, "after Tao gives birth to the heaven and earth" (EDU\_150), it is not the case that "it is all done" (EDU\_151), rather it continues to take care of the world (DU\_152-155). Thus, EDU\_151 is an antithesis of EDU\_152-155.

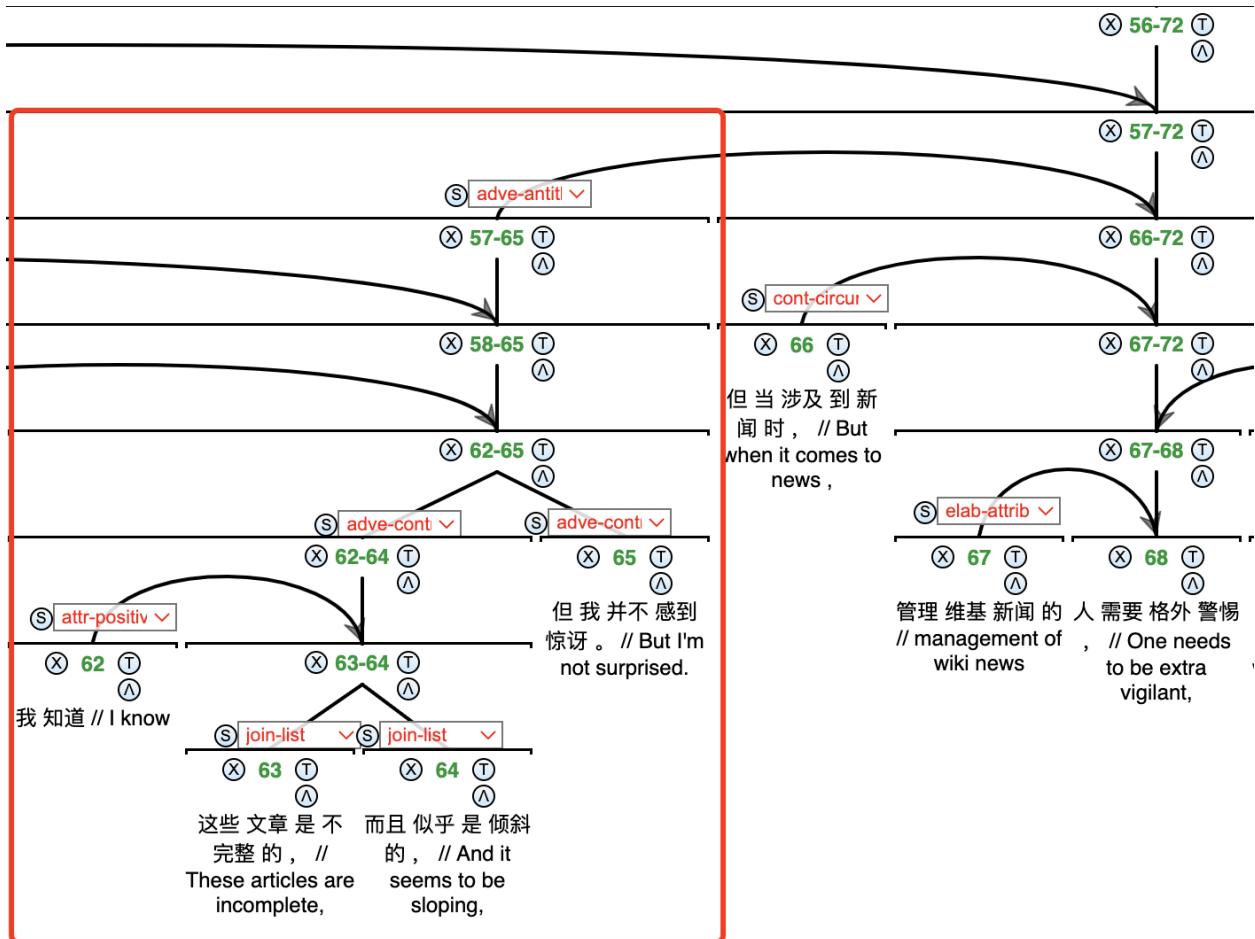
88. EDU\_150 而 道 创生 了 天地 万物 之后 , // And after Tao created the heaven, earth and all things,  
                   EDU\_151 并 非 就 大功告成 , // It's not all done,  
                   EDU\_152 它 还 继续 生养 万物 , // And it continues to beget all things,  
                   EDU\_153 运化 万物 , // transport all things,  
                   EDU\_154 参与 万物 的 流行 变化 , // Participate in the popular changes of all things ,

EDU\_155 养育 和 辅助 万物 的 成长 发育 。 // Nurturing and assisting the growth and development of all things .  
 source: gum\_zh\_academic\_taoist



In the following example, "when it comes news," the unsurprising in EDU\_65 is overridden by the "extra vigilant" in EDU\_68, thus we label DU\_62-65 → DU\_66-72 an adversative-antithesis.

89. EDU\_62 我 知道 // I know  
 EDU\_63 这些 文章 是 不 完整 的 ， // These articles are incomplete,  
 EDU\_64 而且 似乎 是 倾斜 的 ， // And it seems to be sloping,  
**EDU\_65 但 我 并不 感到 惊讶 。 // But I'm not surprised.**  
 EDU\_66 但 当 涉及 到 新闻 时 ， // But when it comes to news ,  
 EDU\_67 管理 维基 新闻 的 // management of wiki news  
 EDU\_68 人 需要 格外 警惕 , // One needs to be extra vigilant,
- source: gum\_zh\_interview\_ward



## 2.1.2 adversative-concession

**adversative-concession:** the Writer admits the Satellite, but still claims the Nucleus.

Concession is the more frequent Nucleus-Satellite adversative relation in GUM. It is usually the scenarios where one acknowledges the factuality of the Satellite but still stands for the Nucleus. Examples below show such preferences:

90. EDU\_82 传统的三角形球框也可以用来摆九球，// The traditional triangular ball frame can also be used to place nine balls,

EDU\_83 但是球之间的空隙比较大。// But the gap between the balls is relatively large .

source: gum\_zh\_whow\_pool



91. EDU\_29 虽然 仍 然 是 三 镜 头 设 计 , // Although it is still a three-lens design,  
 EDU\_30 但 每 颗 镜 头 都 有 明 显 的 进 步 。 // But each lens is a marked improvement .  
 source: gum\_zh\_news\_apple



### 2.1.3 attribution-negative

**attribution-negative:** the Satellite negates the source of information to the Nucleus.

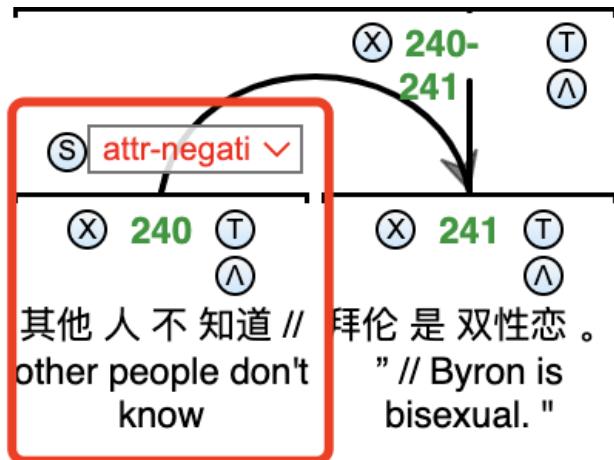
Attribution relations are essential to RST, where one addresses the content of the information more than its source. In other words, who said it is less important than what is said.

Like English GUM, we differentiate between a negative versus a positive source of information. A **attribution-negative** is when the source of information is negated like in the following example.

92. 其他 人 不 知 道 // other people don't know

拜伦 是 双性恋 。 " // Byron is bisexual. "

source: gum\_zh\_bio\_byron



#### 2.1.4 attribution-positive

**attribution-positive:** the Satellite provides a positive source of information to the Nucleus. See Section 1.1.4 for the list of attribution verbs.

On the other hand, we have more frequently attribution-positive as in the following example.

93. EDU\_247 有些 现代 的 医学家 认为 // Some modern medical scientists think that

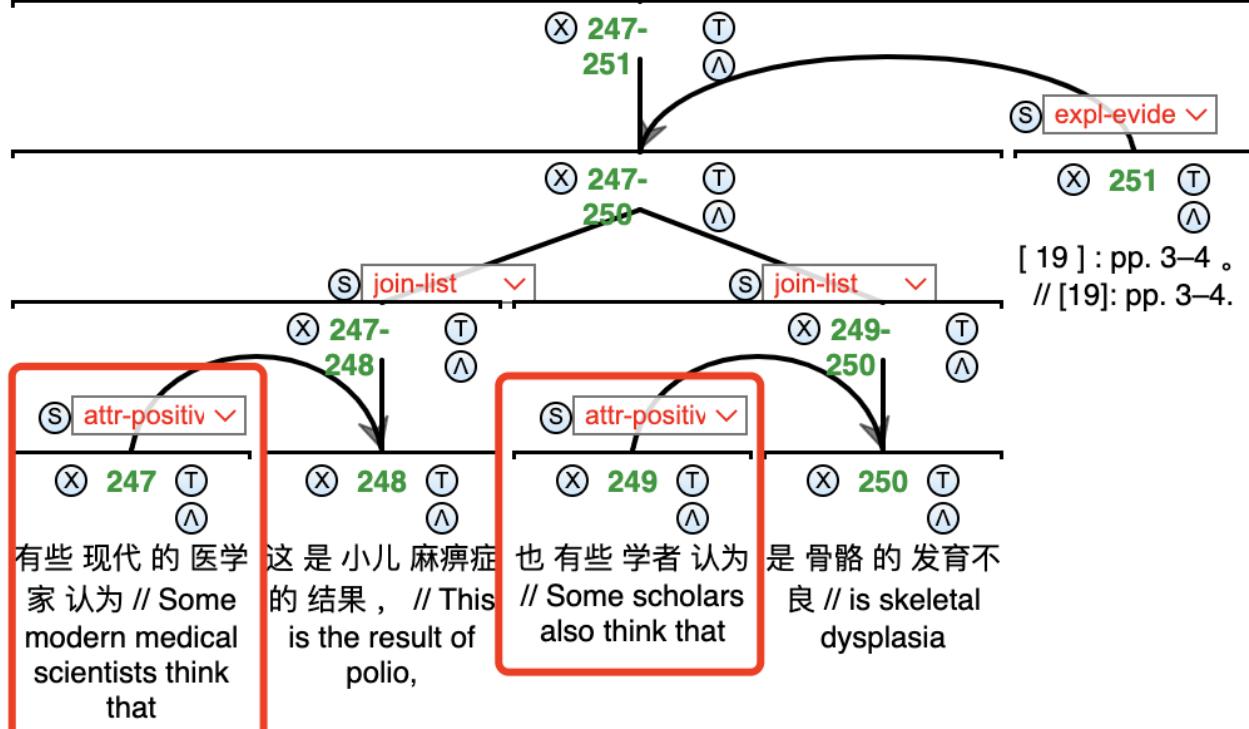
EDU\_248 这 是 小儿 麻痹症 的 结果 , // This is the result of polio,

**EDU\_249** 也 有些 学者 认为 // Some scholars also think that

EDU\_250 是 骨骼 的 发育不良 // is skeletal dysplasia

EDU\_251 [ 19 ] : pp. 3-4 。 // [19]: pp. 3-4.

source: gum\_zh\_bio\_byron



Also note that we still label cognitive verbs with negative emotions as attribution-positive since the source of information is not negated.

#### 94. 因为 他 害怕 // because he is afraid

乌克兰民族运动会与布尔什维克革命相竞争。// The Ukrainian National Games competed with the Bolshevik Revolution.

source: gum\_zh\_news\_famine



Note: the difference between *attribution-positive* and *explanation-evidence* is that the former emphasizes the **saying event** whereas the latter only provides the source of information.

## 2.1.5 causal-cause

**causal-cause: the Satellite causes the Nucleus.**

Causal relations are predominant in RST corpora. Causal-cause is used to label the less prominent cause that modifies the more prominent result.

As in the following example, "spreading the word" is more central than "not having illusion of winning".

95. EDU\_86 在这一 点 上 , 只 是 为 了 帮 助 宣 传 , // At this point, just to help spread the word,

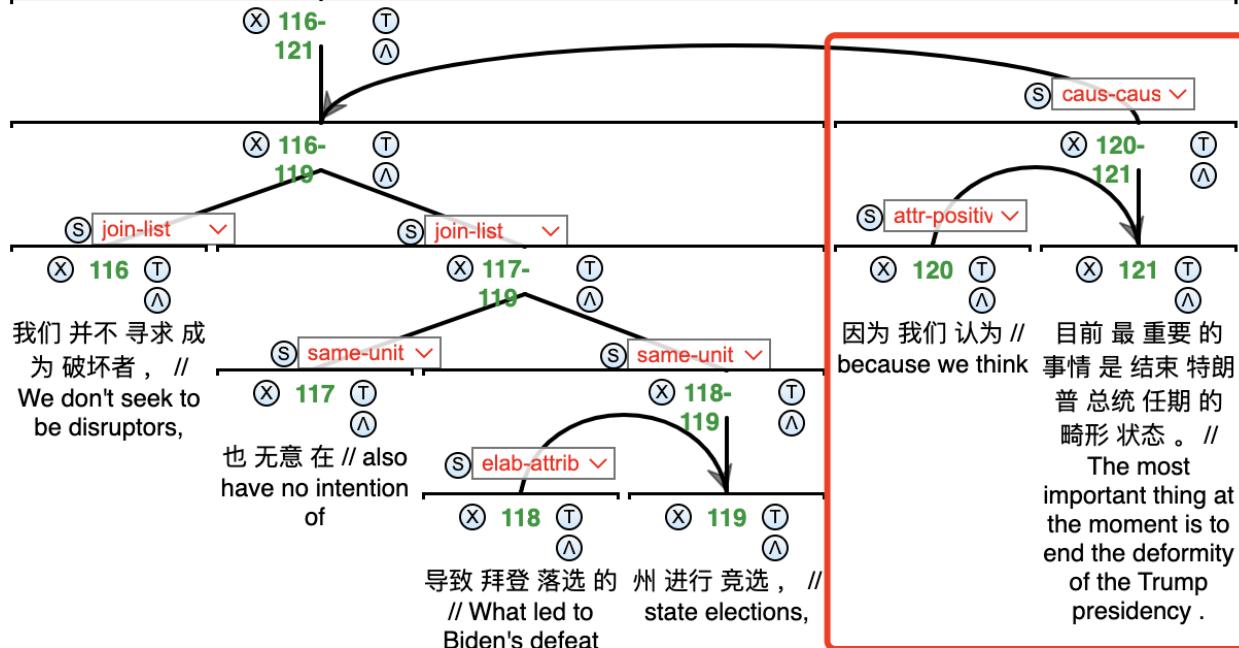
EDU\_87 因 为 我 们 并 不 抱 有 获 胜 的 幻 想 。 // Because we don't have the illusion of winning.

source: gum\_zh\_interview\_graaf



Similarly, "not being a disruptor and not winning" is the more central idea in the example below.

96. EDU\_116 我们 并不 寻求 成为 破坏者 ， // We don't seek to be disruptors,  
 EDU\_117 也 无意 在 // also have no intention of  
 EDU\_118 导致 拜登 落选 的 // What led to Biden's defeat  
 EDU\_119 州 进行 竞选 ， // state elections,  
**EDU\_120 因为 我们 认为 // because we think**  
**EDU\_121 目前 最 重 要 的 事 情 是 结束 特朗普 总统 任 期 的 畸 形 状 态 。 // The**  
**most important thing at the moment is to end the deformity of the Trump**  
**presidency .**
- source: gum\_zh\_interview\_graaf



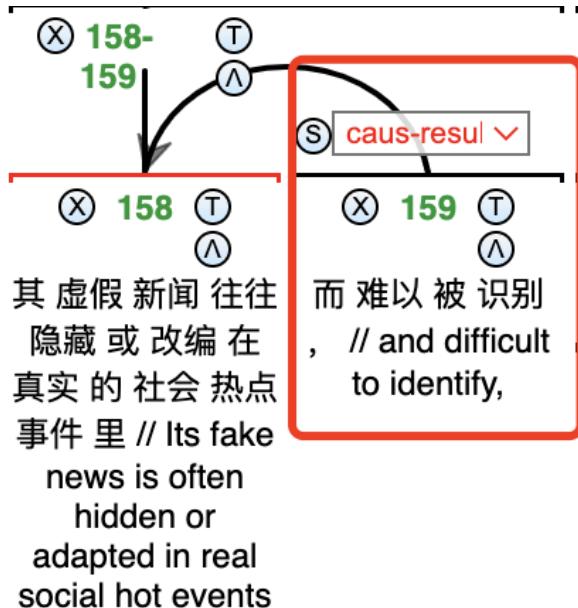
## 2.1.6 causal-result

**causal-result: the Satellite is the result of the Nucleus (inverse of cause).**

In these situations, the cause is more central than the result. For example, the "hiddenness" is more relevant in the context than the "difficulty to identify".

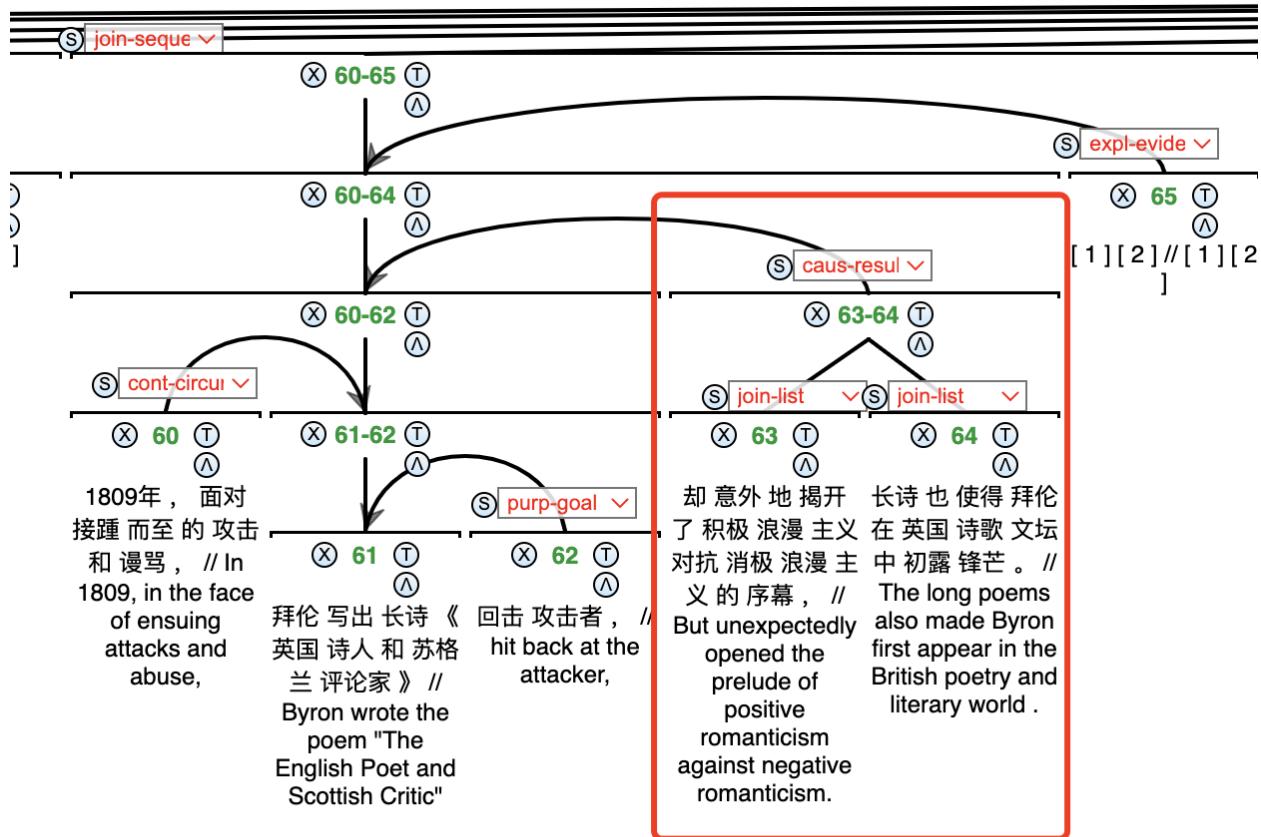
97. EDU\_158 其虚假新闻往往隐藏或改编在真实的社会热点事件里 // Its fake news is often hidden or adapted in real social hot events .

**EDU\_159 而难以被识别，// and difficult to identify,**  
source: gum\_zh\_academic\_supervision



In the following example, "Byron's writing long poem" is among a sequence of events at a higher RST structure, thus the result of his long poem is the satellite cause-result.

98. EDU\_60 1809年, 面对接踵而至的攻击和谩骂, // In 1809, in the face of ensuing attacks and abuse,  
 EDU\_61 拜伦写出长诗《英国诗人和苏格兰评论家》 // Byron wrote the poem "The English Poet and Scottish Critic"  
 EDU\_62 回击攻击者, // hit back at the attacker,  
 EDU\_63 却意外地揭开了积极浪漫主义对抗消极浪漫主义的序幕, // But unexpectedly opened the prelude of positive romanticism against negative romanticism.  
 EDU\_64 长诗也使得拜伦在英国诗歌文坛中初露锋芒。// The long poems also made Byron first appear in the British poetry and literary world .  
 EDU\_65 [1][2]//[1][2]  
 source: gum\_zh\_bio\_byron



## 2.1.7 context-background

**context-background: the Reader needs to know the Satellite to understand the Nucleus.**  
 The Satellite provides the context for the Nucleus and it is essential for the reader to know the satellite in order to understand the nucleus.

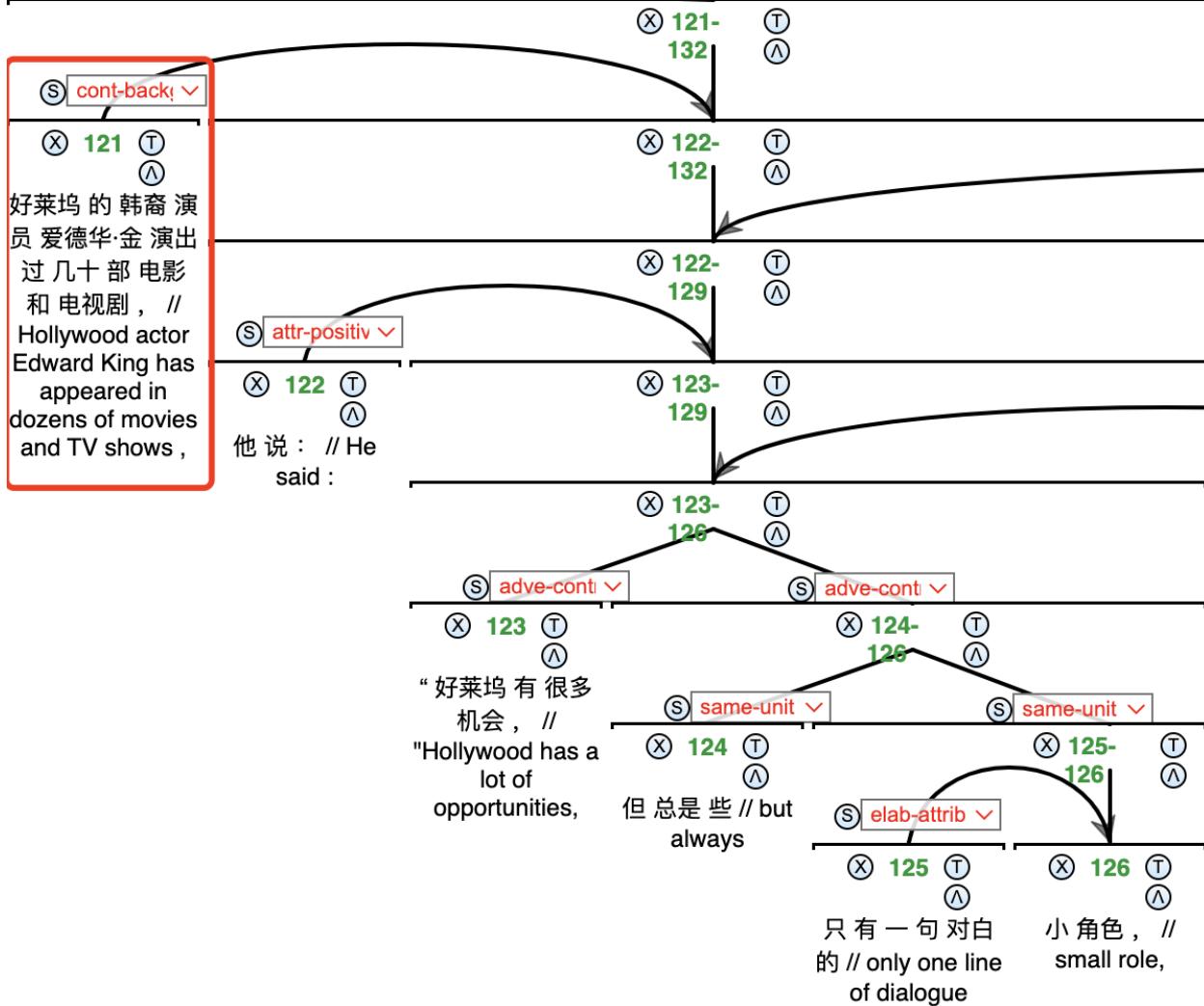
In the following example, the equivalent of English "besides" is a good example of context-background.

99. EDU\_74 除了直接问人家问题，// In addition to asking people questions directly,  
 EDU\_75 还可以给出你自己的看法。// You can also give your own opinion .  
 source: gum\_zh\_whow\_flirt



In this long example, Edward's participating in dozens of movies and shows sets up his reputation, and also makes what he is saying credential.

100. EDU\_121 好莱坞 的 韩裔 演员 爱德华·金 演出 过 几十 部 电影 和 电视剧 ， //  
**Hollywood actor Edward King has appeared in dozens of movies and TV shows ,**  
 EDU\_122 他 说 : // He said :  
 EDU\_123 “ 好莱坞 有 很多 机会 ， // "Hollywood has a lot of opportunities,  
 EDU\_124 但 总是 些 // but always  
 EDU\_125 只 有 一 句 对白 的 // only one line of dialogue  
 EDU\_126 小 角色 , // small role,  
 source: gum\_zh\_news\_five



## 2.1.8 context-circumstance

**context-circumstance: the Satellite gives circumstances, e.g., time, place, of the Nucleus.**

In Chinese, phrases such as 在...的时候 "at ... circumstance" are prototypical signals for marking context-circumstance. The following is an example of "在...时."

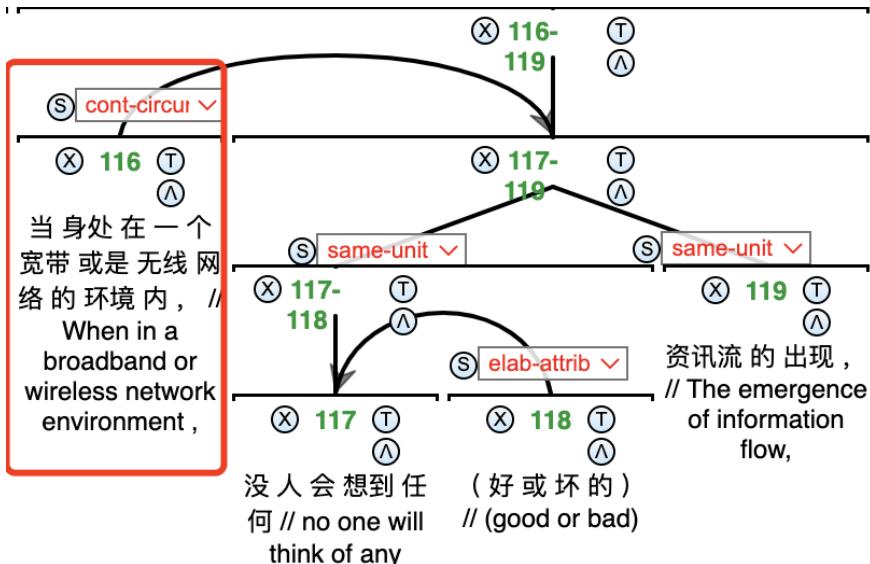
### 101. EDU\_292 在 对话 即将 失去 热度 时 // When the conversation is about to lose steam

EDU\_293 简短 地 说 再见 。 // Say goodbye briefly.

source: gum\_zh\_whow\_flirt



102. EDU\_116 当身处在一个宽带或是无线网络的环境内， // When in a broadband or wireless network environment ,
- EDU\_117 没人会想到任何 // no one will think of any
- EDU\_118 (好或坏的) // (good or bad)
- EDU\_119 资讯流的出现， // The emergence of information flow,
- source: gum\_zh\_interview\_wimax



## 2.1.9 contingency-condition

**contingency-condition: the Satellite is a condition for the Nucleus to happen.**

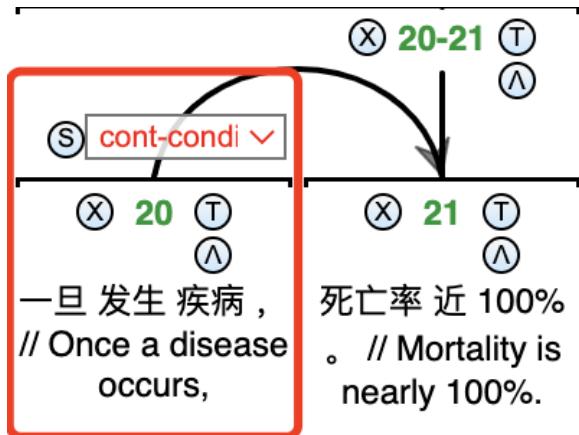
As stated in the RST-DT manual, "the truth of the proposition associated with the nucleus is a consequence of the fulfillment of the condition in the satellite."

In the following example, 一旦 "once" sets up a hypothetical condition where mortality can be nearly 100%.

103. EDU\_20 一旦发生疾病， // Once a disease occurs,

EDU\_21 死亡率近100%。 // Mortality is nearly 100%.

source: gum\_zh\_academic\_rabies



Similarly, the increase of retweets and comments is a condition for Douyin to provide traffic support in the following example.

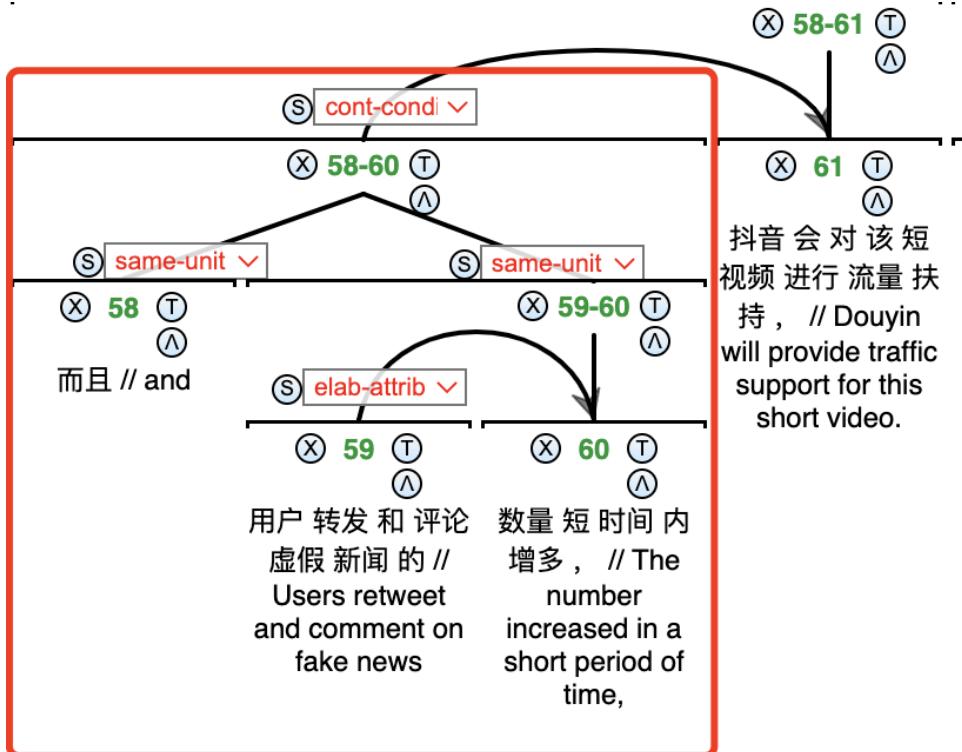
104. EDU\_58 而且 // and

EDU\_59 用户转发和评论虚假新闻的 // Users retweet and comment on fake news

EDU\_60 数量短时间增多， // The number increased in a short period of time,

EDU\_61 抖音会对该短视频进行流量扶持， // Douyin will provide traffic support for this short video.

source: gum\_zh\_academic\_supervision



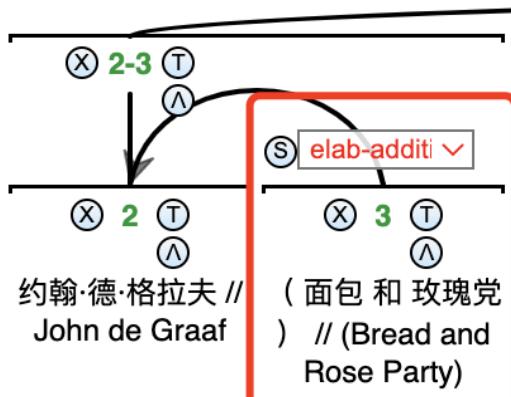
### 2.1.10 elaboration-additional

**elaboration-additional: the Satellite provides more information about the Nucleus.**

This is the most general "last resort" relation when the satellite gives more information about the nucleus. In practice, one annotates a nucleus-satellite relation as elaboration-additional only when other relations are not as suitable.

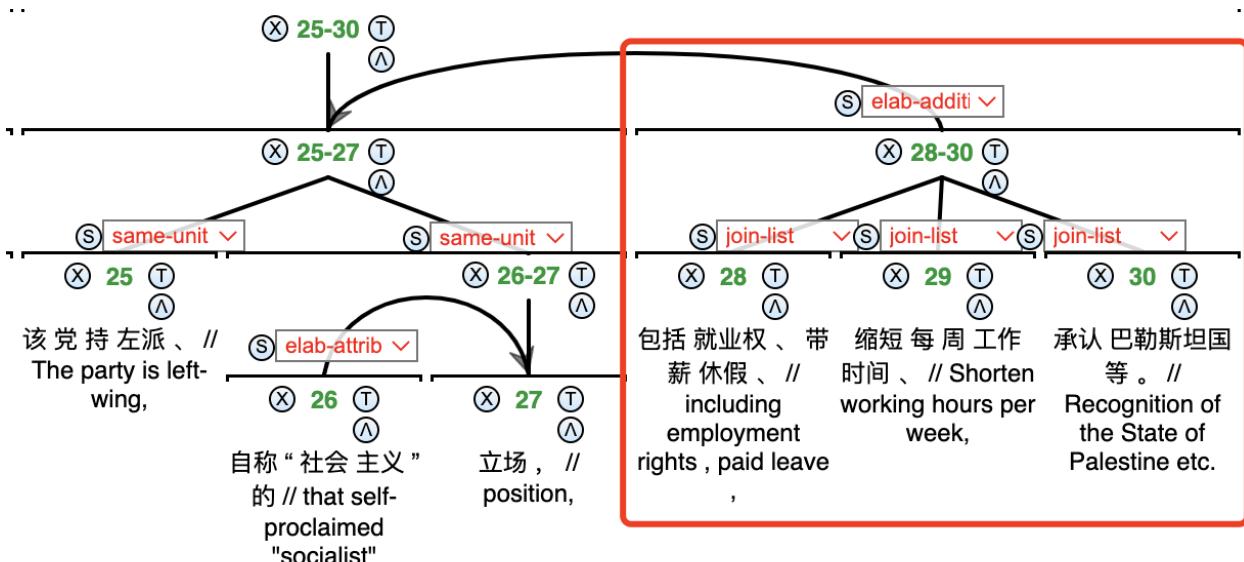
For example, the part-and-whole relation between John de Graaf and the Bread and Rose Party is annotated as elaboration-additional.

105. EDU\_2 约翰·德·格拉夫 // John de Graaf  
**EDU\_3( 面包 和 玫瑰党 ) // (Bread and Rose Party)**  
 source: gum\_zh\_interview\_graaf



More generally, when further explanations are made on a statement, we use the elaboration-additional relation. For example, the satellite in the example below explains what a "left-wing self-proclaimed socialist" means.

106. EDU\_25 该党持左派、 // The party is left-wing,  
       EDU\_26 自称“社会主义”的 // that self-proclaimed "socialist"  
       EDU\_27 立场， // position,  
       **EDU\_28 包括就业权、带薪休假、 // including employment rights , paid leave**  
       ,  
       **EDU\_29 缩短每周工作时间、 // Shorten working hours per week,**  
       **EDU\_30 承认巴勒斯坦国等。 // Recognition of the State of Palestine etc.**
- source: gum\_zh\_interview\_graaf



### 2.1.11 elaboration-attribute

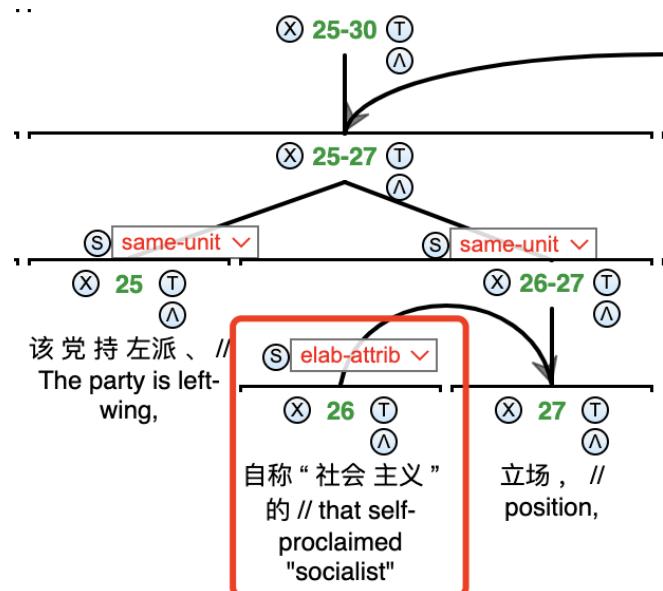
**elaboration-attribute:** the Satellite provides more information about some phrase (rather than the entire clause) in the Nucleus.

Functionally, the satellite provides the same detail to the nucleus. The major difference is that instead of modifying the whole clause of the nucleus, it only modifies a particular phrase, most frequently a noun phrase.

Since Chinese relative clauses are placed before the noun head, the structure of a higher *same-unit* and a lower *elaboration-attribute* is quite common in this dataset.

In the previous example, the phrase "that self-proclaimed socialist" modifies the noun phrase "position", instead of the clause (possession of a position).

107. EDU\_25 该党持左派、 // The party is left-wing,  
EDU\_26 自称“社会主义”的 // that self-proclaimed "socialist"  
EDU\_27 立场 , // position,  
source: gum zh interview graaf

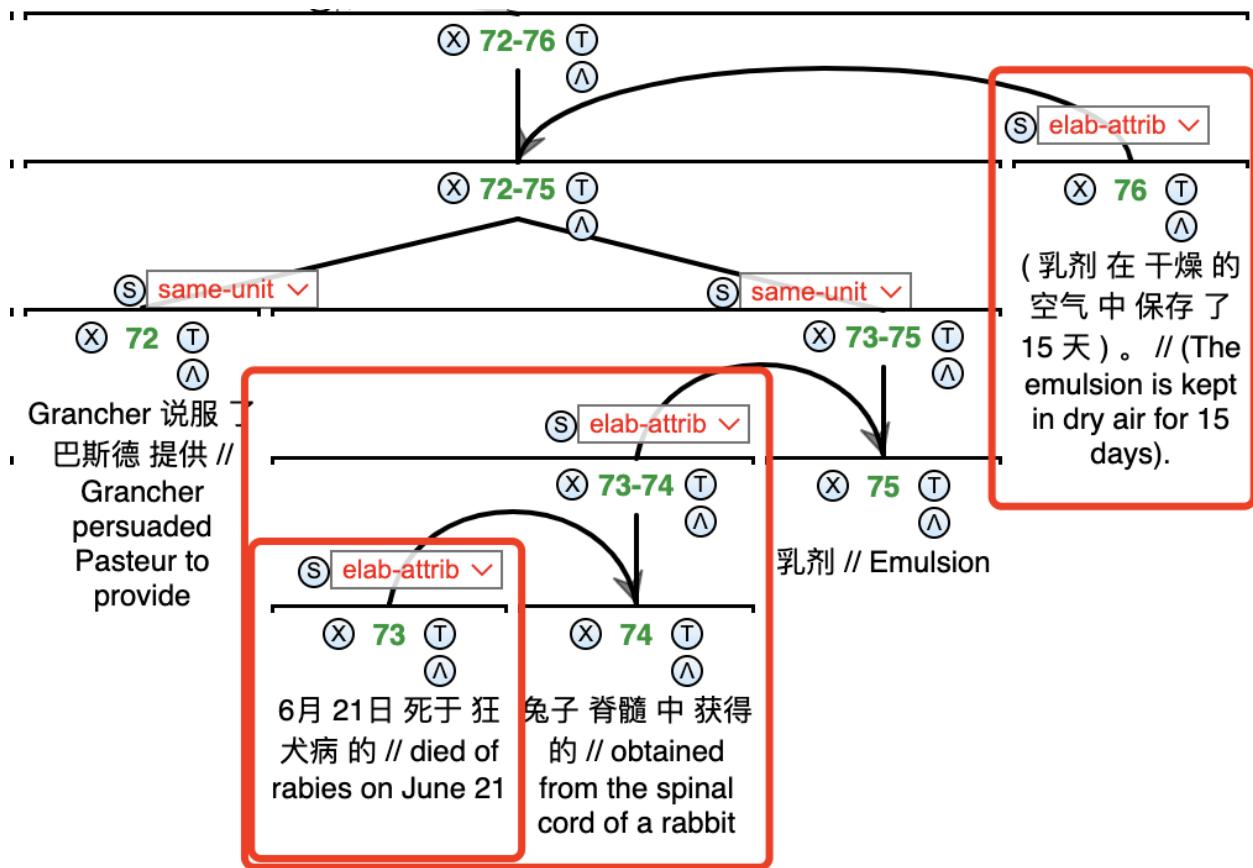


Elaboration-attribute can also label modifiers that do not surface as relative clauses. In the following example, we see three elaboration-attribute relations:

- the rabbit **that died of rabies on June 21**
  - the emulsion **that was obtained from the spinal cord of a rabbit**
  - emulsion was kept in dry air for 15 days

The first two are relative clauses but the third is parenthetical.

108. EDU\_72 Grancher 说服 了 巴斯德 提供 // Grancher persuaded Pasteur to provide  
 EDU\_73 6月 21日 死于 狂犬病 的 // died of rabies on June 21  
 EDU\_74 兔子 脊髓 中 获得 的 // obtained from the spinal cord of a rabbit  
 EDU\_75 乳剂 // Emulsion  
 EDU\_76 ( 乳剂 在 干燥 的 空气 中 保存 了 15 天 )。 // (The emulsion is kept in dry air for 15 days).  
 source: gum\_zh\_news\_rabies



### 2.1.12 evaluation-comment

**evaluation-comment** is a Nucleus-Satellite relation in which the Satellite gives an opinion about the Nucleus (that the Reader does not need to agree with).

In the following example, EDU\_177 evaluates the fact that Zhao Yuanren, as a distinguished professor, humbly asks his students about the Anhui dialect.

109. EDU\_176 赵元任 虚心 求教 安徽 方言 , // Zhao Yuanren humbly learn about the Anhui dialect,

EDU\_177 没有 老师 架子 。 // does not behave like a teacher at all.  
source: gum\_zh\_bio\_chao



When deciding between **attribution-positive** and **evaluation-comment**, the positive sentiment is more essential than the source of information, thus EDU\_29-31 → EDU\_32 is labeled **evaluation-comment**.

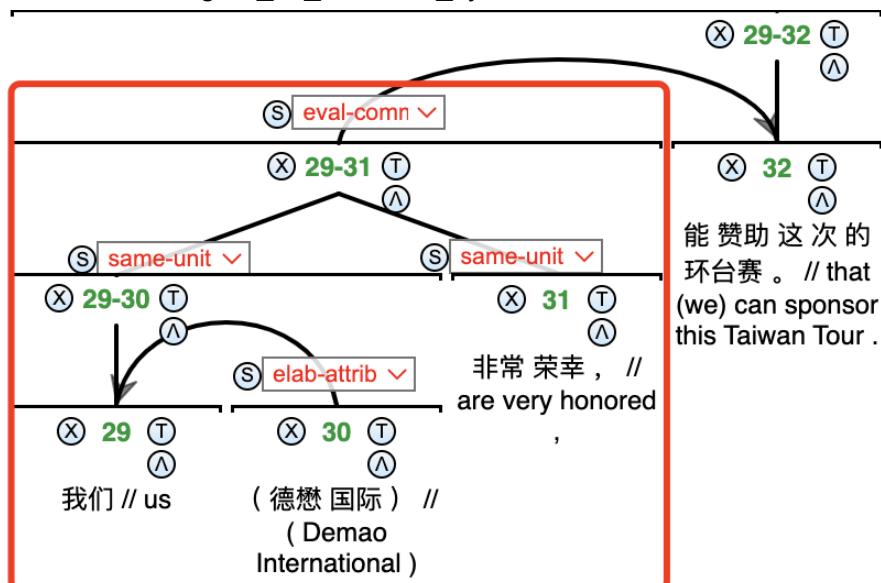
110. EDU\_29 我们 // We

EDU\_30( 德懋 国际 ) // ( Demao International )

EDU\_31 非常 荣幸 , // are very honored ,

EDU\_32 能 赞助 这 次 的 环台赛 。 // that (we) can sponsor this Taiwan Tour .

source: gum\_zh\_interview\_cycle

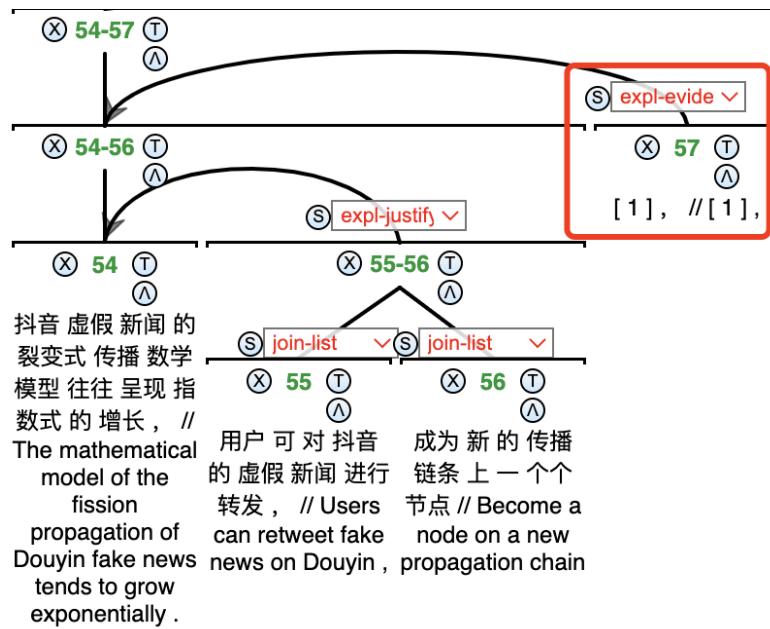


## 2.1.13 explanation-evidence

**explanation-evidence: the Satellite gives evidence that the Nucleus is true.**

One common example of an evidence DU is citation as below, the square-bracketed citations are evidences for the preceding quotes or transliterations.

111. EDU\_54 抖音 虚假 新闻 的 裂变式 传播 数学 模型 往往 呈现 指数式的 增长 , //  
The mathematical model of the fission propagation of Douyin fake news tends to grow exponentially .  
EDU\_55 用户 可 对 抖音 的 虚假 新闻 进行 转发 , // Users can retweet fake news on Douyin ,  
EDU\_56 成为 新的 传播 链条 上 一个 个 节点 // Become a node on a new propagation chain  
**EDU\_57 [1] , // [1] ,**  
source: gum\_zh\_evidence\_supervision

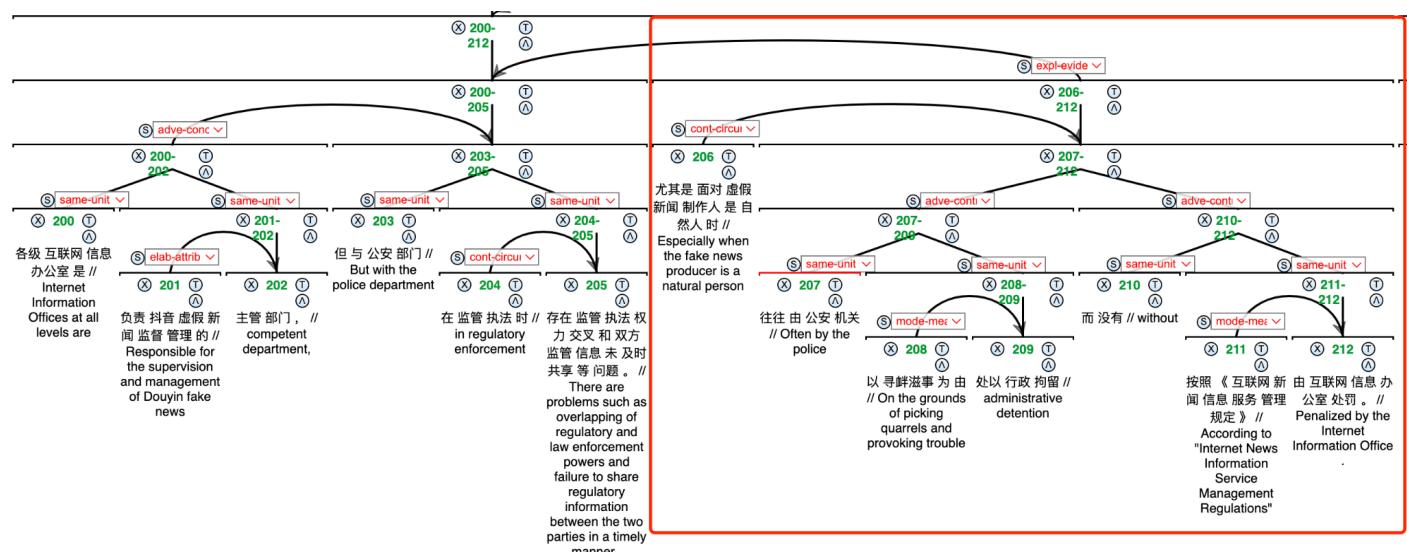


In the following example, the scenario facing fake producers is an example of problematic law enforcement.

112. EDU\_200 各级 互联网 信息 办公室 是 // Internet Information Offices at all levels are  
EDU\_201 负责 抖音 虚假 新闻 监督 管理 的 // Responsible for the supervision and management of Douyin fake news  
EDU\_202 主管 部门 , // competent department,  
EDU\_203 但 与 公安 部门 // But with the police department

- EDU\_204 在 监 管 执 法 时 // in regulatory enforcement
- EDU\_205 存 在 监 管 执 法 权 力 交 叉 和 双 方 监 管 信 息 未 及 时 共 享 等 问 题 。 //  
There are problems such as overlapping of regulatory and law enforcement powers and failure to share regulatory information between the two parties in a timely manner .
- EDU\_206 尤 其 是 面 对 虚 假 新 闻 制 作 人 是 自 然 人 时 // Especially when the fake news producer is a natural person
- EDU\_207 往 往 由 公 安 机 关 // Often by the police
- EDU\_208 以 寻 邪 滋 事 为 由 // On the grounds of picking quarrels and provoking trouble
- EDU\_209 处 以 行 政 拘 留 // administrative detention
- EDU\_210 而 没 有 // without
- EDU\_211 按 照 《 互 联 网 新 闻 信 息 服 务 管 理 规 定 》 // According to "Internet News Information Service Management Regulations"
- EDU\_212 由 互 联 网 信 息 办 公 室 处 罚 。 // Penalized by the Internet Information Office .

source: gum\_zh\_academic\_supervision



### 2.1.14 explanation-justify

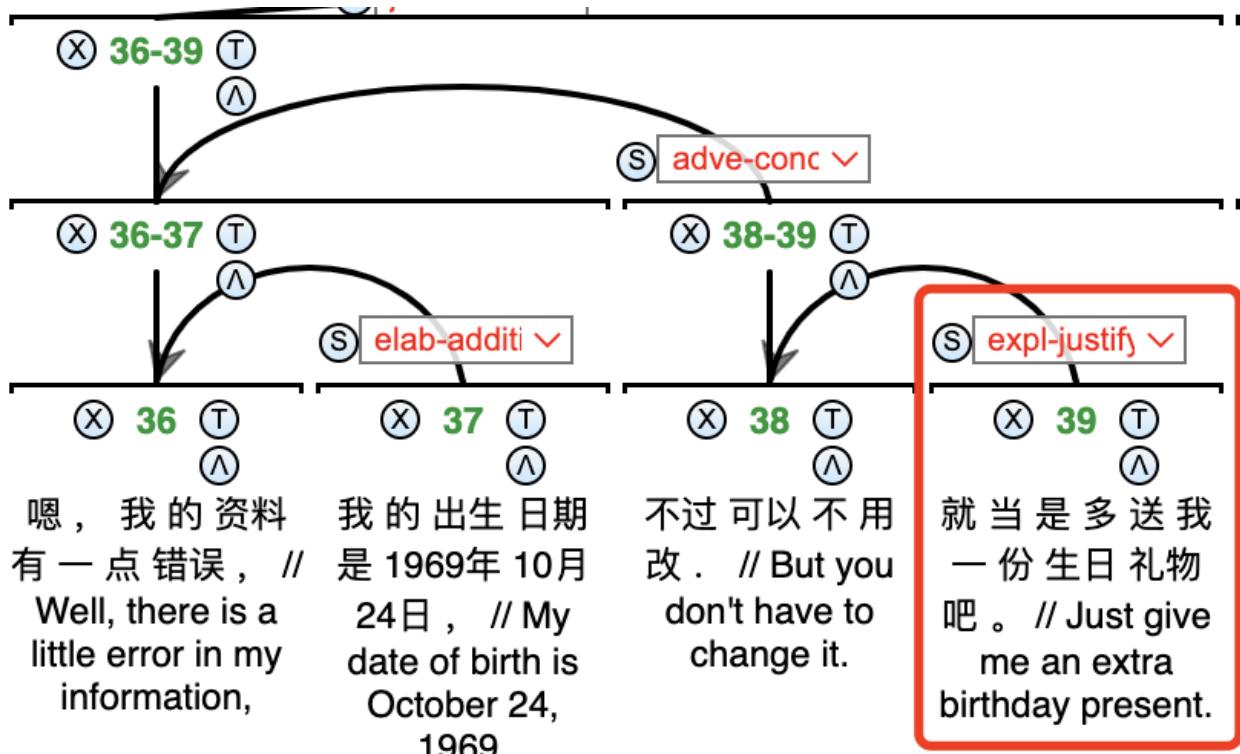
**explanation-justify: the Satellite justifies why the Writer can say the Nucleus.**

The satellite explanation-justify gives further explanation to the reader why the author states the nucleus.

In the following example, "giving me an extra birthday present" justifies why the speaker does not want correct his wrong birthday on Wikipedia.

113. EDU\_36 嗯 , 我的 资料 有 一 点 错 误 , // Well, there is a little error in my information,  
EDU\_37 我 的 出 生 日期 是 1969年 10月 24日 , // My date of birth is October 24, 1969,  
EDU\_38 不 过 可 以 不 用 改 . // But you don't have to change it.  
**EDU\_39 就 当 是 多 送 我 一 份 生 日 礼 物 吧 。 // Just give me an extra birthday present.**

source: gum\_zh\_interview\_keyman

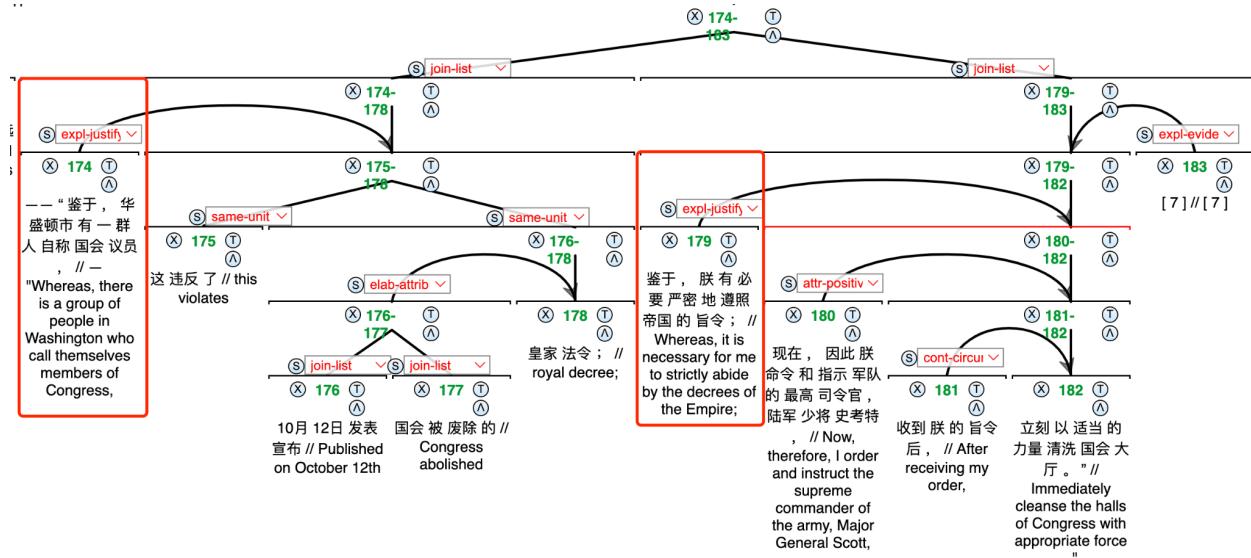


In Chinese, 鉴于 is a word that means "based on..." or "given that..." usually referring to a prescribed rule or principle. Here is an example of two consecutive 鉴于 explanation-justify examples.

114. EDU\_174 —— “ 鉴于 , 华盛顿市 有一 群 人 自 称 国 会 议 员 , // — "Whereas, there is a group of people in Washington who call themselves members of Congress, ||  
EDU\_175 这 违 反 了 // this violates ||  
EDU\_176 10月 12日 发 表 宣 布 // Published on October 12th ||

EDU\_177 国会被废除的 // Congress abolished ||  
 EDU\_178 皇家法令； // royal decree; ||  
**EDU\_179 鉴于，朕有必要严密地遵照帝国的旨令； // Whereas, it is necessary for me to strictly abide by the decrees of the Empire; ||**  
 EDU\_180 现在，因此朕命令和指示军队的最高司令官，陆军少将史考特，// Now, therefore, I order and instruct the supreme commander of the army, Major General Scott,  
 EDU\_181 收到朕的旨令后， // After receiving my order, ||  
 EDU\_182 立刻以适当的力量清洗国会大厅。" // Immediately cleanse the halls of Congress with appropriate force . " ||  
 [ 7 ] // [ 7 ] ||

source: gum\_zh\_bio\_emperor



### 2.1.15 explanation-motivation

#### explanation-motivation: the Satellite motivates the Reader to do the Nucleus.

explanation-motivation rarely occurs in the written or formal text where the writer tends not to interact directly with the reader. In this corpus, organization-motivation occurs more frequently in the wikihow genre.

In this short example, asking whether the readers want to DIY their own glowstick motivates them to continue reading this wikihow article.

**115. EDU\_13 还是很想制作荧光棒吗？ // Still want to make glow sticks？**

EDU\_14 那就继续阅读吧。 // Then read on.

source: gum\_zh\_whow\_glowstick



In the following example, the fact that "rats are usually tempted by food" motivates the readers to hide food from rats.

116. EDU\_18 老鼠跑到屋子里来一般都是受到食物的诱惑。// Rats are usually tempted by food when they come into the house.

EDU\_19 如果家里什么吃的都没有，// If there is nothing to eat at home,

EDU\_20 老鼠自然也就不怎么呆的下去了。// The mice naturally didn't stay much longer.

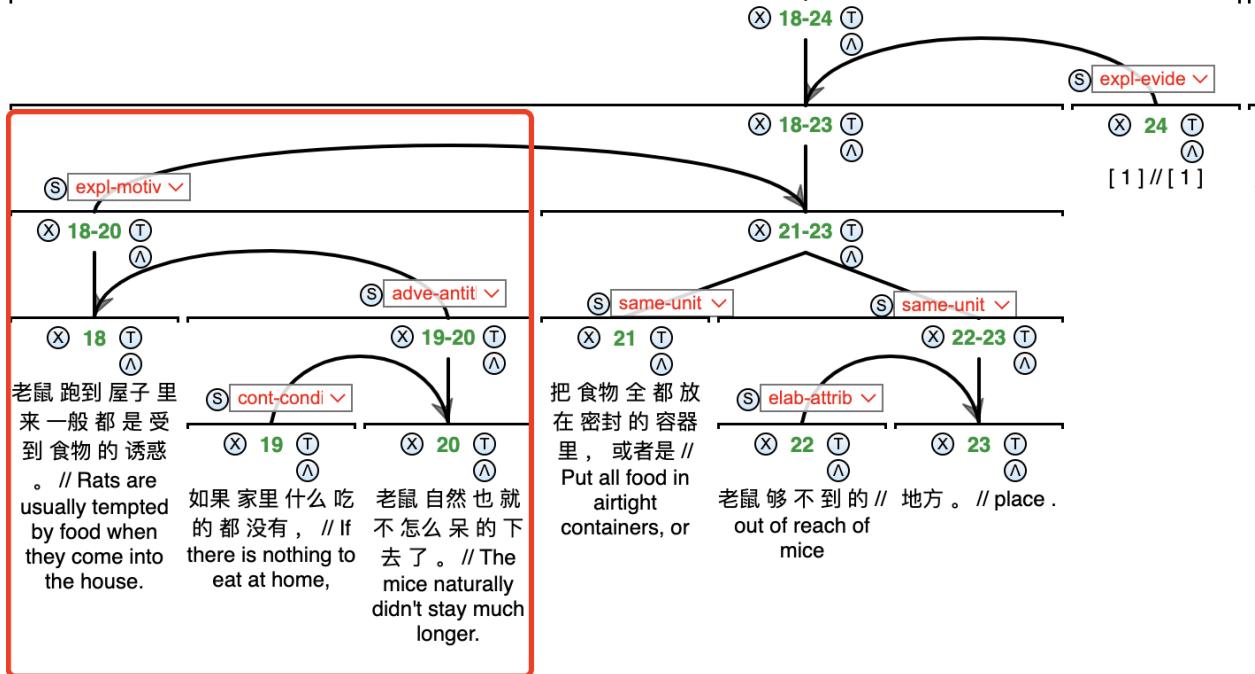
EDU\_21 把食物全都放在密封的容器里，或者是 // Put all food in airtight containers, or

EDU\_22 老鼠够不到的 // out of reach of mice

EDU\_23 地方。// place .

EDU\_24 [1]// [1]

source: gum\_zh\_whow\_mice



### 2.1.16 mode-manner

**mode-manner:** the Satellite gives the manner of how the Nucleus happened.

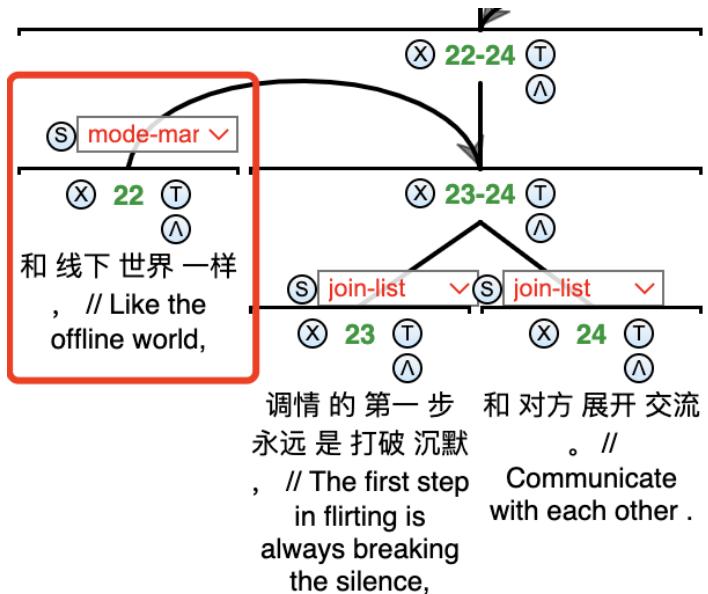
Based on Carlson 2003, "A manner satellite explains the way in which something is done. (It sometimes also expresses some sort of similarity/comparison.) The satellite answers the question "in what manner?" or "in what way?". **A MANNER relation is less "goal-oriented" than a MEANS relation, and often is more of a description of the style of an action."**

117. EDU\_22 和线下世界一样，// Like the offline world,

EDU\_23 调情的第一步永远是打破沉默，// The first step in flirting is always breaking the silence,

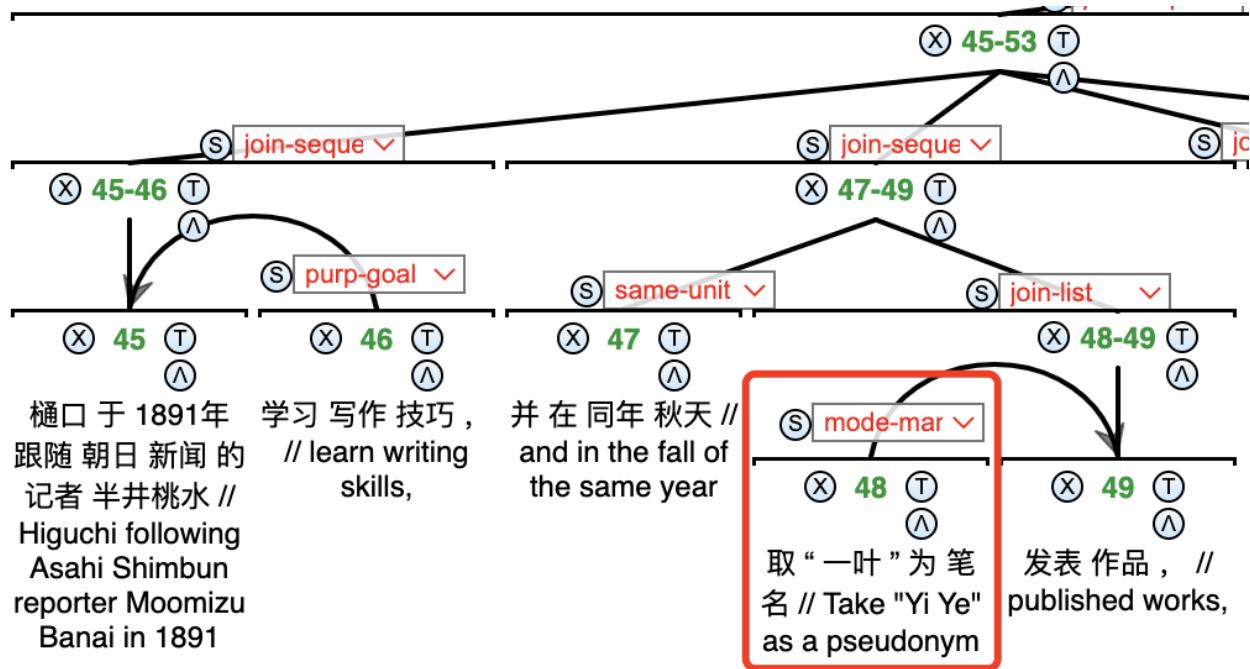
EDU\_24 和对方展开交流。// Communicate with each other .

source: gum\_zh\_whow\_flirt



In the following example, "taking a pseudonym" specifies how Higuchi publishes works.

118. EDU\_45 樋口 于 1891年 跟随 朝日 新闻 的 记者 半井桃水 // Higuchi following Asahi Shimbun reporter Moomizu Banai in 1891  
 EDU\_46 学习 写作 技巧 , // learn writing skills,  
 EDU\_47 并 在 同年 秋天 // and in the fall of the same year  
**EDU\_48 取 “一叶” 为 笔名 // Take "Yi Ye" as a pseudonym**  
 EDU\_49 发表 作品 , // published works,  
 source: gum\_zh\_bio\_higuchi



## 2.1.17 mode-means

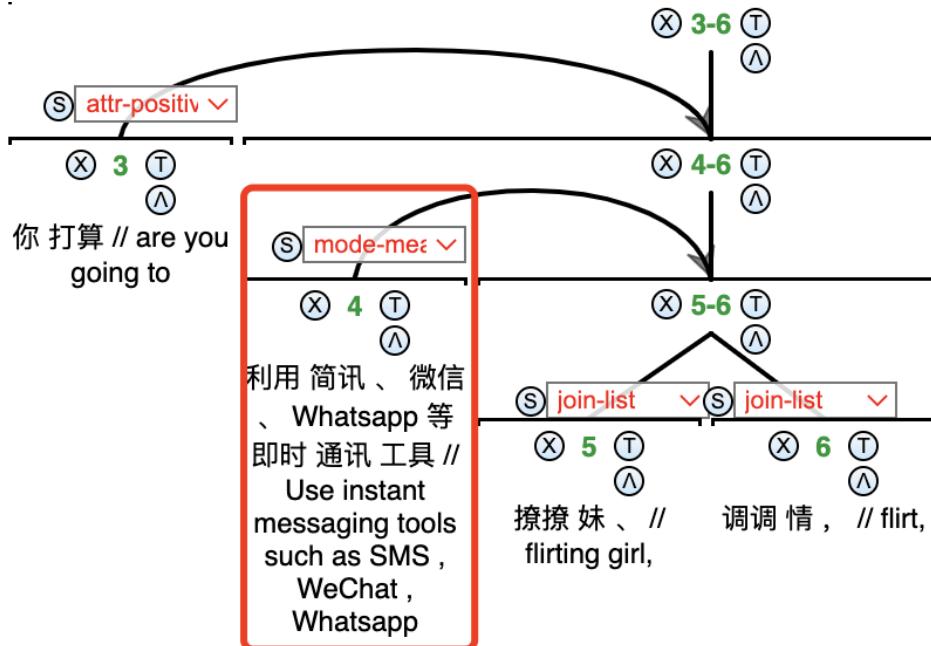
**mode-means: the Satellite indicates means by which the Nucleus happened.**

Based on Carlson 2003, "a means satellite specifies a **method, mechanism, instrument, channel or conduit for accomplishing some goal**. It should tell you how something was or is to be accomplished. In other words, the satellite answers a "by which means" or "how" question that can be assigned to the nucleus. **It is often indicated by the preposition by.**"

In Chinese, means can be paraphrased as 用...的方式 (by means of ...)

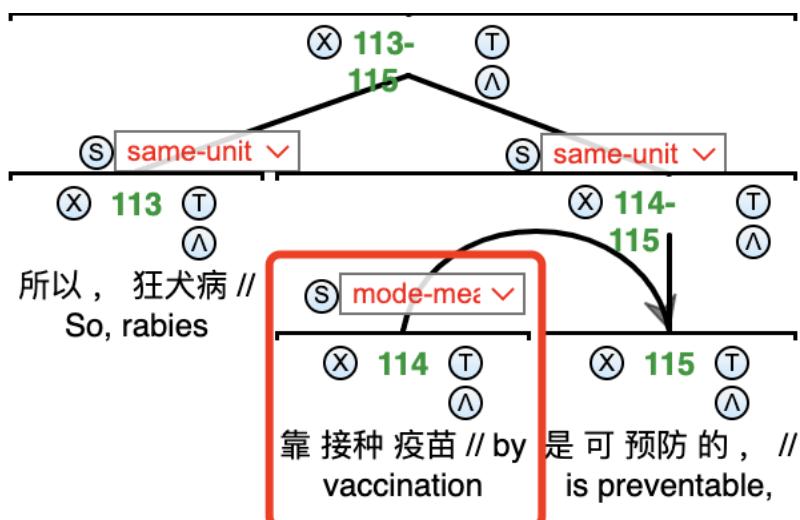
in other words, different from manner, means is when you cannot accomplish the nucleus without the method mentioned in the satellite.

119. EDU\_3 你打算 // are you going to  
 EDU\_4 利用 简讯、微信、Whatsapp 等即时通讯工具 // Use instant messaging tools such as SMS, WeChat, Whatsapp  
 EDU\_5 撩撩妹 // flirting girl,  
 EDU\_6 调调情 // flirt,
- source: gum\_zh\_whow\_flirt



Similarly, "injecting vaccines" is the way to cure rabies.

120. EDU\_113 所以，狂犬病 // So, rabies  
**EDU\_114 靠 接种 疫苗 // by injecting vaccines**  
 EDU\_115 是 可 预防 的， // is preventable,  
 source: gum\_zh\_academic\_rabies



## 2.1.18 organization-heading

**organization-heading: the Satellite is graphically arranged to prepare for the Nucleus.**

Organization-heading satellites can be easily distinguished by graphical traits. They are usually headings and there is a line break between heading and the main content. These include document titles, section and subsection headings.

In the following example, "family background" and "life" are two section titles.

## 121. EDU\_21 家世 // family background

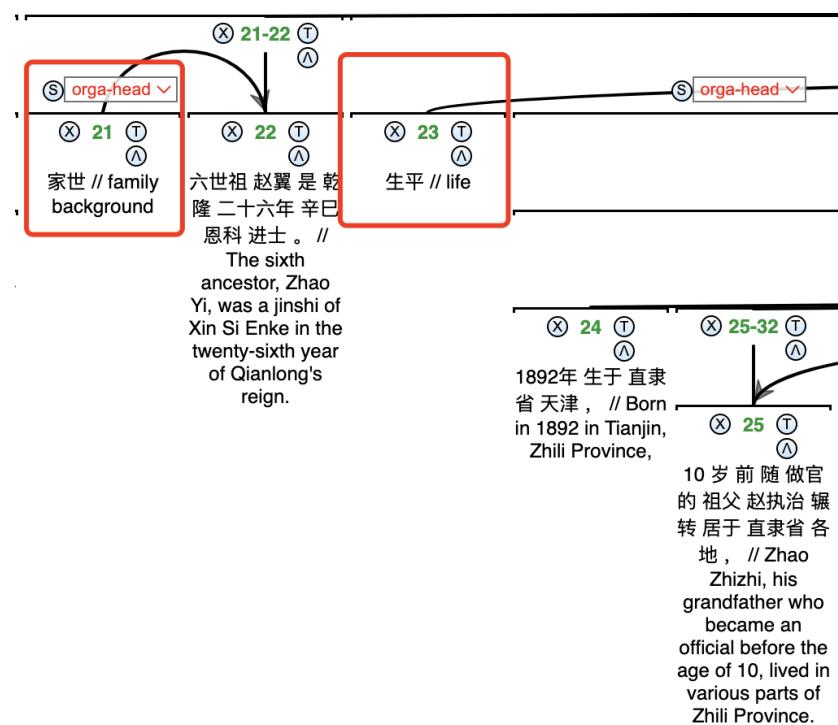
EDU\_22 六世祖 赵翼 是 乾隆 二十六年 辛巳 恩科 进士 。 // The sixth ancestor, Zhao Yi, was a jinshi of Xin Si Enke in the twenty-sixth year of Qianlong's reign.

## EDU\_23 生平 // life

EDU\_24 1892年 生于 直隶省 天津 , // Born in 1892 in Tianjin, Zhili Province,

EDU\_25 10岁 前 随 做官 的 祖父 赵执治 辗转 居于 直隶省 各 地 , // Zhao Zhizhi, his grandfather who became an official before the age of 10, lived in various parts of Zhili Province.

source: gum\_zh\_bio\_chao



Similar parallelism of subsections are shown with different apple products in the following example.

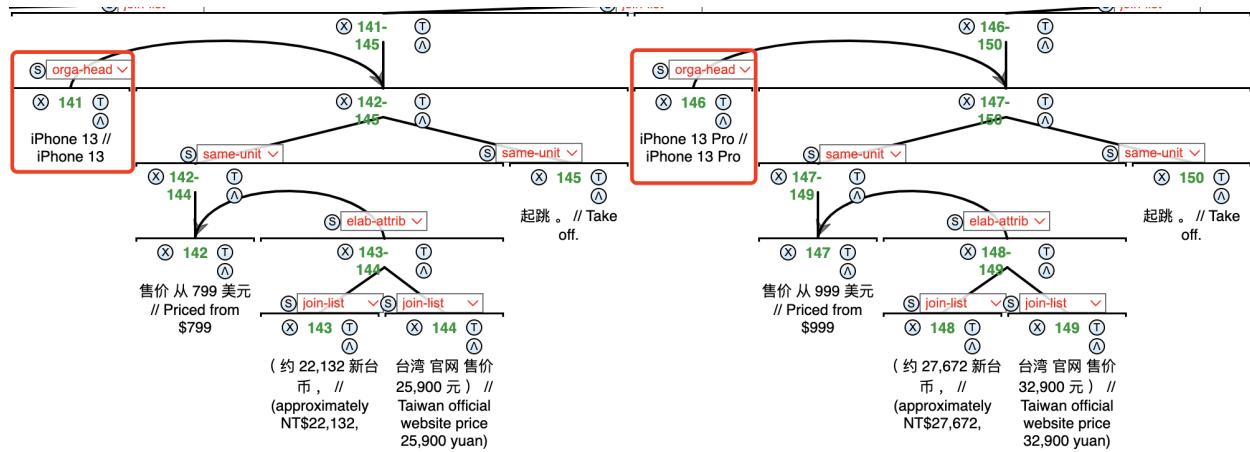
## 122. EDU\_141 iPhone 13 // iPhone 13

EDU\_142 售价 从 799 美元 // Priced from \$799

EDU\_143 (约 22,132 新台币 , // (approximately NT\$22,132,

- EDU\_144 台湾官网售价 25,900 元) // Taiwan official website price 25,900 yuan)  
 EDU\_145 起跳。// Take off.  
**EDU\_146 iPhone 13 Pro // iPhone 13 Pro**  
 EDU\_147 售价从 999 美元 // Priced from \$999  
 EDU\_148 (约 27,672 新台币, // (approximately NT\$27,672,  
 EDU\_149 台湾官网售价 32,900 元) // Taiwan official website price 32,900 yuan)  
 EDU\_150 起跳。// Take off.

source: gum\_zh\_news\_apple



## 2.1.19 organization-phatic

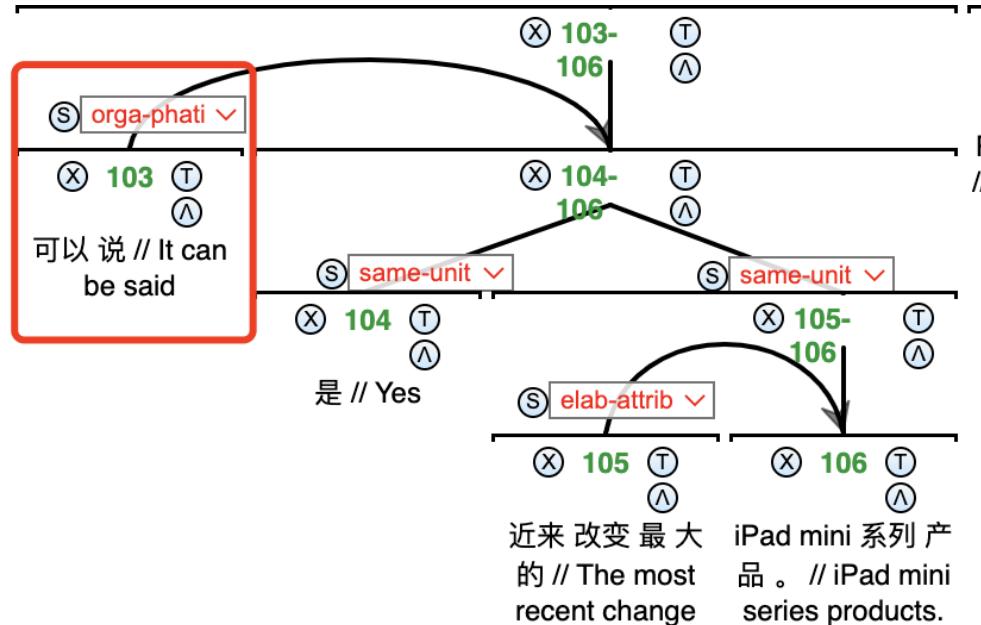
**organization-phatic: the Satellite holds the floor for the Nucleus, with no semantic value.**

This label is usually applied to language disfluencies within a text.

In English, "see," "you know" are examples of such phatic expressions. In Chinese, such examples include "这么说" (saying this way), "啥" (what), etc.

### 123. EDU\_103 可以说 // It can be said

- EDU\_104 是 // it is  
 EDU\_105 近来 改变 最 大 的 // that changed the most recently  
 EDU\_106 iPad mini 系列 产品。 // iPad mini series products.  
 source: gum\_zh\_news\_apple



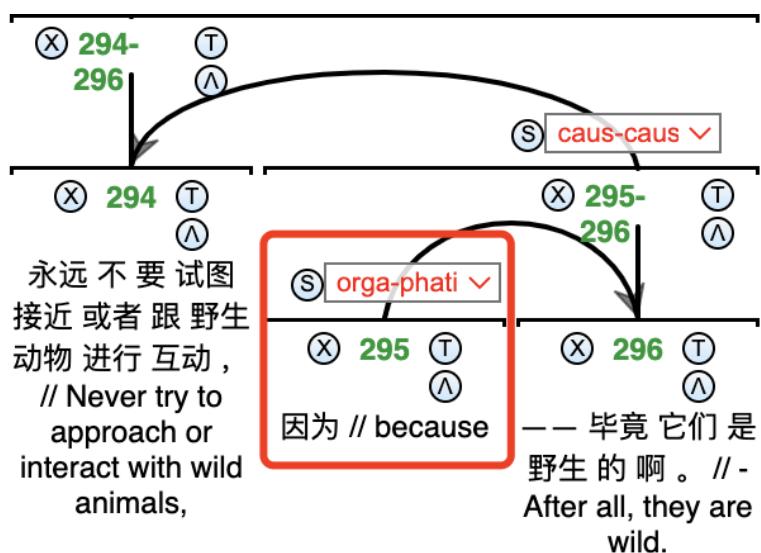
Organization-phatic is also used for self-corrections in speech. In the following example, 毕竟 (nevertheless) replaces 因为 (because), a more smooth connective between "they are wild animals" and "do not interact with them."

124. EDU\_294 永远不要试图接近或者跟野生动物进行互动， // Never try to approach or interact with wild animals,

**EDU\_295 因为 // because**

EDU\_296 —— 毕竟它们是野生的啊。 // - After all, they are wild.

source: gum\_zh\_whow\_hiking



## 2.1.20 organization-preparation

**organization-preparation: the Satellite prepares the Reader for the Nucleus.**

Compared to context-background, organization-preparation contributes minimal information and simply serves the purpose of bridging discourse sections with a document.

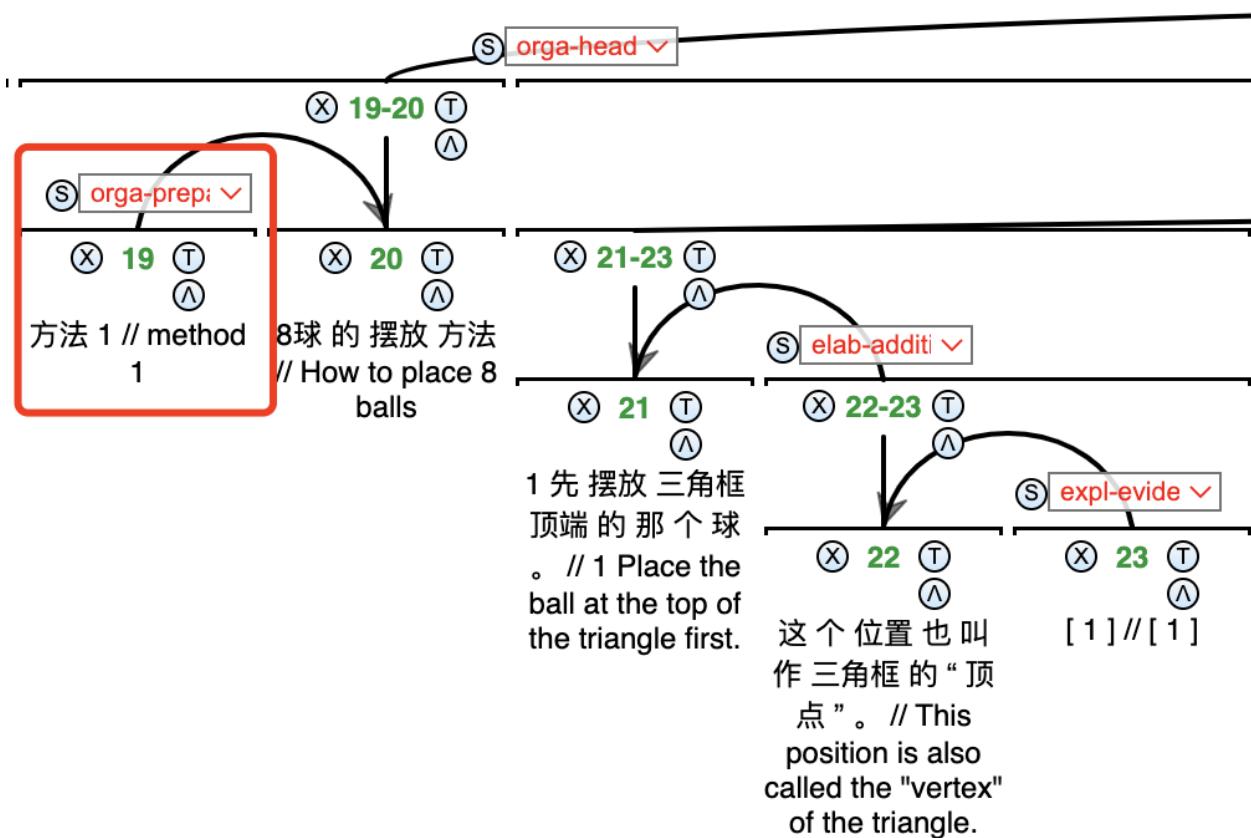
We label those graphically not distinguished headers organization-preparation.

In the following example, "method 1" is an organization-preparation for "how to place 8 balls" since there is no graphical disfluency between the two. Whereas they as a whole is the header of the following section so DU\_19-20 functions as organization-heading.

### 125. EDU\_19 方法 1 // method 1

EDU\_20 8球 的 摆放 方法 // How to place 8 balls

source: gum\_zh\_whow\_pool



organization-preparation can also be found within the main texts. For example, this relation is used to label the beginning of a document, section or paragraph that continues from the preceding one.

In the following example, the targets of the genocide prepares for Ikhlov's claim that the intention is to eliminate rich peasants.

126. EDU\_81 伊赫洛夫 说 : // Ikhlov says:

EDU\_82 “尽管大饥荒中饿死的以乌克兰人和哈萨克人居多，// "Although most of the people who starved to death in the Great Famine were Ukrainians and Kazakhs,

EDU\_83 但那场民族灭绝行动不仅仅针对单一民族，// But that genocide wasn't just for a single people.

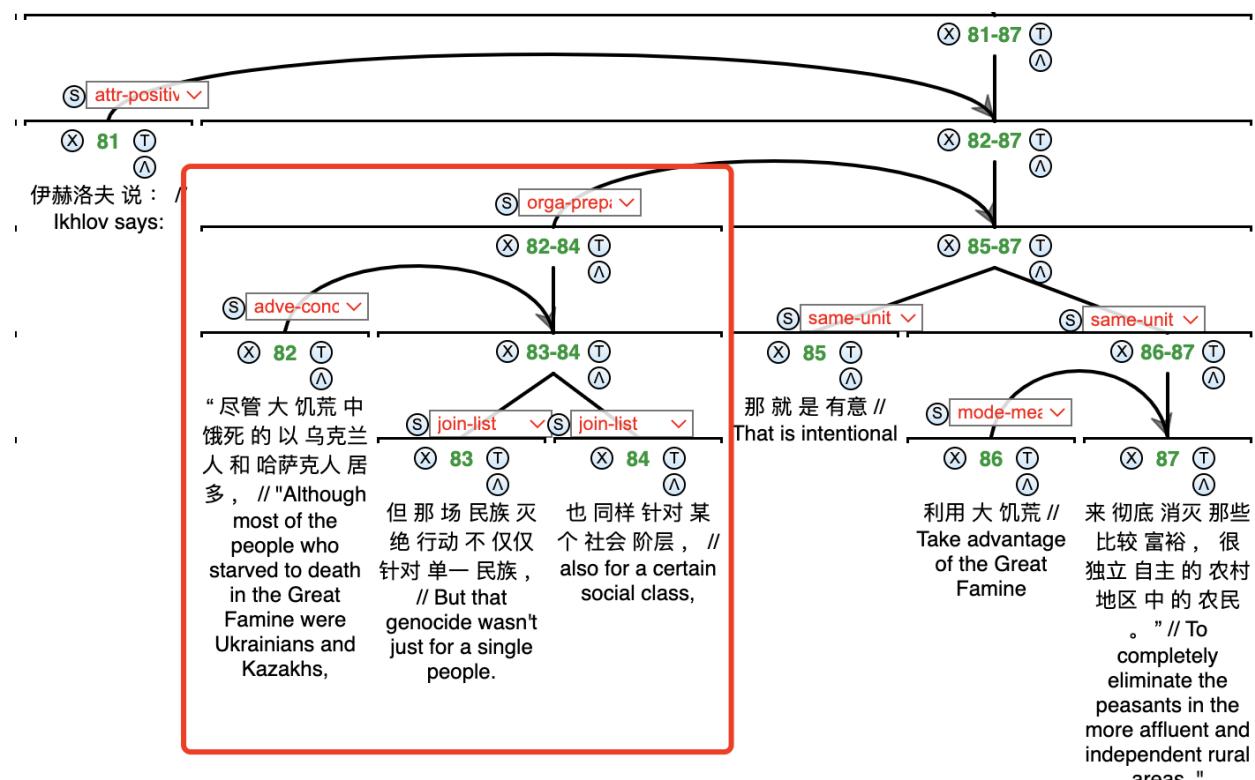
EDU\_84 也同样针对某个社会阶层，// also for a certain social class,

EDU\_85 那就是有意 // That is intentional

EDU\_86 利用大饥荒 // Take advantage of the Great Famine

EDU\_87 来彻底消灭那些比较富裕，很独立自主的农村地区中的农民。" // To completely eliminate the peasants in the more affluent and independent rural areas. "

source: gum\_zh\_news\_famine



## 2.1.21 purpose-attribute

**purpose-attribute: only a part of the Nucleus (a phrase rather than the entire clause) occurs in order for the Satellite to happen.**

*Purpose-attribute* is the nominal-modifier counterpart of *purpose-goal* just like *elaboration-attribute* for *elaboration-additional*.

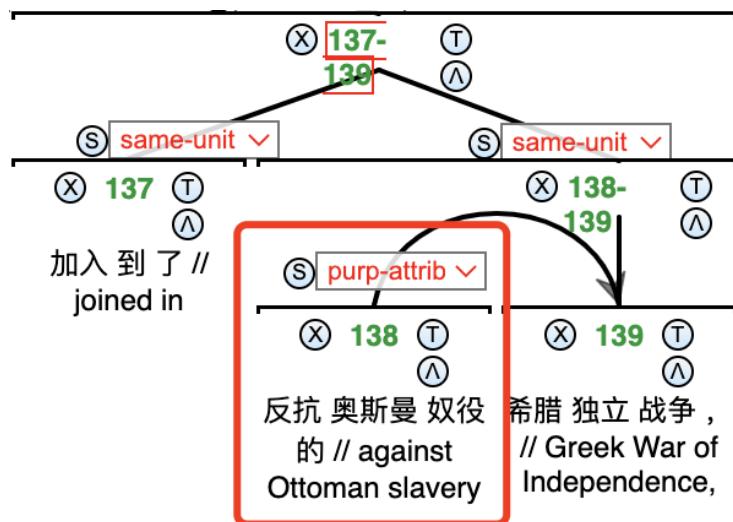
In the following example, "the Greek War of Independence" is for the purpose of resisting Ottoman slavery.

127. EDU\_137 加入 到 了 // joined in

EDU\_138 反抗 奥斯曼 奴役 的 // against Ottoman slavery

EDU\_139 希腊 独立 战争 , // Greek War of Independence,

source: gum\_zh\_bio\_byron



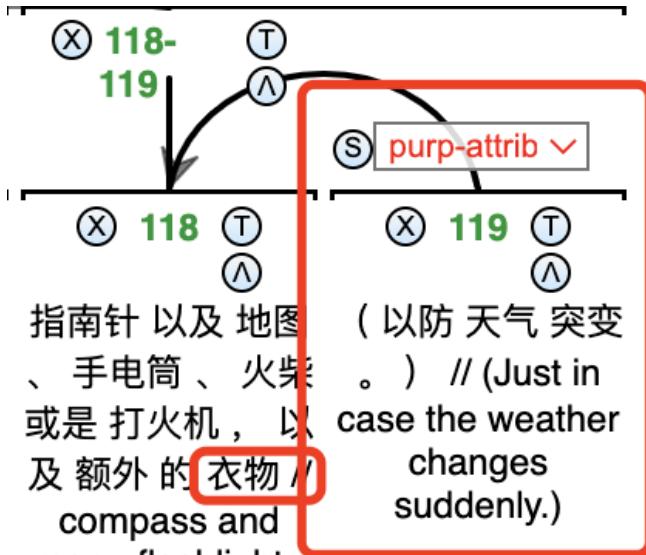
Similar to *elaboration-attribute*, there are also a few occurrences of *purpose-attribute* that are not relative clauses.

In the following example, only the "extra clothing" is for preventing weather change.

128. EDU\_118 指南针 以及 地图 、 手电筒 、 火柴 或是 打火机 , 以及 额外 的 衣物 // compass and map , flashlight , matches or lighters , and extra clothing

EDU\_119 (以防 天气 突变。 ) // (Just in case the weather changes suddenly.)

source: gum\_zh\_whow\_hiking



### 2.1.22 purpose-goal

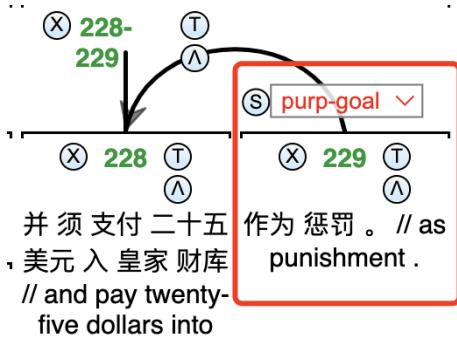
**purpose-goal:** the Nucleus occurs in order for the Satellite to happen.

In the following example, the "\$25" is for the purpose of punishment.

129. EDU\_228 并须支付二十五美元入皇家财库 // and pay twenty-five dollars into the royal treasury ||

EDU\_229 作为惩罚。// as punishment . ||

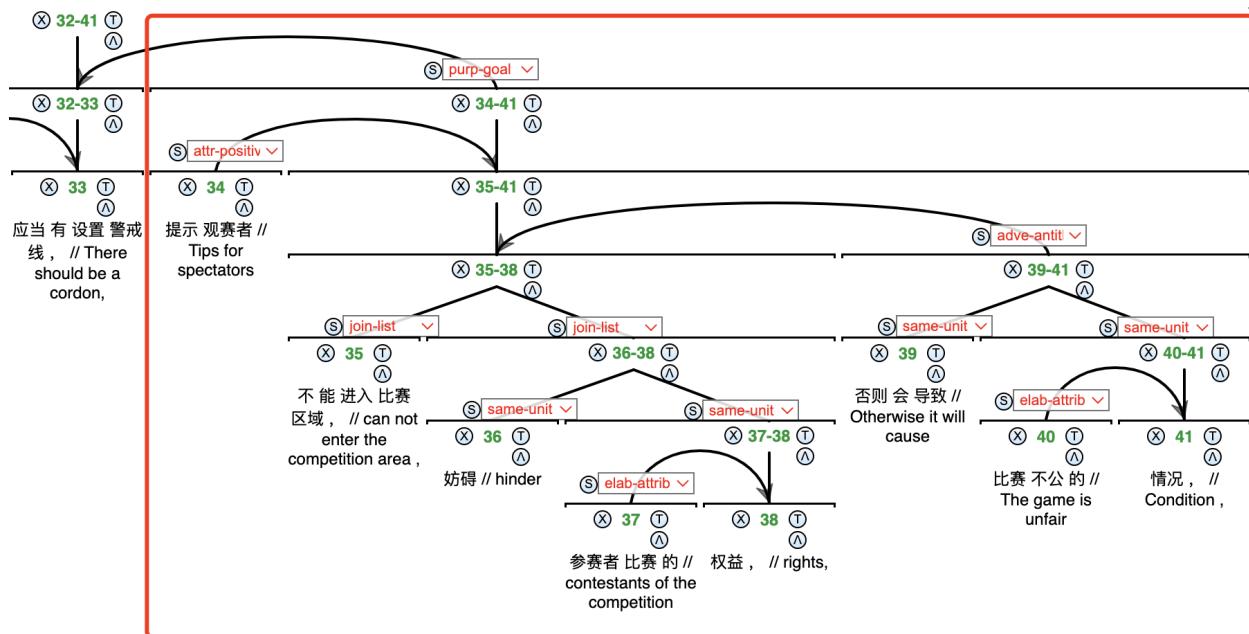
source: gum\_zh\_bio\_emperor



In the following example, the purposes of "setting up a cordon" are reminding the spectators to not enter the area, not hinder the rights, etc.

130. EDU\_33 应当有设置警戒线，// There should be a cordon,

- EDU\_34** 提示 观赛者 // Tips for spectators  
**EDU\_35** 不能进入 比赛 区域 , // can not enter the competition area ,  
**EDU\_36** 妨碍 // hinder  
**EDU\_37** 参赛者 比赛 的 // contestants of the competition  
**EDU\_38** 权益 , // rights,  
**EDU\_39** 否则 会 导致 // Otherwise it will cause  
**EDU\_40** 比赛 不公 的 // The game is unfair  
**EDU\_41** 情况 , // Condition ,  
source: gum\_zh\_interview\_game



### 2.1.23 restatement-partial

**restatement-partial: the Satellite reiterates part of the Nucleus.**

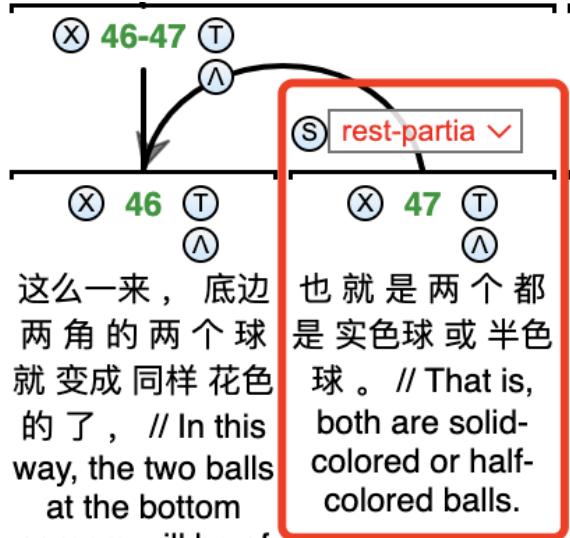
If the relation is a full repetition, please use the multinuclear relation: restatement-repetition.

In the following example, "two ball are both solid or strip/half" is a repetition of "two balls of the same kind" in EDU\_46.

131. EDU\_46 这么一来 , 底边 两 角 的 两 个 球 就 变 成 同 样 花 色 的 了 , // In this way, the two balls at the bottom corners will be of the same suit.

**EDU\_47** 也 就 是 两 个 都 是 实 色 球 或 半 色 球 。 // That is, both are solid-colored or half-colored balls.

source: gum\_zh\_whow\_pool

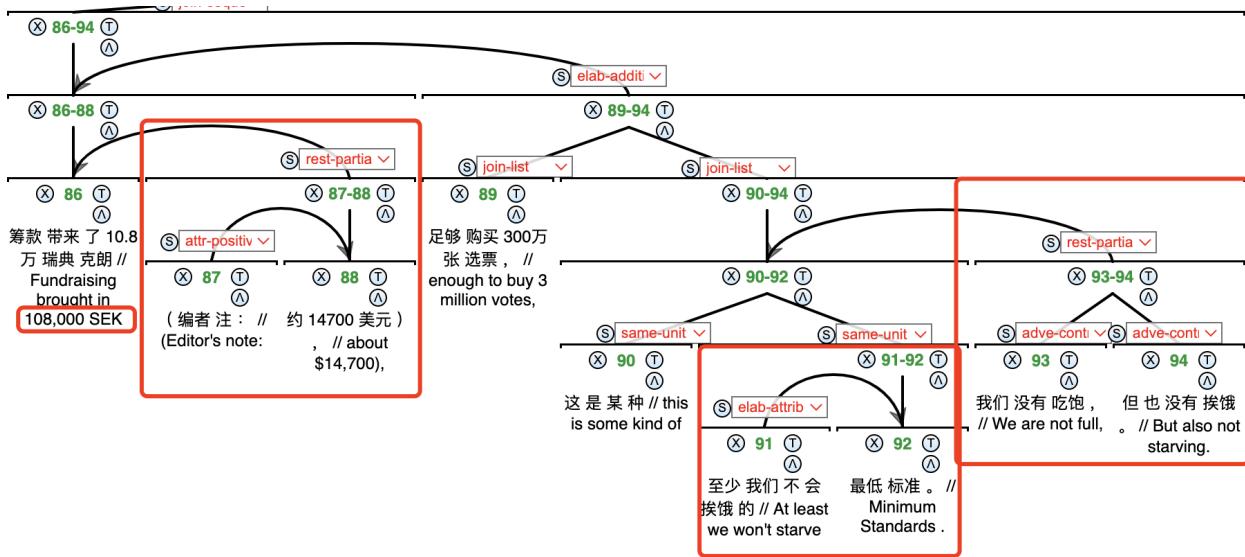


这么一来，底边  
两角的两个球  
就变成同样花色  
的了， // In this  
way, the two balls  
at the bottom  
corners will be of  
the same suit.

也就是两个都  
是实色球或半色  
球。 // That is,  
both are solid-  
colored or half-  
colored balls.

The following example includes two occurrences of *restatement-partial*: "14,700 USD" is a repetition of "108,000 SEK" and "not full but not starving" is a repetition of "minimum standard for not starving."

132. EDU\_86 筹款带来了 10.8万 瑞典 克朗 // Fundraising brought in 108,000 SEK  
 EDU\_87 (编者注： // (Editor's note:  
 EDU\_88 约 14700 美元 ) , // about \$14,700),  
 EDU\_89 足够 购买 300万 张 选票 , // enough to buy 3 million votes,  
 EDU\_90 这 是 某 种 // this is some kind of  
 EDU\_91 至少 我 们 不 会 挨 饿 的 // At least we won't starve  
 EDU\_92 最低 标 准 。 // Minimum Standards .  
 EDU\_93 我 们 没 有 吃 饱 , // We are not full,  
 EDU\_94 但 也 没 有 挨 饿 。 // But also not starving.  
 source: gum\_zh\_interview\_falkvinge



Note: we draw the following distinctions between **restatement** versus **elaboration**.

In general, restatement does not provide additional knowledge and are interpreted as equivalent given the context (as well as world knowledge).

The followings are **restatement**:

- "today" ← "(May 11)"
- Synonym of the same entity in the language
- when the latter part can be conducted from the former, "the number went from 50 to 40"  
← "it decreased by 10"

In contrary, the followings are **elaboration**:

- Same phrase in different languages (i.e. translations) are not restatements; e.g. "Apple"  
← "( German : Apfel )"

## 2.1.24 topic-question

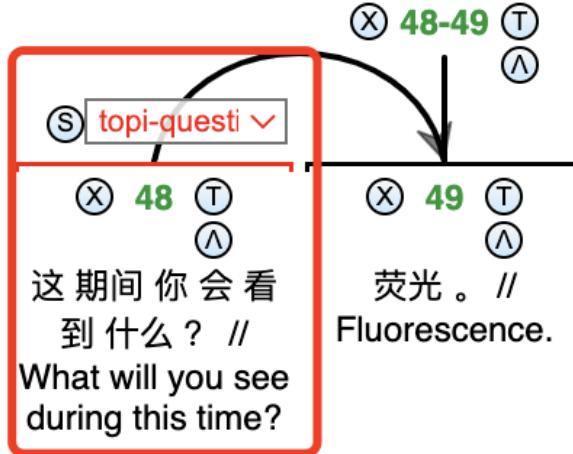
**topic-question: the Satellite requests the information in the Nucleus.**

In the following example, the question satellite asks for what "you will see during this time."

**133. EDU\_48** 这期间你会看到什么? // What will you see during this time?

EDU\_49 荧光。 // Fluorescence.

source: gum\_zh\_whow\_glowstick



In the following example, we see two question-answer pairs. DU\_146-147 asks for the interviewee's experience, and EDU\_148 is a clarification question regarding "Tour Taiwan Tournament."

**134. EDU\_146** 对于今年让环台赛、自行车展、体育用品展三合一的 // For this year's trip to make the Tour of Taiwan , the bicycle show and the sporting goods show

**EDU\_147** 看法 ? // view?

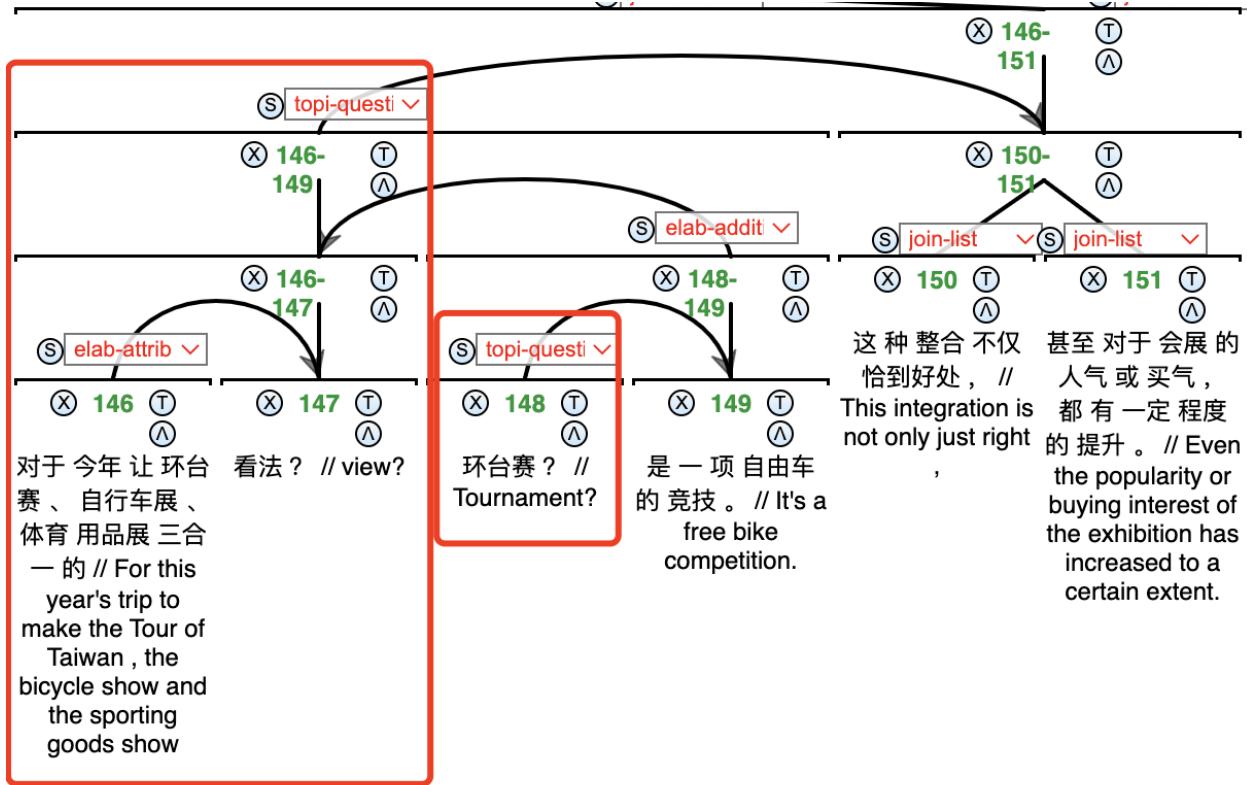
**EDU\_148** 环台赛 ? // Tour Taiwan Tournament?

EDU\_149 是一项自由车的竞技。// It's a free bike competition.

EDU\_150 这种整合不仅恰到好处，// This integration is not only just right ,

EDU\_151 甚至对于会展的人气或买气，都有一定程度的提升。// Even the popularity or buying interest of the exhibition has increased to a certain extent.

source: gum\_zh\_interview\_ideal



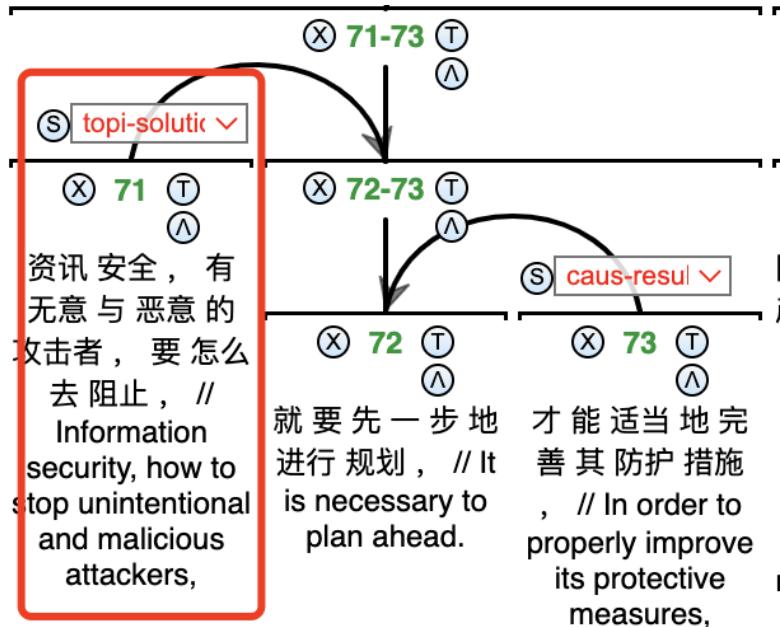
## 2.1.25 topic-solutionhood

**topic-solutionhood: the Nucleus is the answer to a problem in the Satellite.**

In other words, the Satellite poses a problem and the nucleus presents a solution.

In the following example, "planning ahead" is the solution to "preventing attacks."

135. EDU\_71 资讯安全，有无意与恶意的攻击者，要怎么去阻止，//  
**Information security, how to stop unintentional and malicious attackers,**  
 EDU\_72 就要先一步地进行规划，// It is necessary to plan ahead.  
 EDU\_73 才能适当地完善其防护措施，// In order to properly improve its  
 protective measures,  
 source: gum\_zh\_interview\_wimax



topic-solutionhood is a relative infrequent one in Chinese, similarly, *elaboration-additional* occurs ~45 times more frequent than *topic-solutionhood* in GUM.

## 2.2 Multinuclear relations

### 2.2.1 adversative-contrast

**adversative-contrast:** the Writer presents similar units with contrast.

Compared to *adversative-concession* or *adversative-antithesis*, *adversative-contrast* is multclear meaning that the contrastive units are of the same importance.

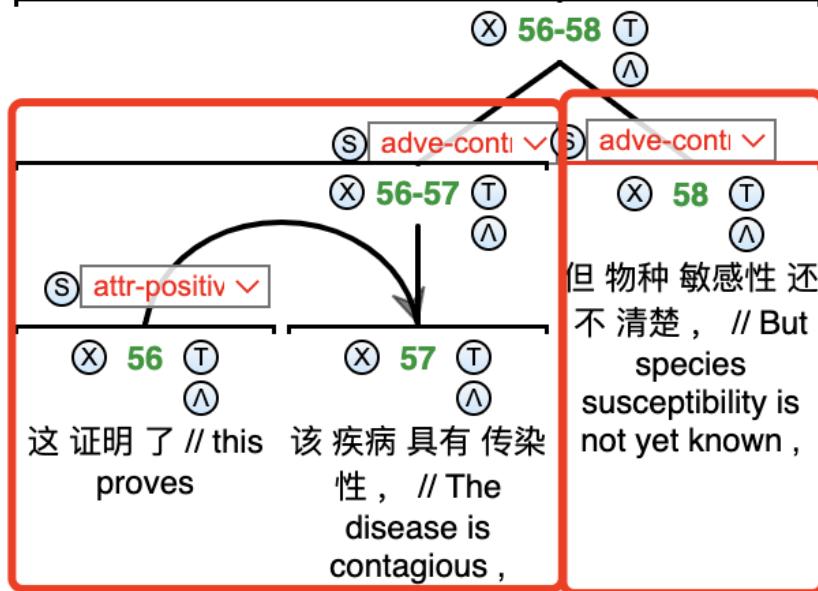
In the following example, "being contagious" and "unclear susceptibility" are adversarial to each other but of equal importance.

#### 136. EDU\_56 这证明了 // this proves

EDU\_57 该疾病具有传染性 , // The disease is contagious ,

EDU\_58 但物种敏感性还不清楚 , // But species susceptibility is not yet known ,

source: gum\_zh\_academic\_rabies



In the following example, we see a contrast between two larger DUs, DU\_118-127 "DEV that has less severe" versus DU\_128-129 "DEV that is not as immunogenic to prevent rabies."

### 137. EDU\_118 严重 反应 较 少 的 // less severe reaction

EDU\_119 另 一 个 疫 苗 是 鸭 胚 疫 苗 // Another vaccine is the duck embryo vaccine

EDU\_120 ( duck embryo vaccine // (duck embryo vaccine

EDU\_121 简 称 DEV ) , // DEV for short),

EDU\_122 该 疫 苗 // the vaccine

EDU\_123 通 过 // pass

EDU\_124 在 受 孕 鸭 蛋 里 传 播 的 // Spread in the egg of a pregnant duck

EDU\_125 病 毒 // Virus

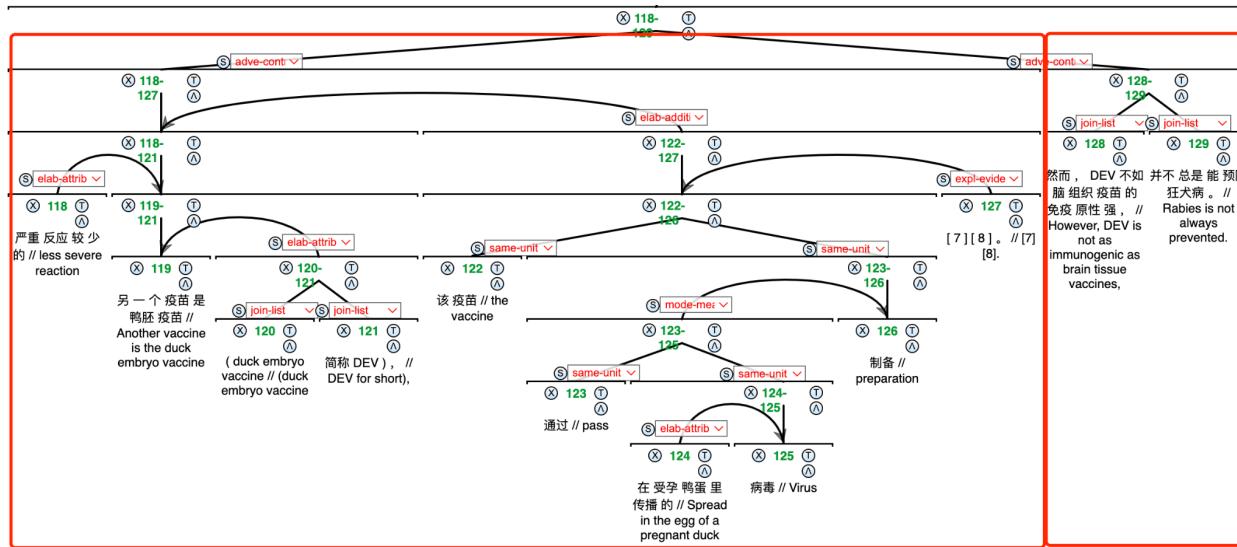
EDU\_126 制 备 // preparation

EDU\_127 [ 7 ][ 8 ] . // [7][8].

EDU\_128 然 而 , DEV 不 如 脑 组 织 疫 苗 的 免 疫 原 性 强 , // However, DEV is not as immunogenic as brain tissue vaccines,

EDU\_129 并 不 总 是 能 预 防 狂 犬 病 。 // Rabies is not always prevented.

source: gum\_zh\_academic\_rabies



## 2.2.2 joint-disjunction

**joint-disjunction: the Writer presents a set of alternatives.**

Different from *joint-list*, the set of alternative are in complementary distribution where I should choose one among them.

In the following example, the hiking trail is either a back-and-forth trail or a loop. Thus, the hiking distance is either twice the distance from the start to the end or the distance of the loop.

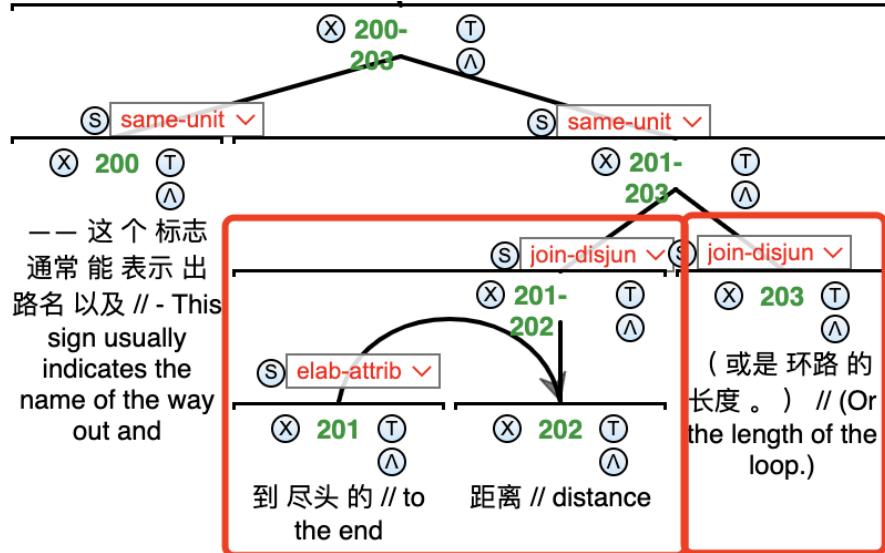
138. EDU\_200 —— 这个标志通常能表示出路名以及 // - This sign usually indicates the name of the way out and

**EDU\_201 到尽头的 // to the end**

**EDU\_202 距离 // distance**

**EDU\_203 (或是环路的长度。) // (Or the length of the loop.)**

source: gum\_zh\_whow\_hiking



The following longer example presents two alternatives: "go to your desired place according to road sign" or "according to map and rockfill landmarks if there is no sign."

### 139. EDU\_223 请 // Please

EDU\_224 依照 // according to

EDU\_225 路标上标的 // marked on a road sign

EDU\_226 路名 // road name

EDU\_227 前往 // go to

EDU\_228 你要去的 // you are going

EDU\_229 地方。// place .

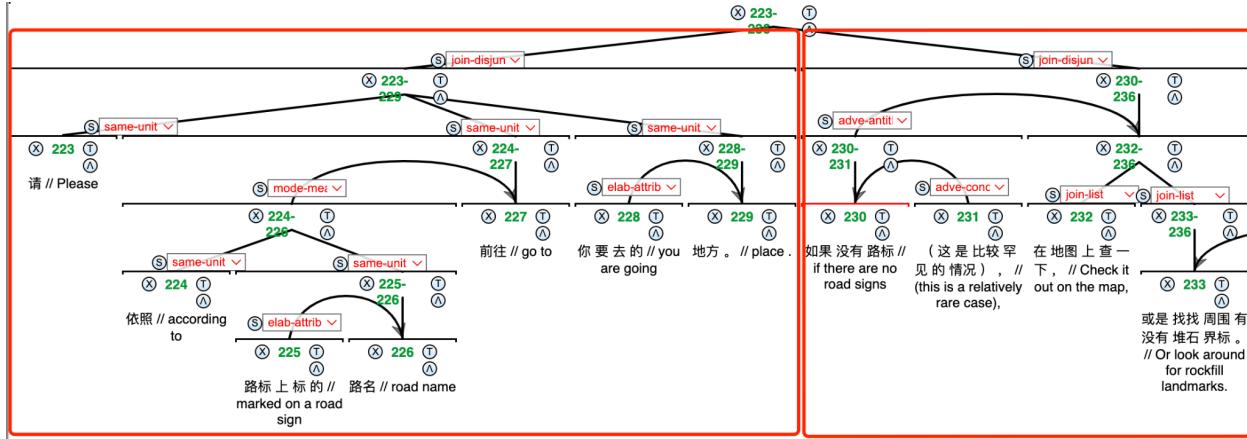
EDU\_230 如果没有路标 // if there are no road signs

EDU\_231 (这是比较罕见的情况) , // (this is a relatively rare case),

EDU\_232 在地图上查一下 , // Check it out on the map,

EDU\_233 或是找找周围有没有堆石界标。 // Or look around for rockfill landmarks.

source: gum\_zh\_whow\_hiking



### 2.2.3 joint-list

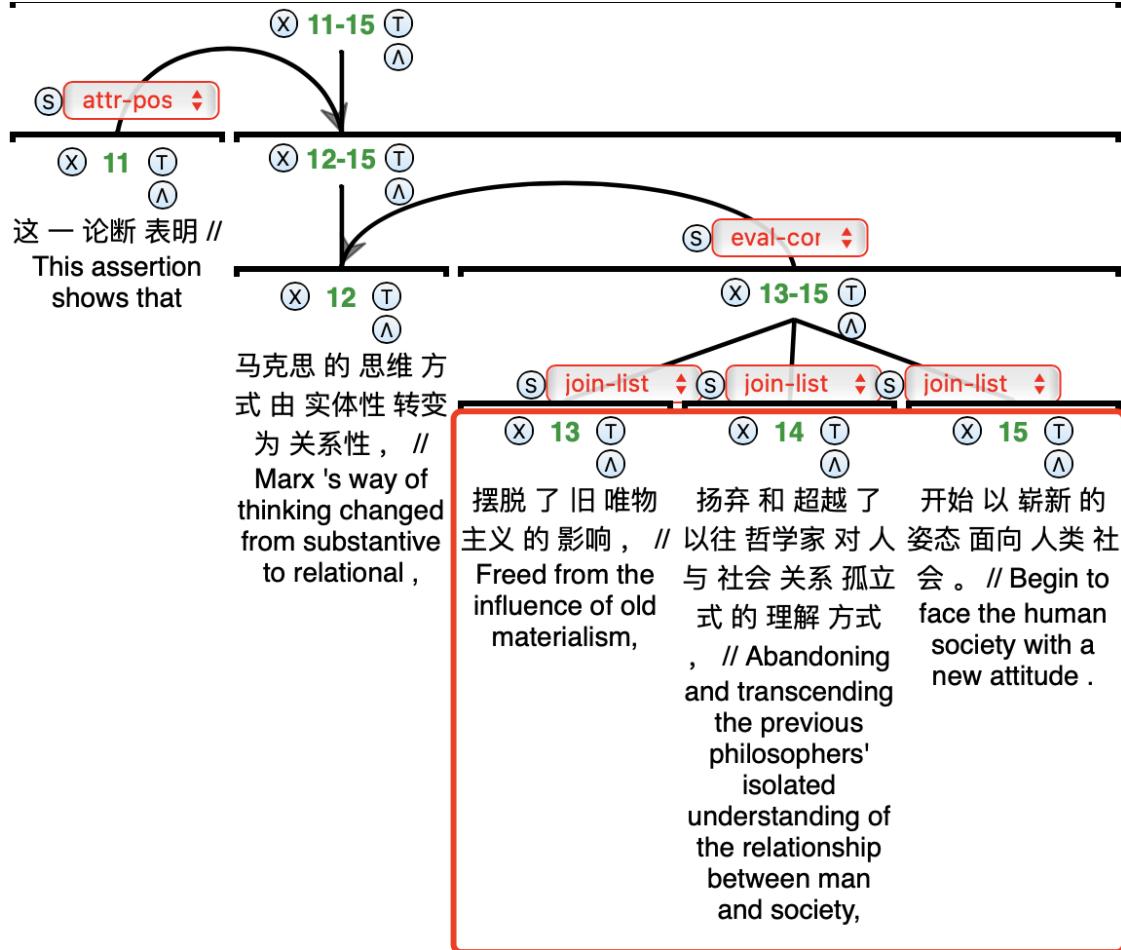
**joint-list:** the Writer presents coordinated and alike units.

The following scenarios are common instances of *joint-list*:

- listed coordinating conjunctions within a sentence
- enumerations such as:
  - Method 1, Method 2
  - Section numbers: 2.1, 2.2

In the following example, "freeing," "abandoning and transcending" and "starting" are the significance of the change of Marx's way of thinking.

140. EDU\_11 这一论断表明 // This assertion shows that  
 EDU\_12 马克思的思维方式由实体性转变为关系性，// Marx's way of thinking changed from substantive to relational,  
**EDU\_13 摆脱了旧唯物主义的影响，// Freed from the influence of old materialism,**  
**EDU\_14 扬弃和超越了以往哲学家对人与社会关系孤立式的理解方式，// Abandoning and transcending the previous philosophers' isolated understanding of the relationship between man and society,**  
**EDU\_15 开始以崭新的姿态面向人类社会。// Begin to face the human society with a new attitude .**
- source: gum\_zh\_academic\_socialized



In the following example, we see three groups of *joint-list* coordinations at different hierarchical levels:

- between DU\_42-46 "socialized human being revealing" and DU\_47-52 "ideal of community answering ... and providing ..."
- among EDU\_43 "taking people's ...", EDU\_44 "economy's ..." and EDU\_45 "ideas'..."
- between DU\_47-51 "answering" and EDU\_52 "providing."

- 141. EDU\_42** “社会化的人类”概念就像一把哲学钥匙，// The concept of “socialized human being” is like a philosophical key,  
**EDU\_43** 以“人的社会化”为逻辑支点，// Taking “people's socialization” as the logical fulcrum,  
**EDU\_44** 以“经济的社会化”为思想内核，// Taking “socialization of the economy” as the core of thought,  
**EDU\_45** 以“观念的社会化”为价值诉求，// Taking “the socialization of ideas” as the value appeal,

**EDU\_46** 揭示了人类社会发展的必然趋势和追求目标。// It reveals the inevitable trend and pursuit of the human society development .

**EDU\_47** 人类命运共同体思想作为新时代重大的理论创新成果，// The idea of a community with a shared future for mankind is a major theoretical innovation in the new era .

**EDU\_48** 以“社会化的人类”概念作为思想来源，// Using the concept of "socialized human beings" as a source of thought ,

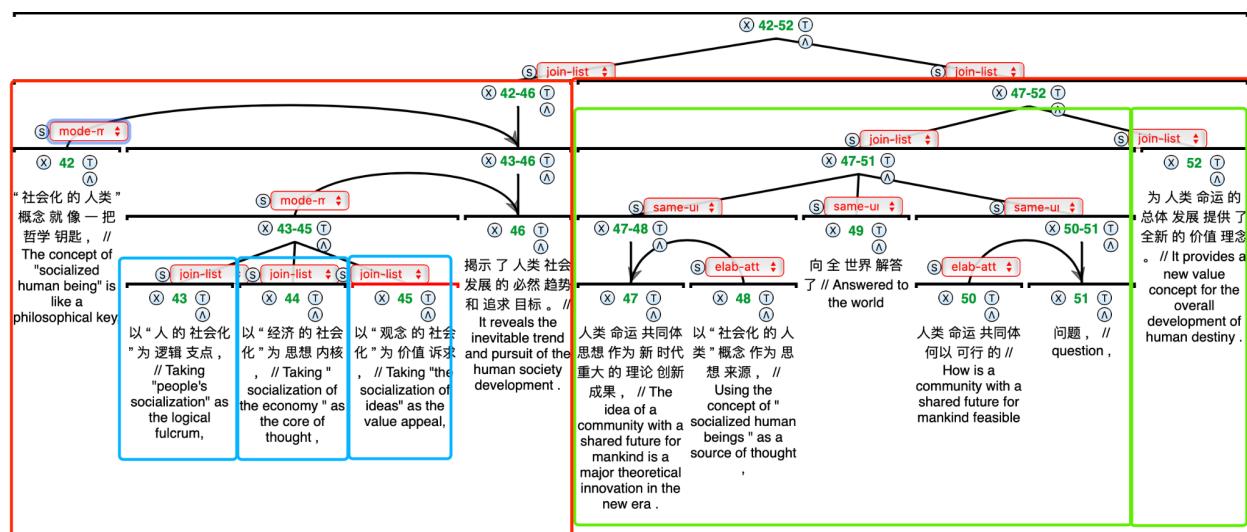
**EDU\_49** 向全世界解答了 // Answered to the world

**EDU\_50** 人类命运共同体何以可行的 // How is a community with a shared future for mankind feasible

**EDU\_51** 问题，// question ,

**EDU\_52** 为人类命运的总体发展提供了全新的价值理念。// It provides a new value concept for the overall development of human destiny .

source: gum\_zh\_academic\_socialized



## 2.2.4 joint-sequence

**joint-sequence: the Writer presents EDUs of chronological sequence.**

Coordinating conjunctions that follow each other in a time sequence. For example:

- coordinations signaled by date or time
- section titles ordered by a person's growing to death or an event's preparation to completion

The following example DU\_16-18 shows symptoms of rabies from the beginning to death.

142. EDU\_14 狂犬病影响大脑和脊髓 // Rabies affects the brain and spinal cord

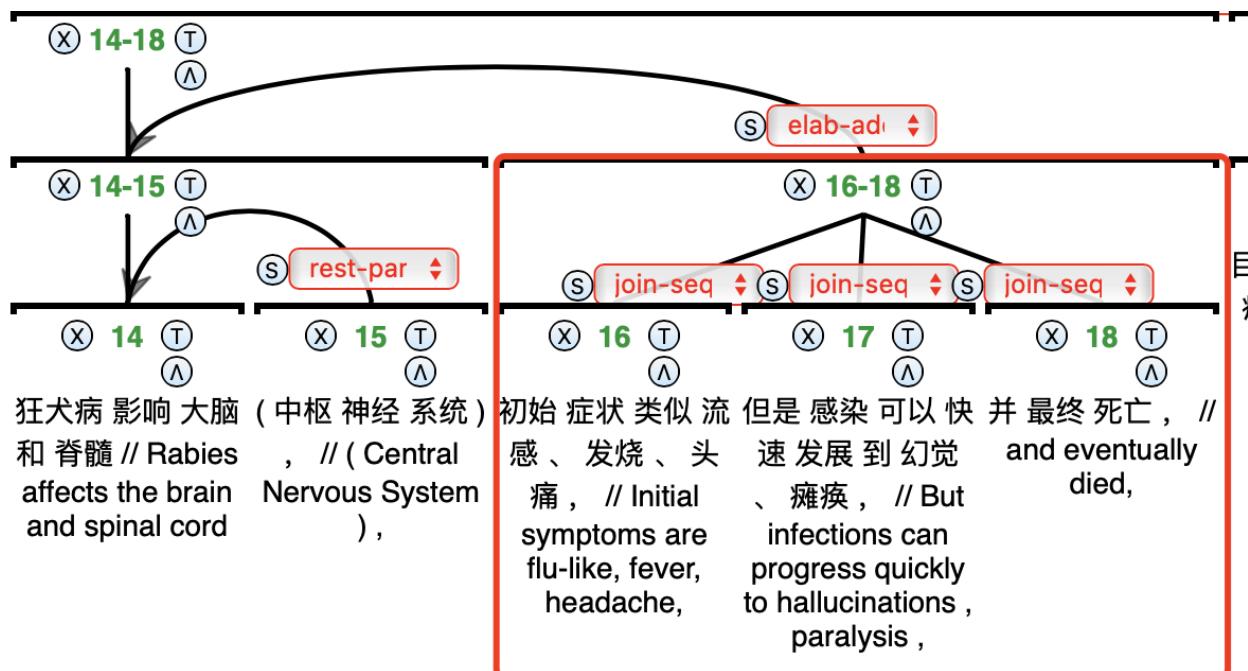
EDU\_15 (中枢神经系统)，// (Central Nervous System) ,

EDU\_16 初始 症状 类似 流感 、 发烧 、 头痛 , // Initial symptoms are flu-like, fever, headache,

EDU\_17 但是 感染 可以 快速 发展 到 幻觉 、 瘫痪 , // But infections can progress quickly to hallucinations , paralysis ,

EDU\_18 并 最终 死亡 , // and eventually died,

source: gum\_zh\_academic\_rabies



In the following example, we see three groups of *joint-sequence*:

- Among "Brighton High School," "Bachelor in Suffolk University" and "Master in Boston University"
- between "moving to Boston" and "studying at Suffolk University"
- between "transferring to Boston University" and "obtaining master degree"

143. EDU\_32 她 中学 就读于 布莱顿 高中 , // She attended Brighton High School in secondary school,

EDU\_33 2004年 毕业 后 , // After graduating in 2004,

EDU\_34 穆雷 搬到 波士顿 // Murray moves to Boston

EDU\_35 就读 萨福克 大学 的 心理学系 // Studied psychology at Suffolk University

EDU\_36 [8] 。 // [8].

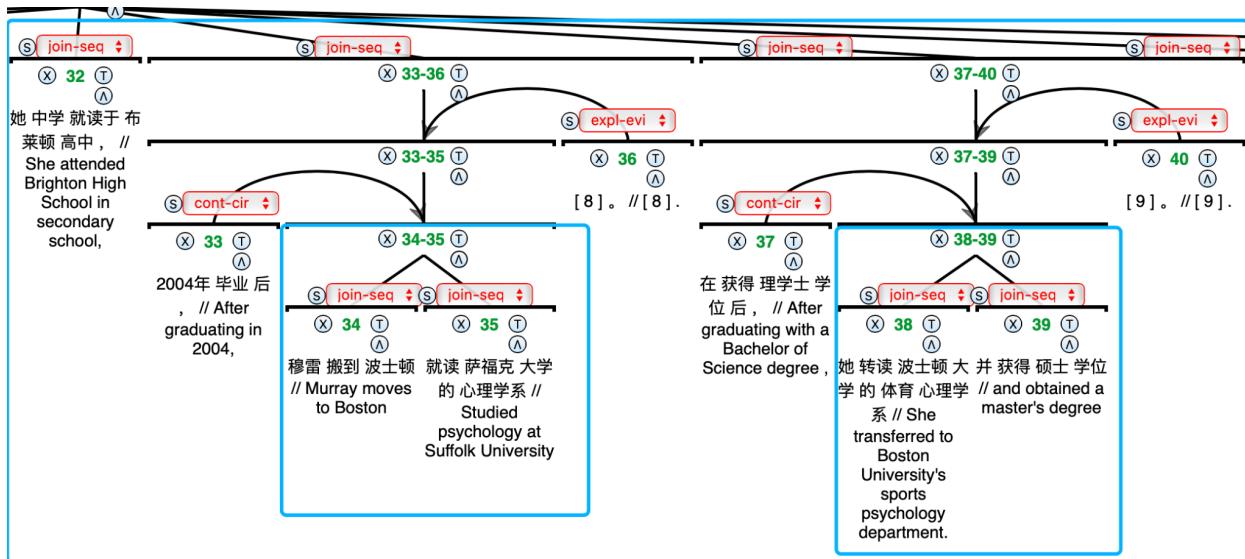
EDU\_37 在 获得 理学士 学位 后 , // After graduating with a Bachelor of Science degree ,

EDU\_38 她 转读 波士顿 大学 的 体育 心理学系 // She transferred to Boston University's sports psychology department.

EDU\_39 并 获得 硕士 学位 // and obtained a master's degree

EDU\_40 [9] 。 // [9].

source: gum\_zh\_bio\_marbles



## 2.2.5 joint-other

**joint-other: the Writer presents unlike units with no other relation.**

This is the last resort for multinuclear relations. When conjoined elements are neither listed nor of sequential order, we use this joint-other label. For example:

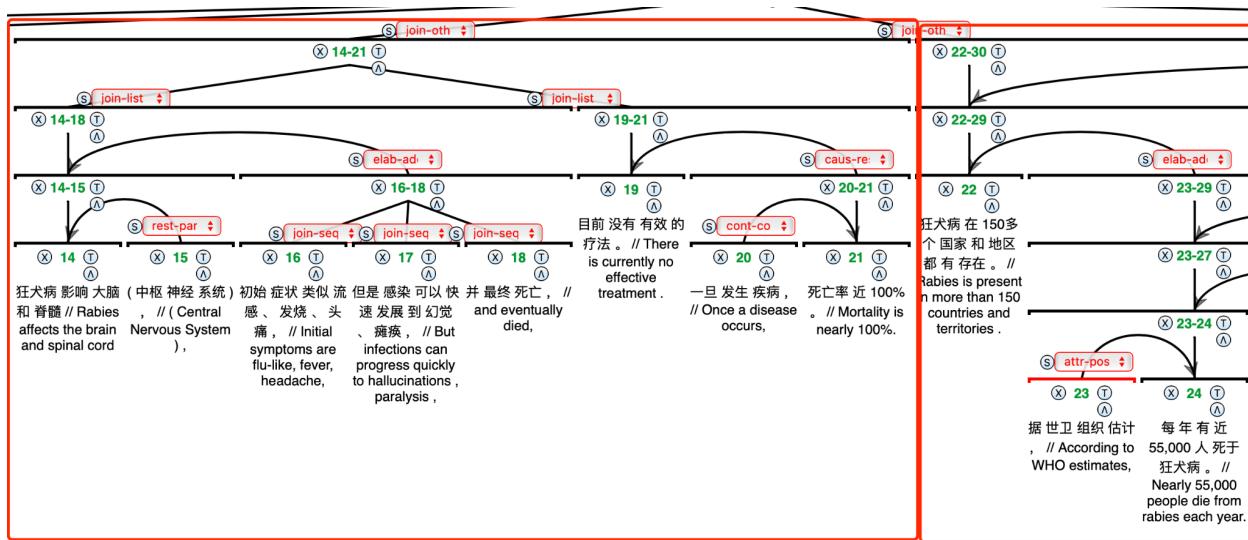
- section and subsection heads without enumeration or time sequence

*Joint-other* usually occurs between larger discourse units. Due to the length of these larger DUs, it is difficult to present a full example. The following is a snippet in a sequence of *joint-other* DUs and here we show a *joint-other* relation between DU\_14-21 and DU\_22-30 (only the nucleus part in the snippet).

144. EDU\_14 狂犬病影响大脑和脊髓 // Rabies affects the brain and spinal cord  
 EDU\_15 (中枢神经系统)， // (Central Nervous System),  
 EDU\_16 初始症状类似流感、发烧、头痛， // Initial symptoms are flu-like, fever, headache,  
 EDU\_17 但是感染可以快速发展到幻觉、瘫痪， // But infections can progress quickly to hallucinations, paralysis ,  
 EDU\_18 并最终死亡， // and eventually died,  
 EDU\_19 目前没有有效的疗法。 // There is currently no effective treatment .  
 EDU\_20 一旦发生疾病， // Once a disease occurs,  
 EDU\_21 死亡率近100%。 // Mortality is nearly 100%.  
 EDU\_22 狂犬病在150多个国家和地区都有存在。 // Rabies is present in more than 150 countries and territories .  
 EDU\_23 据世卫组织估计， // According to WHO estimates,

**EDU\_24 每年有近 55,000 人 死于 狂犬病 。// Nearly 55,000 people die from rabies each year.**

source: gum\_zh\_bio\_marbles



## 2.2.6 restatement-repetition

**restatement-repetition: the Writer presents equivalent or redundant units.**

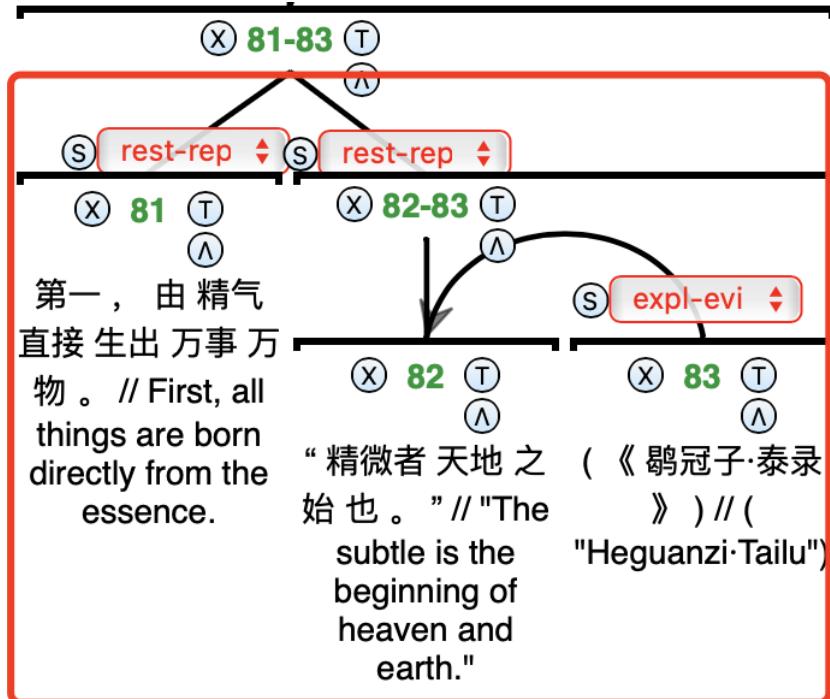
In the following example, EDU\_81 is a modern Chinese paraphrase of the ancient Chinese quote EDU\_82.

**145. EDU\_81 第一，由 精气 直接 生出 万事 万物 。// First, all things are born directly from the essence.**

**EDU\_82 “精微者 天地 之 始 也 。” // "The subtle is the beginning of heaven and earth."**

**EDU\_83 (《 鹬冠子·泰录 》 ) // ( "Heguanzi·Tailu" )**

source: gum\_zh\_academic\_taoist



In the following example, we see a repetition of "rare to see People Daily's comments" between DU\_141-143 and EDU\_144.

146. EDU\_140 2.2.2. 很少 回复 大众 评论 // 2.2.2. Few replies to public comments

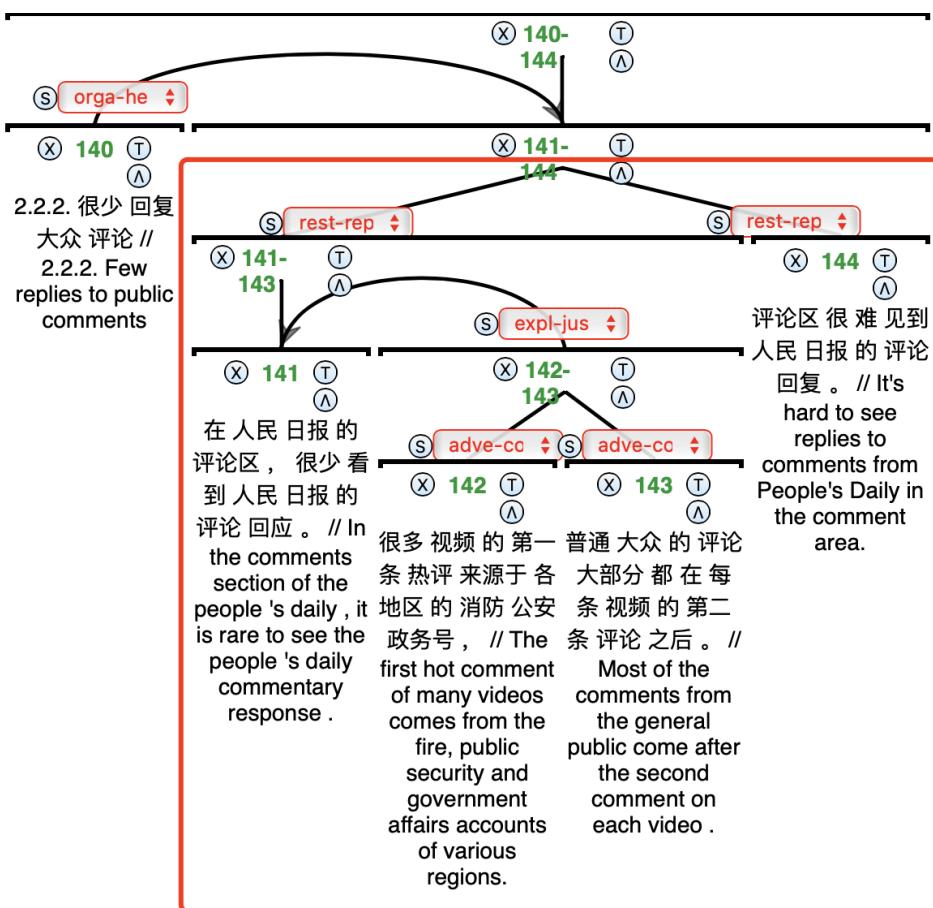
**EDU\_141** 在 人 民 日 报 的 评 论 区 , 很 少 看 到 人 民 日 报 的 评 论 回 应 。 // In the  
comments section of the people 's daily , it is rare to see the people 's daily  
commentary response .

**EDU\_142** 很 多 视 频 的 第 一 条 热 评 来 自 各 地 区 的 消 防 公 安 政 务 号 , // The  
first hot comment of many videos comes from the fire, public security and  
government affairs accounts of various regions.

**EDU\_143** 普 通 大 众 的 评 论 大 部 分 都 在 每 条 视 频 的 第 二 条 评 论 之 后 。 // Most  
of the comments from the general public come after the second comment on each  
video .

**EDU\_144** 评 论 区 很 难 见 到 人 民 日 报 的 评 论 回 复 。 // It's hard to see replies to  
comments from People's Daily in the comment area.

source: gum\_zh\_academic\_peoples

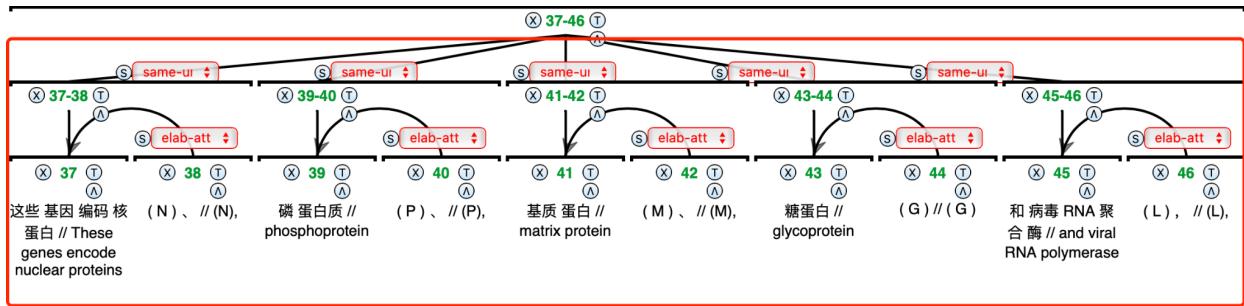


## 2.3 same-unit

**same-unit:** this is a technical device for interrupted EDUs.

In the following example, we see a single EDU broken up by five parenthetical abbreviations, forming a series of *elaboration-attribute* + *same-unit* relations.

147. EDU\_37 这些 基因 编码 核 蛋白 // These genes encode nuclear proteins
- EDU\_38 ( N ) , // (N),
  - EDU\_39 磷 蛋白质 // phosphoprotein
  - EDU\_40 ( P ) , // (P),
  - EDU\_41 基质 蛋白 // matrix protein
  - EDU\_42 ( M ) , // (M),
  - EDU\_43 糖蛋白 // glycoprotein
  - EDU\_44 ( G ) // ( G )
  - EDU\_45 和 病毒 RNA 聚合 酶 // and viral RNA polymerase
  - EDU\_46 ( L ) , // (L),
- source: gum\_zh\_academic\_rabies



Sometimes, due to the structure of modifications, it is more reasonable to have hierarchical *same-unit* groups than a one-level *same-unit* grouping.

In the following example, we see two levels of *same-unit* due to the *attribution-positive* to DU\_70-72 and parenthetical repetitions of EDU\_70.

148. EDU\_68 这说法 // this statement ||

EDU\_69 和罗拔·高云 (1923年)声称 // and Robert Gowan (1923) claimed ||

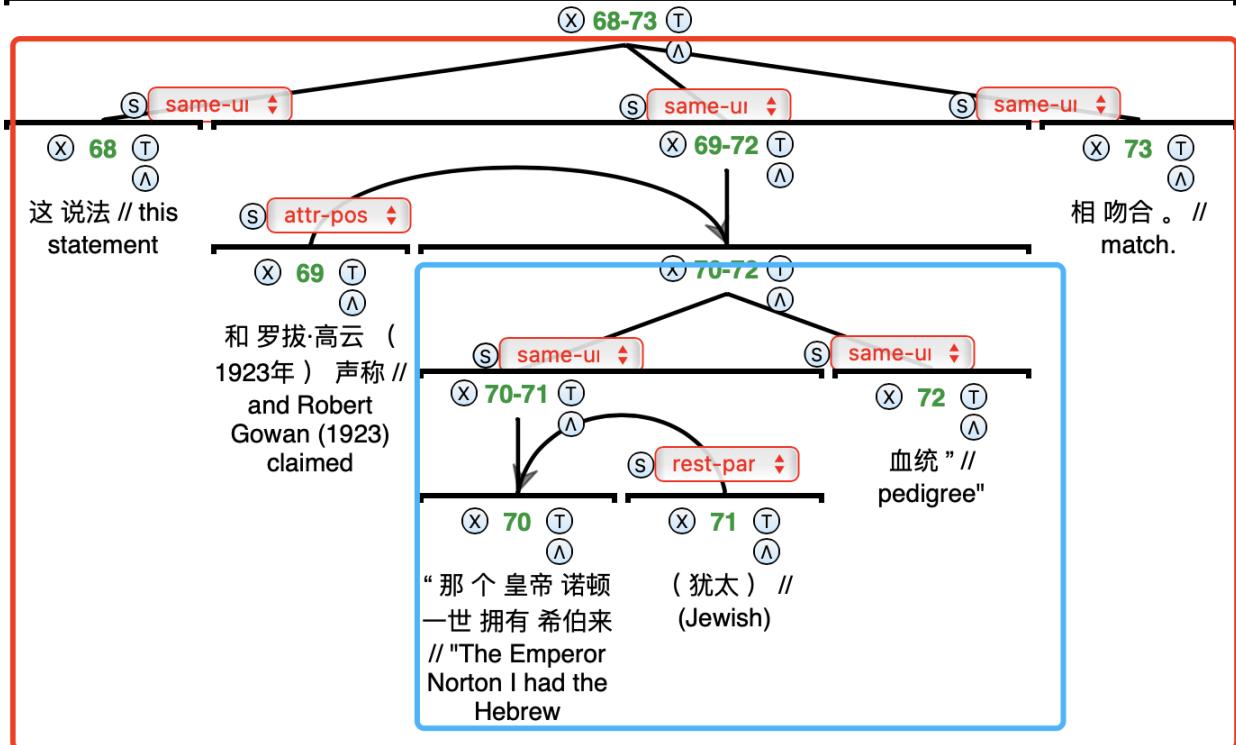
EDU\_70 “那个皇帝诺顿一世拥有希伯来 // "The Emperor Norton I had the Hebrew ||

EDU\_71 (犹太) // (Jewish) ||

EDU\_72 血统 // pedigree" ||

EDU\_73 相吻合。// match. ||

source: gum\_zh\_bio\_emperor



### 3. Others

#### 3.1 Questions for Amir

#### 3.2 TODOs

TODO: revise whow\_procrastination (wrong headedness (should head-initial for steps))  
also revise whow\_flirt

TODO: revise bio\_galois (French: XXX should be prep)

TODO: revise interview\_game for non-flat question-answer relations (when a follow-up question follows a detail from the previous answer)

c.f.

[1 Q: What food do you like? ]

[2 A: I like chickpeas. ]

[3 Q: what is chickpea? ]

[4 A: ... ]

[3-4] would be background to [2]

TODO: PP+的 should not be segmented! similar to English (they are un-tense-able)

- write a python script to search for :
- P + no-Vs + de and merge them with close by EDUs

TODO: Relative clause need to have predicate structure verb+noun

- e.g. adv + v is not relative clause

TODO interview\_keyman is a good document for @Amir to look over relation names...