

阅读报告

The Semantics of Comparatives and Other Degree Constructions

刘柱 中国语言文学 2022312212

摘要

本阅读报告是关于文献 (Schwarzschild 2008) 的阅读评述和心得体会。其中包括论文重述、汉语拓展和对大语言模型的评估。论文重述部分,一方面使用图的方式更清晰地展现了文中出现的重点概念“阈值”,并对文中涉及程度结构相关的关键例句逐一进行了分析;另一方面尝试使用经典的形式语义学

(Heim&Krämer 1998) 语言对论文中的关键部分进行重述。第二部分汉语拓展则考虑到汉语中的例子与文中的现象做对比分析。第三部分则评估了大语言模型对于程度结构语义的理解程度。

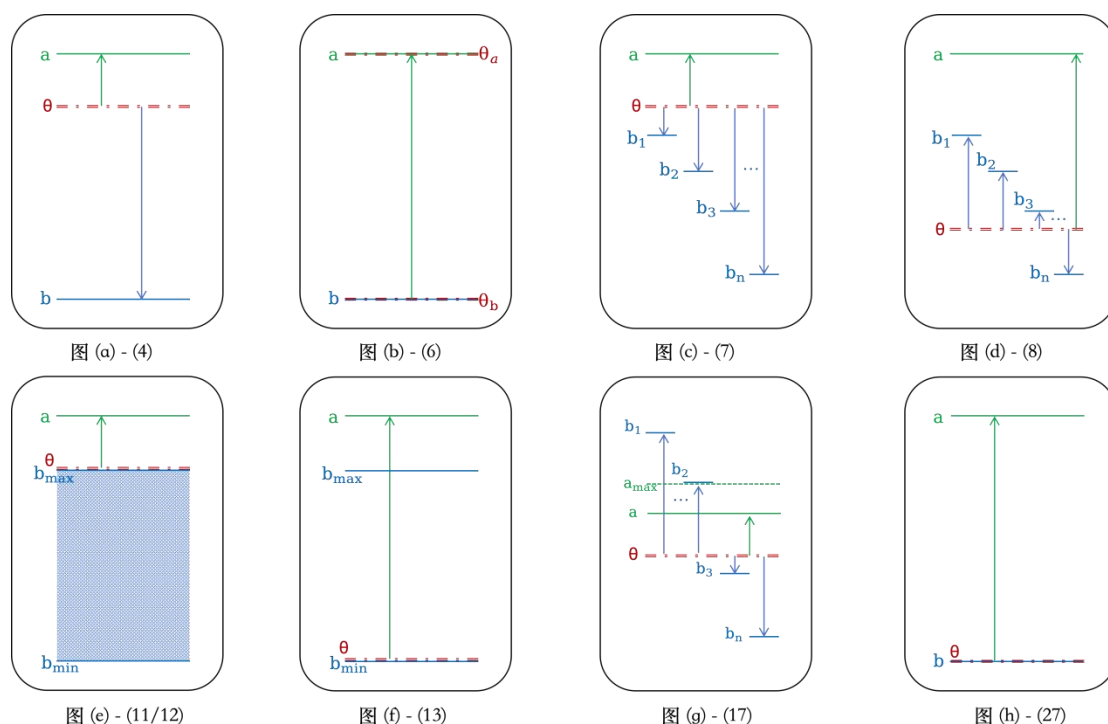


图 1: 典型的程度结构形式语义示意图¹

(一) 论文重述²

比较级结构 作者在形容词 A 里引入了阈值 θ 的概念, 来精准量化形容词 A 的程度。具体来说, 使用 $A(a, \theta)$ 来表示“实体 a 在 A 所描述的维度上满足或者超过了 θ ”。将这一符号引入到比较级结构中, 从而形成了公式 (4) 中的“A-NOT-A”的结构分析, 即存在某一阈值 θ , a 大于等于 θ 但是 b 小于 θ 。

这一结构可以使用图 (a) 来清晰地可视化出来。图中使用绿色和蓝色的实

¹ 每个图编号后的数字表示该图说明的典型例子在论文中的编号。

² 为方便描述和节约空间, 无特别说明, 本报告直接引用原论文中公式或例子编号。同时不论具体例子中涉及的具体名词, 使用小写字母 a 和 b 分别表述句子主语和补语中的被比较对象。文中的 meet or exceed 在报告中被表达为等价的“大于等于”, 其反面被称作“小于”。与原例句对比的新例句, 报告在原有编号后增加了 a 以表示区别。

线分别表示实体 a 和实体 b 在形容词刻画的属性上的数值，其中上面刻度的数值大于下面的，红色虚线表示的数值为阈值 θ ，箭头方向反映了实体反映的数值高于（向上箭头）或者低于（向下箭头）阈值，长度则反映了高过或者低于阈值的程度。图（a）体现了“A-NOT-A”的分析，两个实体针对同一阈值具有不同的方向。

比较级结构另外一种表达方式采用的是与形容词名物化以及双阈值的方式。针对例（5）中的语义，可以将其中的 longer 转变为 length，而实体 a 的 length 的语义可以使用“存在的最大阈值”或者“a 满足/等于的那个值”来表达。因此涉及的两个实体都具有相应的阈值，而例（5）自然可以表达为“a 的长度/高度大于 b 的”（例 6）。图（b）体现了转化后的语义比较，其中 θ 的下标反映了各自实体对应的那个“最大阈值”，在图中体现为红色虚线和实体实际的值重合。

笔者认为第二种表达（图 b）较之第一种表达（图 a）更为自然，反映了比较级一般和直接地理解，但额外引入了更多的阈值变量。而第一种表达则引入了否定算子，让分析更加复杂。但是表达方式仅涉及一个公共的阈值变量，更加简约，且后文举到的其他现象都可以用否定算子更精准地表达出来，因此还是一种更占优势的方案。

例（7）（8）则涉及被比较对象 b 不仅是一个实体，而是多个时相关表达的语义。其中（7）对应的否定表达为（9），其中 on any other day 出现在否定算子 not 的外部，而（8）对应的否定表达为（10），其中 at least one day 出现在否定算子 not 的内部，因此前者要比后者更加程度深。

图（c）和图（d）分别表达了例子（7）和例子（8）的语义。图中涉及的 b 的 N 个对象假设按照值大小从大到小排序，在图中显示为从高到低。每个对象下的索引代表在排序中的大小位置。图（c）反映了例子（7）对应的否定转化（9）中其他所有的对象都小于阈值 θ ；图（d）则反映了例子（8）对应的否定转化（10）中提到的至少有一个对象小于阈值 θ 。

笔者认为，从图中很容易观察到，（7）和（9）的语义表达可以通过引入最大值 MAX 算子和 A-not-A 分析来等价表达，（8）和（10）的语义表达则可以引入最小值算子 MIN 来等价表达，该算子作用于集合 b，用于找出形容词属性下的相应的最大值和最小值。相应的等价表达可以参考下面的（7a）和（7b）：

$$(7a) \exists \theta \text{ high}(a, \theta) \wedge \neg \text{high}(\text{MAX}(b), \theta)$$

$$(7b) \exists \theta \text{ high}(a, \theta) \wedge \neg \text{high}(\text{MIN}(b), \theta)$$

也就是说，当一个值比一个集合中最大的值还大时，表明该值大于集合中的所有值；如果一个值至少要比集合中的一个值大，那么它一定大于集合中的最小值。值得注意的是，该算子与例（28）中要引入的 MAX 算子作用的对象不同，后者作用于阈值。

后面的例子涉及否定算子后 than 后补语中情态动词的范围关系。对于例（11）中的 allowed to，它在 not 算子的作用域之内，即 not allowed to。而并不是 allowed to not。图（e）反映了这两种区别。图（e）中的阴影部分为假定实体 a 被允许的范围，它被 b 的最小值 min 和最大值 max 限制住，其中阴影包括 min 和 max 边线。not allowed to 对应于 max 往上的区域（不包含 max），它也是例子（11）中阈值应该在的区域。而 allowed to not 则处于阴影之内，而正确的语义告知我们阈值不应落在这一区域。同样地，例子（12）中的 should 也对应图（e）中阈值的

取值。而例子 (13) 中的 have to 中的阈值则对应图 (e) min 线往上的区域, 即图 (f)。注意图 (f) 中的阈值虚线并未与 min 对应的蓝色实线重合。因此相对于 should 而言, than have to 的结构表达更弱。例 (15) (16) 中的 said, promise 的分析与 allowed to 的分析类似。

同样地, 当被比较的对象涉及多个时候, 需要考虑到它们的排序。例 (17) 中的 (17b) 表达得太弱了以至于无法准确捕捉到 (17a) 就是因为没有考虑到被比较的对象 b 处于最低的高度, 而比较的对象 a 处于最高的高度时候, 阈值无法刻画 exactly 的含义。论文中给出的方案是将 (17b) 中的阈值 θ 更改为 a 的长度, 因此 a 和被比较的对象之间就无法再有其他对象了。

笔者提供了另一种修改思路。如图 (g) 所示, 当按照数值高低次序排列好之后, 发现当阈值处于倒数第 2 和第 3 之间 (可以触碰第 3, 但是不能触碰第 2) 的时候, 也可以如实反映 (17a) 中的语义, 因此规则改为如下 17c, 其中 n, n-1, n-2 分别表示依照数值大小倒数第 1, 第 2, 第 3:

$$(17c) \exists \theta \text{ high}(a, \theta) \wedge \neg \text{high}(b_n) \wedge \neg \text{high}(b_{n-1}) \wedge \text{high}(b_{n-2})$$

其他程度相关的构式 在分析了典型的比较级结构后, 作者又分析了其他和程度相关的结构。在等同类的 as 中 (27) 反映了 as...as 的形式语义。其中将 A-not-A 中的 not 去掉, 并将 b 的阈值修改为 b 的最大阈值, 或者高度。可以参考图 (h)。这主要是考虑到单用 b 的阈值无法刻画被比较的基准线。

需要注意的是 as...as 的表达并非如直觉理解的“和……一样” (exactly as...as), 而是反映的是 taller or exactly as...as, 文中提到这一理解是 not as...as 的取反的结果, not as...as 就是 less taller than 的含义。确实在中文的一些语境下“和……一样”也可以直接理解为 taller or exactly as...as。例如运动员在看到比他更高的选手时候, 不由感慨: “我和他一样高就好了”, 那么应理解为该运动员希望拥有不低于对方的身高值。

另一方面, 笔者认为这也提示了在分析一般的比较级结构时候采用了 not 算子而不直接使用例 (27) 中肯定表达的一个原因: 用于排除掉 exactly as...as 的情况。在图 (a) 展示的图中可以清晰地看到, 由于一个阈值的存在, a 的数值线不会和 b 的数值线重合。

其余程度副词例如 (36) 所列举的分析类似于 as 的分析, 这里就不加赘述。

第二类过量结构中使用了 too, 表示“太……以至于不能……”, 这里除了否定算子的运用, 还额外增加了因果结果 BECAUSE, 同时受限于 too 表达的情景, 其中与否定词相搭配的情态动词也有所不同。如 (42) 到 (45) 所表达的。

笔者认为 too...to...结构是汉语对译中最可以反映否定词 not 存在的证据, 理由是汉语在翻译 too to 的结构时候不得不显式地将否定词“不”甚至情态动词补充出来, 下面罗列了 (42) - (45) 中的中文翻译:

(42a) 太阳太热了而不能触摸。

(43a) 对他来说伤口太小了, 因此不可能是使用霰弹枪造成的。

(44a) 杰克太懒了而不愿意讲解步骤。

(45a) 她太年轻了不应该结婚。

可以看到它们虽然在英文中没有任何 not 和情态动词的显性表现, 在汉语中却有着很明显的标记。推而广之, 也可以作为 not 在表示程度结构中存在的理由。

量 论文之后分析了与形容词的程度量相关的副词，例如 a bit, much 等，它们可以视作是对阈值的范围的宽窄做了规定，或者直接规定了阈值的选取集合，例如 2 pounds heavy 之类的具体数量。

非形容词的比较结构 除了形容词外，还有比较级加在副词、名词、动词等词类上的，但是作者在英语的案例中倾向于认为名词、动词都是形容词 many 和副词 much 的比较级，即 more coffee 相当于 much-er coffee。在汉语中，存在比较级直接修饰名词或者动词的情况，例如对于比较句中常见的“更”来说：

- (a) 更有利于心理健康
- (b) 更像是机器人说话
- (c) 更想听大家的互动讨论
- (d) 有美育，更清华
- (a-c) 都可以视作“更”修饰动词，(d) 视作“更”修饰名词。

综上，本报告采用图示的方式，更加直观地刻画了阈值以及 A-not-A 的分析思路，同时还可以引发诸如 (7a) (7b) (17c) 的新思考。

Heim&Krarzer 形式语言重述 (Heim&Krarzer1998) 中使用 λ 算子来表述形式语言，本报告对论文中的比较级所涉及的形式语言用 λ 算子重新表达。本文对原文例 (2) 中的例句进行两种语义分析：

(2) A is more expensive than B

第一种将 more expensive than 看作是整体，它相当于一个及物动词，是一个 $\langle e, et \rangle$ 类型的函数。其中 more expensive than 的语义可以表示为 (e)：

(e) $\| \text{more expensive than} \| = \lambda x. \lambda y.$ 存在一个阈值，y 的价格大于等于该阈值，同时 x 的价格小于该阈值

其中我们采用汉语作为元语言。“价格”表示在 expensive 属性上的名词化表达。根据 (e) 很容易将例 (2) 的语义分析出来。

$\| (2) \| = \| A \text{ is more expensive than } B \|$
 $= \| A [VP \text{ is } [AP \text{ more expensive than } B]] \|$
 $= \| VP \text{ is } [AP \text{ more expensive than } B] \| (A)$
 $= \| \text{is} \| (\| AP \text{ more expensive than } B \|) (A)$
 $= \| AP \text{ more expensive than } B \| (A)$
 $= \| \text{more expensive than} \| (B) (A)$
 $= 1 \text{ iff 存在一个阈值，A 的价格大于等于该阈值，同时 x 的价格小于该阈值。}$
最终结果符合正式的语义。

第一种方式没有把形容词单独拿出来，是一种简约的表达。另一种方法则将每一部分都视作是独立的语义，笔者采用 (59) 中的语义表达式，它将 than 后的句子补充完整。

(59) A is more expensive than B is expensive

首先需要明确此刻词库中的词汇：

- (a) 实体：A 和 B
- (b) 真值：0 和 1

(c) 为了定义形容词 expensive 我们引入新的类型，命名为：程度 d，它是一个大于 0 的连续值。

由此, expensive 可以被当作是 $\langle d, \langle e, t \rangle \rangle$ 类型的, 其定义如下:

$\| \text{expensive} \| = \lambda d. \lambda x. x$ 的价格大于等于值 d 。

(d) more 是一个 d 类型的, 目的在于引入针对形容词的程度

(e) than 是一个 $\langle t, \langle d, et \rangle \rangle$ 类型的, 其定义如下:

$\| \text{than} \| = \lambda t. \text{not } t$

这一定义主要是考虑到 than 之后的补充语成分 (即第一个参数) 中隐式包含了 not

(e) is 是一个恒等函数。

此外采用如下的 Logical Form 来刻画结构:

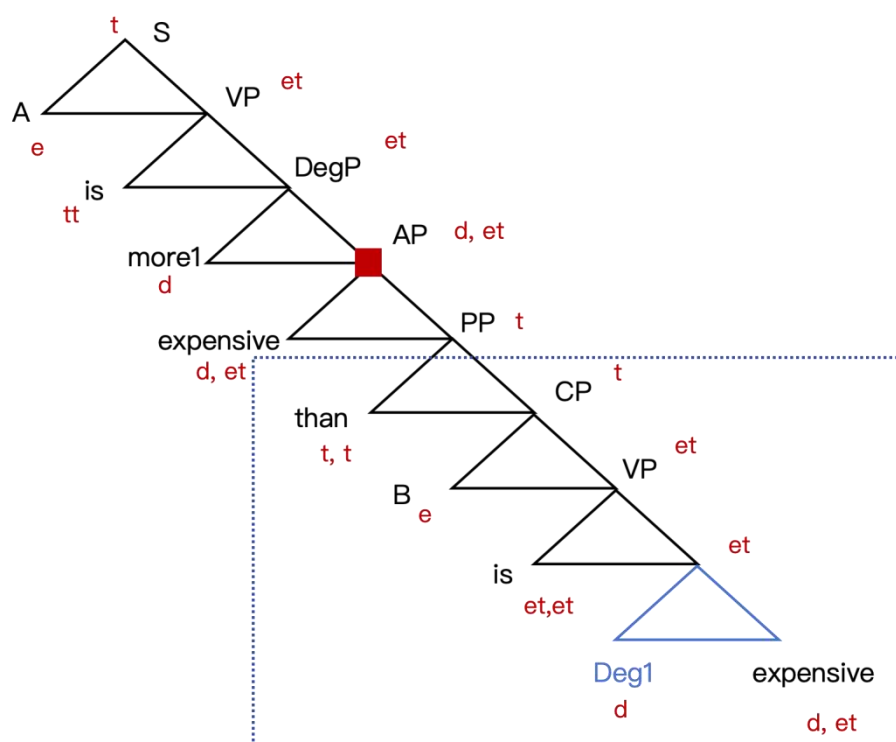


图 2: 例 (59) 比较级结构中的 Logical Form

图 2 画出了例 (59) 的句法结构树³, 除了句法结点外, 还额外标识了结点类型。注意在 than 之后的从句中, expensive 前额外补充一个 trace, 命名为 Deg1, 它是程度类型, 与之前 more 中提供的程度一样。同时为了使得组合顺利进行下去, 我们还需要在红色方形标识的 AP 处, 额外引入一条类似与 Predicate Modification 的规则, 这里称之为 Degree Predicate Modication (DPM) 。

(DPM) Degree Predicate Modication

如果 A 是一个分支结点, B 和 C 是它的子女结点, 且 $\|B\|$ 是 $\langle d, et \rangle$ 类型, 而 $\|C\|$ 是 t 类型的那么 $\|A\|$ 是类型同样是 $\langle d, et \rangle$ 类型的, $\|A\| = \lambda d. \lambda x. \|B\|(d)(x) = 1 \text{ and } \|C\| = 1$ 。

这样我们便可以通过规则来组合生成句子 (59) 的语义了。首先将虚线部分的 PP 的真值求出来:

³ 三角形仅为作图方便, 并非句法树中省略子树的标记。

$$\begin{aligned}
\|PP\|^{d=1} &= \| \text{than} [CP B [VP [is [Deg1 expensive]]]] \|^{d=1} \\
&= \| \text{than} \| (\| CP B [VP [is [Deg1 expensive]] \|) \|^{d=1} \\
&= \| \text{than} \| (\| CP \| is [Deg1 expensive] \|) \| (B)^{d=1} \\
&= \| \text{than} \| (\| CP \| is \| (\| Deg1 expensive \|) \|) \| (B)^{d=1} \\
&= \| \text{than} \| (\| CP \| Deg1 expensive \|) \| (B)^{d=1} \\
&= \| \text{than} \| (\| CP \| expensive \| (d) \|) \| (B)^{d=1} \\
&= \| \text{than} \| (\| expensive \| (d) \|) \| (B)^{d=1}
\end{aligned}$$

$$\begin{aligned}
\|S\| &= \| A [VP is [DegP [more [AP [expensive PP]]]] \| \\
&= \| VP \| (A) \\
&= \| DegP \| (A) \\
&= \| AP \| (\| expensive \| (d)(A), \| PP \|^{d=1}) \\
&= \| AP \| (\| expensive \| (d)(A), \| \text{than} \| (\| expensive \| (d)(B) \|)) \text{ (trace rule)} \\
&= \| expensive \| (d)(x). \text{iff } x \text{ 的价格大于等于 } d, \text{ 且 } \| \text{than} \| (\| expensive \| (d)(B) \|) = 1 \| (d)(A) \\
&= 1. \text{iff } A \text{ 的价格大于等于 } d, \text{ 且 } B \text{ 的价格小于 } d
\end{aligned}$$

(二) 汉语中的程度结构 程度结构广泛存在于各个语言中，汉语中也具有相应的表达形式，下面列举了文中提到的程度结构所对应的汉语

- (1) 比较结构：A 比/相较于/比起/甚于 B 更/还 Adj
- (2) 形容词的量化修饰语：很少，一点儿，很多，相当多，2 厘米等
- (3) 补语的修饰语：任意一个，某一些，某一个，刚好一个等
- (4) 其他程度结构：太……以至于（不能），和 B 一样 Adj，最 Adj，没有 B Adj

通过使用论文中阈值的方式同样也可以分析汉语的比较级语义，例如下面的例句：

(a) 他比我高

可以分析为：存在一个阈值，他的高度值大于等于该阈值，而我的高度值小于该阈值。（图 a 所展示的）

汉语中的“太……以至于（不能）”结构将否定词“不”显式地表达了出来，这也为隐含的否定词提供了证据。

此外在论文中非形容词性的比较结构中，汉语也存在修饰更加广泛的情况，详见上面分析。

(三) 大语言模型是否可以理解到论文所表示的形式语义？ 笔者使用当前性能最好的 ChatGPT⁴进行测试。参考自然语言推理任务（Natural Language Inference），设置了如下的提示语。

大模型的输入提示语：

Can <P> mutually infer that <H>?

Give me a short answer, and a short explanation.

其中<P>是推理的前置条件句，<H>是推理的假设句，mutually infer 表明在探测“相互语义等价”的关系。

本报告选取了论文中的一些代表例句进行分析，如该句完全出现在原始论文

⁴ <https://chatgpt.com/>

中，为了节约空间，使用（）里填充句子编号的形式，实际输入模型的是完整的句子。以下推理中，使用紫色标注了原本答案应该是正确的，结果模型认为是错误的；红色标注了原本答案是错误的，结果模型认为是正确的。同时绿色标出了模型在解释中出现的错误。为了节约空间，仅罗列错误答案，全部答案请访问该网址⁵。之后有笔者简要分析。

(LLM1) Can (2) mutually infer (3)?

Short answer: No.

Short explanation: Saying “A is more expensive than B” only compares their relative prices; it does not **imply any absolute threshold of expense**—both could be cheap, both could be expensive, or any combination. There’s no necessary “cutoff” involved.

[笔者] 模型把阈值认为是一种绝对度量，根本上不认可这种方法。

(LLM2) Can (5a) mutually infer (6a)?

(LLM3) Can (7) mutually infer (9)?

Short answer: No.

Explanation: The first sentence only says today’s height is the highest relative to other days; it does not require a specific threshold θ to exist. Inferring a threshold θ and claiming the balloon did not meet it on other days adds an extra, unnecessary assumption.

(LLM4) Can (8) mutually infer (10)?

(LLM5) Can (11) mutually infer “it meets or exceeds a threshold that it is not allowed to meet or exceed” ?

(LLM6) Can (11) mutually infer “it meets or exceeds a threshold that it is allowed not to meet or exceed” ?

Short answer: Yes.

Explanation: Saying “The balloon is higher than it is allowed to be” implies it has surpassed the maximum allowed height. This is logically equivalent to stating it meets or exceeds a threshold it is **not supposed to** reach.

[笔者] 模型似乎不理解 not 和 allowed 位置关系表达的语义差异，把 allow not to 和 not supposed to 等同了起来。

(LLM7) Can (16a) mutually infer (16b)?

(LLM8) Can (17a) mutually infer (17b)?

Short answer: Yes.

Short explanation: The English sentence means there exists a height (θ) where balloon A is at θ , and exactly two other balloons are below that height. This **directly**

⁵ https://juniperliuzhu.netlify.app/projects/res_appendix.pdf

corresponds to “ $\exists \theta \text{ high}(a, \theta) \wedge$ exactly two other balloons y such that $\neg \text{high}(y, \theta)$.”

(LLM9) Can (26) mutually infer (27a)?

Short answer: No.

Explanation: “A is as tall as B is” only states a relative comparison between A and B—it means ($\text{height}(A) = \text{height}(B)$). It does not specify any absolute threshold or the maximal height that B meets. Inferring an existential statement about a “highest threshold” for B goes beyond what the comparison alone guarantees.

[笔者]与一般的直觉类似，as as 也被理解为 exactly as as. (27b)和 (27c) 也是一样的问题

(LLM10) Can (26) mutually infer (27b)?

(LLM11) Can (26) mutually infer (27c)?

(LLM12) Can (31) mutually infer (32)?

Short answer: No.

Short explanation: The first statement compares A’ s current height to the heights of other balloons specifically at six o’ clock, not to their maximum heights over all times. The second statement claims A’ s height equals the maximum height of every other balloon at some time θ , which is stronger and not guaranteed by the first.

(LLM13) Can (39) mutually infer (40)?

综上，模型整体回答的准确率并不高，13 个案例中仅有 5 个完全正确，同时观察 LLM1 和 LLM3 的评测结果可以推断出模型从根本上不太认可阈值和否定的方法，更倾向于从正面理解比较句。尽管这也是一种直观的理解方式，但是没有本文提出的方案这么优雅和普适性。

余论

本篇读书报告分析了文献 (Schwarzschild 2008) 中对于程度结构的语义分析，首先使用可视化的方式更直观清晰地展现了文献的内容，其次使用课堂所学的形式语义学文献 (Heim&Kratzer 1998) 中的术语对比较级进行了描述，之后针对汉语中的现象举例进行说明，最后设计提示语对 ChatGPT 模型进行了评估。本文提出的方法通过使用阈值和否定的形式，简洁明了地归纳了程度结构的语义，较之一般分析更加优雅和对称。该方法同样适用于汉语中相关结构的分析。大模型的结果表明在深层次理解语义方面，仍存在缺陷。

参考文献

(Heim&Kratzer 1998) Heim I, Kratzer A. Semantics in generative grammar[J]. 1998.
(Schwarzschild 2008) Schwarzschild R. The semantics of comparatives and other degree constructions[J]. Language and linguistics compass, 2008, 2(2): 308–331.