

句子生成顺序与三种 Wh-移位异常: 熵的视角*

李亚非

提要 本文讨论两类句法现象:(1)所有已知的语言都是先说结构位置高的成分,后说位置低的成分。(2)英语 wh 疑问句的造句规则有三个例外:所有者为疑问词时不能单独移位,必须把包含它的整个 DP 拖带到句首;主语为疑问词时不使用无语义助动词 do;结构位置在 VP/vP 之外的所有副词类都没有疑问形态。第一个现象基本上被目前的句法理论所忽视。第二类现象所涉及的三种句法结构从表面来看似乎互不相关,但是都分享一个特征:符合疑问句造句规则的都是从实词环境向虚词环境移位;而一旦移位的起点和终点都是虚词环境就会发生上述各种例外。本文通过热力学第二定律(即能量自然耗散)和句法运算的经济原则,对以上各种现象提出一个统一的解释。

关键词 句子生成方向 wh 疑问句 拖带移位 副词类 热力学第二定律 句法运算的经济原则

Universal Direction of Sentence-Generation and Three Odd Cases of Wh-movement: An Entropic Perspective

LI Yafei

Abstract Two types of syntactic phenomena have no explanations. The first is the fact, made explicit in Kayne (1994), that structurally higher constituents are always produced earlier in actual sentence-generation. Kayne's own effort to explain this, i.e. his tentative answer to why the Linear Correspondence Axiom holds to begin with, amounts to shifting the question to several assumptions that are based on dubious analogies or whose plausibility is difficult to assess independently. The second phenomenon consists of three exceptions to the general rules of English wh-question formation: (i) a wh-possessor must pied-pipe the entire DP, (ii) a wh-subject does not trigger do-insertion, and (iii) all adverb classes from the subject-oriented upward have no legitimate wh-counterparts (cf. *how carefully* vs. * *how evidently*). (i-iii) all involve the wh-word originating inside a functional context (DP, TP, various FPs, respectively) and aiming at another functional context (CP), a generalization not yet made in the literature. There is also reason to believe

* 感谢陈哲博士(美国威斯康星大学语言学系)和李钦策副研究员(哈尔滨工业大学计算机科学与技术学院)分别对本文的语言学部分和物理学部分所提出的建议。文章的行文及表述得益于《当代语言学》责任编辑王伟副研究员及中国科学院一位匿名热力学专家的宝贵建议。特此致谢。

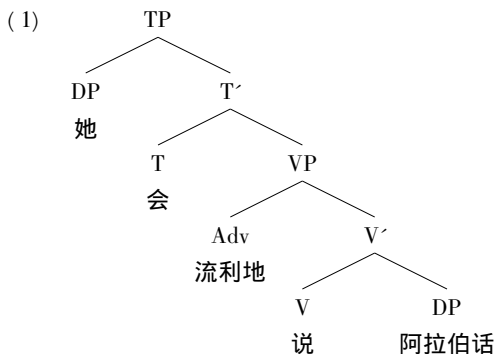
that these “oddities” are not specific to English. Possessor-pied-piping, for instance, has been suggested to be parametric (Gavruseva 2000), with Tzotzil, Chamorro and Hungarian presumably allowing the *wh*-possessor to move alone. A closer look at the data reported in the original works, however, turns out extra facts in all these languages that question whether the claimed extracted possessor indeed originates inside the possessee DP: Tzotzil possessors, *wh* or not, trigger indirect object agreement on the verb (Aissen 1996); Chamorro possessor movement requires the possessee phrase to have no determiner (Chung 1991) and looks more like Massam’s (2001) pseudo-noun-incorporation; the well-known Hungarian dative *wh*-possessor actually has a missing link that no non-*wh* pronominal possessor is permitted at the left edge of the possessee DP (É. Kiss 2002), making one wonder where the extracted *wh*-counterpart gets its dative case in the first place. Drawing on the fact that lexical words occur at an earlier age than functional words in language acquisition, this paper hypothesizes that the former type is more basic and takes less energy to process than the latter type. When combined with the copy theory of movement and the minimalist assumption that syntactic operations are sensitive to computational cost, we prove mathematically that under the second law of thermodynamics, top-down sentence-generation is always more energy-efficient than bottom-up and therefore must be favored. The afore-mentioned exceptions to *wh*-movement as well as other related facts are shown to follow from this theory too.

Keywords direction of sentence-generation, *wh*-questions, pied-piping, adverb classes, 2nd law of thermodynamics, sensitivity to computational cost

本文的内容可以归纳为一、二、三。目前知道的所有自然语言都按照同一个顺序造句，以通常的句法树形图来说就是永远自上而下，此乃一。物理学中定义熵的法则是热力学第二定律，权作二。而文中讨论的异常 *wh*-移位有三类，是为三。

1. 高先低后的句子生成顺序

首先声明，这里所讨论的句子生成顺序，是指实际的造句过程，而非最简方案理论里的由小到大逐步融合造句成分的理论顺序。以汉语为例(暂时忽略无关节):



主语“她”最先被说出，同时也是结构图里位置最高的成分；助动词“会”是句子中的第二个词，位置仅仅低于主语；依此类推。问题在于，似乎没有任何语

言和汉语同构却以反向顺序由低向高造句(即“阿拉伯话说流利地会她”)。一些语言(比如爱尔兰语)的基本语序是动-主-宾(相当于“说她阿拉伯话”),但近年的研究表明,这并非从低处的动词先说起,而是把动词上移到结构上高于主语的位置,句子的第一个词仍旧是最高的成分。为了便于讨论,我们把这个现象称为“高先低后”。

高先低后在语言中没有例外^①,所以其背后肯定有一个根本性的原因。在目前的理论中,这个原因是未知的。有人可能会说,每个句子中先说的是旧信息,后说的是新信息。但是焦点信息是讲话者设定对方尚且不了解的内容,是新信息。而近年来的句法制图理论(cartographic syntax; 可参考 Rizzi(2004)中关于这个题目的多篇文章)有大量证据说明,焦点在句子中的结构位置非常高,在主语之上,比如“你的那篇文章我已经读过了,她的还没有。”

另一种可能性是先行词要“先行”。这个说法不但有合理之处,而且可以从更深的层次获得支持。Julien(2000)做过一番统计,发现被调查到的语言中,基本句的构成要么需要把 vP 中的主语上移到 TP 的位置(比如英语和汉语),要么需要把动词上移到 T 的位置(比如爱尔兰语),要么是两者都上移(比如法语),唯独没有语言完全不移动,即基本句的语序是主-动-宾(因为所有的成分都停留在 vP 和 VP 中的初始位置),而带有助动词的句子语序是助-主-动-宾(助动词的初始位置在 vP 之外,其余同基本句)。Baker(2002)和 Li(2005)都试图对此加以理论上的解释。但无论最终的原因是什么,从 Julien(2000)的发现中可以得出一个结论:即人类语言中的句子(clause)都涉及成分移位。如果用 Chomsky(1995)的拷贝理论(copy theory)来表述,则每个句子必然包含下述结构:

$$(2) \text{ [句子} \cdots X \cdots \text{ [}_{vP} \cdots X\text{-拷贝} \cdots]]}$$

其中 X 是被移动的成分, X-拷贝对应于形式理论中经常用到的语迹(trace)。用更通俗的术语来说, X 是 X-拷贝的先行词。此外,已知移位的终点必需成分统治(c-command)起点,即 X 必需成分统治 X-拷贝。如果(2)是所有语言的句子结构,而先行词必须先行,那么可以由此推论出句子只能由高向低生成,因为先行词 X 注定高于 X-拷贝。

① 后核心语言(=宾语在动词之前)并非明显的反例,尽管宾语内部的成分肯定比动词的位置更低。为了理解这一点,我们先看做主语的名词短语。其内部成分肯定不会高于句子中的某些成分(比如 T),但出现顺序却先于 T。正如 Kayne(1994)所澄清的,真正参与高低比较的不是内部的成分有多低,而是整个主语短语作为句子中的一个主要成分有多高。我们不妨设想,这些做主语的名词短语都是(同样按照高先低后顺序)单独形成的,成为短语之后再参与到构句过程之中。换句话说,动词 V 和宾语 DP 其实是等高的。也正是因为这个原因,人类语言才大致上会有一半前核心语言一半后核心语言(Tomlin 1986)。另一个可能来自 Kayne,即所有语言的底层结构都是前核心。所谓的后核心语言只是在句中统一地把宾语提升到限定语位置而已。

我们认为，这个思路在方向上是正确的，但是在细节上有明显缺陷，尚不足以解释高先低后。首先，并非所有的先行词都只能先行：

- (3) a. Her_k friends all love Linda_k.
b. Those rumors about themselves_k made Jack and Marcy_k very upset.
(4) 那些圈子里流行的关于彼此_k的传说让她们两人_k特别兴奋。

这类例子一直被用来讨论约束条件 (binding conditions) 的精确定义，其中每组共指关系 (以下标 k 表示) 都是语义上的先行词后置，比如 (4) 里的“彼此”回指“她们两人”。其次，句子中的任何量化成分 (quantifier) 都具有超越自身位置的量化域 (scope)。以汉语的疑问代词为例：

- (5) a. 同事问你昨天去了哪儿。
b. 同事问 [_{CP} 疑问算子 x , x =地点, 你昨天去了 x]

动词“问”的宾语从句必须是问句，而问句在其他语言中是通过在句子最高结构 (=CP) 里包含疑问算子来实现的，即 CP 是疑问代词“哪儿”的量化域。以此推断，(5a) 应该在某个造句节点上表达为 (5b)。如何得到 (5b)，在汉语学界是一个大问题。Huang (1982)，Cheng (1991)，Tsai (1994) 和 Hu (2002) 既借鉴前人又有推陈出新的不同见地。如果把争论的细节暂时搁置而接受 (5b) 为正确的表达式，那么 [疑问算子 x , x =地点] 就跟句末的 x 构成一个依存关系。问题是，这个依存关系中被听到的一方，从结构上说位于低端；而在汉语中，位于低端等同于在造句过程中出现得更晚。换句话说，在一个依存关系中，显性成分可以晚于隐性成分出现。既然如此，为什么没有一种语言在结构上和英语一样构成问句 Where did you go t yesterday，但是把其中的各种依存关系 (包括 where... t) 反过来表达，先处理隐性的 t 后说显性的 where？

Kayne (1994) 把高先低后表述为线性对应公理 (Linear Correspondence Axiom, LCA) 并试图解释其原因。但他的所谓解释依赖于一系列既没有必然性也不清楚如何验证的预设 (详见原著第4.3节)：

- (6) Kayne 所需要的假设
- 既然任何句法结构 S 都必有一个根节点 (root node, 标作 $R_{\text{合}}$) 包含其余所有的节点，那么 S 也应该有一个“根节点” $R_{\text{统}}$ 成分统治其余所有的节点。
 - $R_{\text{统}}$ 是一个“抽象”节点，其“语义内容”是某个时间点。
 - S 中的其他节点 n_i 也都各自对应于一个时间点 t_i ，具体方式是每个 t_i 对应于从句子开始到 n_i 的全部词项 (lexical items, 大致相当于造句的词)。
 - $R_{\text{统}}$ 和其他词项的线性关系在每一 t_i 都必须保持恒定。

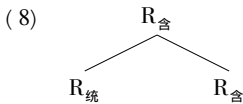
简要地说，(6a) 加上 LCA 使 $R_{\text{统}}$ 成为比实际句子中的任何成分都高都早的节点。如果句子本身是由词项 $abcd$ 组成的线性系列，那么根据 (6b-c)，从高到低的生成句子会产生 (7a) 中的词项组合 (其中 t 用来代表 $R_{\text{统}}$ 所包含的“抽象

词项”)，而从低到高产生(7b)：

(7) 时间点	t_1	t_2	t_3	t_4	
a.	ta	tab	tabc	tabcd	
b.	d	dc	dcb	dcba	dcbat

只有(7a)满足(6d)，因为每一组合中的 t 都处在所有词项的最前方；而在(7b)中， t 在绝大部分组合中都没有得到表达，更无从谈起在每一时间点 t_n 都跟其他词项保持恒定的线性关系。结论是，任何句子必须从成分统治所有成分的(即位置最高的) $R_{统}$ 开始，也就是高先低后。

我们首先注意到的是预设太多，而且其中大都缺乏独立的证据，也没有逻辑上或常识上的必然性。以(6c)为例，为什么对应于每个时间点的不是在那一刻出现的词项而必须是包括此前的全部词项？除了能够得出(7a)之外还有什么用途或者出处？即使把这层疑问暂时搁置，(6)自身也包含逻辑漏洞。该理论的核心是(6a)，即通过强调包含和成分统治在行为上的相似性，从 $R_{含}$ 类推出 $R_{统}$ 。但是类推并不能成为证据，因为表面相似的事物未必相同(比较鸟和蝙蝠)。事实上，按照 Kayne 的 LCA， $R_{含}$ 和 $R_{统}$ 不可能在行为上完全并行。论证如下(篇幅所限，无法解释全部技术细节；请读者参考原著)。(8)是 Kayne 所赋予这两个根节点的结构关系：



因为 $R_{统}$ 存在，包含性根节点 $R_{含}$ 被分离为两个“片段”(segments)。成分统治被 Kayne(1994:18) 定义为只适用于完整节点而不适用于任何片段，所以(8)其实只有两个完整节点： $R_{统}$ 和两个 $R_{含}$ 片段叠加在一起构成的 $R_{含}$ 。 $R_{统}$ 完全排除 $R_{含}$ ，即后者完全不被前者包含(Kayne 1994:16)，因而在(8)这样的结构中成分统治后者。而 $R_{含}$ 并不完全排除 $R_{统}$ (至少有一个 $R_{含}$ 的片段仍旧包含 $R_{统}$)，按照定义不能成分统治 $R_{统}$ 。由此得出，这两个完整根节点的关系是非对称成分统治关系，即单向的 $R_{统}$ 成分统治 $R_{含}$ 。再看包含关系。作为完整根节点， $R_{含}$ 并不包含 $R_{统}$ (至少一个 $R_{含}$ 片段不包含 $R_{统}$)。可见对于包含关系来说，真正包含所有成分的根本节点必须只是一个片段，不可能是完整的 $R_{含}$ ；而成分统治关系中的根节点依照定义只能是完整节点。两个关系并非在所有细节上完全具备可比性。

既然这两个关系注定不会完全并行，(6a)本身就值得商榷。 $R_{含}$ 包含所有句子成分，就一定要造出一个跟句子自身完全无关而且只能相似不能相同的抽象 $R_{统}$ 吗？如果 $R_{统}$ 和 $R_{含}$ 的类比已经不完整，要求前者成分统治其余所有节点

还有多少逻辑上的合理性？进一步说，如果 $R_{统}$ 的存在都有问题，其余预设的逻辑的正当性又如何判断？比如我们完全可以认为 $R_{统}$ 并不存在，然后把 (6c-d) 倒过来，从位置最低的词项 d 开始说起，每一时间点仍旧对应于此前全部词项，并且要求 d 和其他词项的线性关系保持一致：

- (9) 时间点 t_1 t_2 t_3 t_4
 d dc dcB dcba

(9) 在思路上仍旧沿循 Kayne (1994) 的 (6c-d)，而结果却是从低向高生成句子。总之，Kayne (1994) 突出了高先低后，必须给以肯定。但是他对这个现象的解释更像是把问题转嫁给一组既无独立证据而且自身的逻辑也不完整的假设。

2. 英语 wh-问句的三个例外情况

英语 wh-问句的基本法则很直观：wh-疑问词上行到句首，同时把助动词提前到主语和疑问词之间。如果句子自身没有助动词，则以 do 代替。这个造句法则有三个例外，而且看起来互不关联。

第一个例外：如果疑问词指主语，那么替代助动词的 do 必须不能出现：

- (10) a. Who hired his buddy?
 b. * Who did hire his buddy?

注意，(10b) 中的 did 不可重读，否则有可能被理解为表强调的 do。后者是货真价实的助动词，另当别论。

第二个例外：如果名词短语 DP 中的所有者是疑问代词，则整个 DP 必须移位：

- (11) a. Whose buddy did Jack hire?
 b. * Whose did Jack hire buddy?

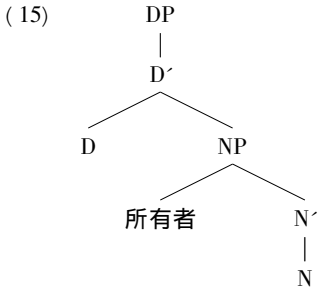
这个现象在句法中被叫做“拖带”(pied-piping)。问题是为什么拖带必须发生？必须指出，这个现象所牵涉的问题远远超出其表面的简单程度。首先，有证据表明所有者在 DP 内部是可以移动的。我们从下面几组例子说起：

- (12) 意大利语
 la mia casa, il vostro bambino
 定冠词 我的 房子 定冠词 你的 孩子

- (13) a. She is a very pretty girl.
 b. How pretty a girl she is!

- (14) a. Sarah wonders [_{CP} whose articles people are talking about]. (Kayne 1994)
 b. * Sarah wonders [_{CP} articles by whom people are talking about].

从(12)得知，所有者至少在某些前核心(head-initial)语言中位于冠词 D 和名词 N 之间。正是这类语料成为用 DP 表示名词性短语的证据，也表明所有者的初始位置应该在 NP 里边：



再看(13)。修饰名词的形容词短语 AP(=very pretty)位于不定冠词 a 和名词之间,在(15)这样的结构里也应该位于 NP 里面,按照标准分析应该附加在 N 的位置。但是当 AP 是 wh 形式时却必须出现在冠词左侧。这个现象在形式语法理论中有一个简单的解释:wh-AP 上移到 DP 的限定语位置(Spec of DP),通过限定语-核心一致性(Spec-head agreement)让 D 也携带 wh 特征。因为 D 是 DP 的核心,所以整个 DP 也是 wh,从而满足特定句式(比如疑问句和感叹句)对句首 wh 成分的要求。按照这个逻辑,(14b)不可接受是可以预期的。毕竟疑问代词 whom 是介词 by 的宾语,位置比(13a)中的 AP 还“深”,自然不可能让包含它的整个 DP(articles by whom)变成 wh,因而不能满足疑问从句对 wh 的要求。而(14a)可以接受,显然证明 whose 位于 DP 的限定语位置。但是已知所有者的初始位置在 NP 里面,可见 whose 是经过移位,从 NP 内部上移到 DP 里边的。这个结论让(11)所提出的问题更“尖锐”:如果 wh 所有者可以在 DP 里面移位,为什么不能进而直接移动到句首而偏偏要拖家带口?

有的学者认为(11)所表现出来的限制并不适用于所有语言。比如 Gavruseva(2000)把 Tzotzil 语、Chamorro 语和匈牙利语归入所有者可以单独移动到句首的类型。但是当我们仔细查看有关文献,就会发现这些所谓的例外都各自具有这样或那样的可疑之处。Gavruseva 的 Tzotzil 语料来自 Aissen(1996):

- (16) a. I-k-il-be s-tot li Xun-e.
 完成体-1-看见-间宾 3-父亲 定冠词 Xun-词缀
 ‘我看到了 Xun 的父亲。’
- b. Buch'ú av-il-be s-tot?
 谁 2-看到-间宾 3-父亲
 ‘你看到了谁的父亲?’
- c. I-y-il tz'í' li Xun-e.
 完成体-3-看见 狗 定冠词 Xun-词缀
 ‘Xun 看到了狗。’

(16a) 反映陈述句的基本语序。在(16b)的问句中,所有者被提前而宾语的核心名词 tot ‘父亲’仍旧留在初始位置,看起来相当于英语中不可接受的(11b)。

但是 Gavruseva 忽视了 Aissen 本人的注释 7，其中特别提到 (16b) 未必表明句首的所有者直接来自宾语内部，因为 (16a-b) 的动词都携带间接宾语的一致后缀 -be。与此形成对照的是 (16c)，里边的同一及物动词 il ‘看见’ 后面没有 -be。Aissen 还提到，在她早年的文章中 (Aissen 1987) 是把 (16a) 处理成所有者提升的 (possessor-ascension)。而所有者提升现象在目前尚没有广为接受的分析，连是否来自移位都有争议^② (Guéron 1985, 2006; Yoon 1990; Authier 1991; Vergnaud and Zubizarreta 1992; Kempchinsky 1992; Hole 2005; Huang, et al. 2009)。如果所有者是直接生成的另一个宾语 (见脚注^②)，那么全句就含有两个宾语，多出来的 -be 词缀也就有了简单的解释。而所谓的所有者移位，其实不过是其中一个宾语移动到句首，即使在英语里也是可接受的。

Chamorro 语料主要来自 Chung (1991)：

- (17) a. Ti man-mäguf i famagon-na si Jose.
 不 时态+复数-开心 定冠词 孩子们-一致词缀 Jose
 ‘Jose 的孩子们不高兴。’
- b. Hayi ti man-mäguf (* i) famagon-na?
 谁 不 时态+复数-开心 定冠词 孩子们-一致词缀
 ‘谁的孩子们不高兴?’

(17a) 是基本句式，(17b) 被用来证明所有者单独移位。但是正如 Chung (1991) 所指出的，(17b) 需要两个条件：和所有者相关联的名词是宾语或者非宾格动词的论元，以及该论元不可携带定冠词 i。在形式语法里，这些条件固然和从宾语中外移相匹配，但也满足其他完全不同的句法结构。比如 Massam (2001) 根据 Niuean 语所提出的假性名词合并 (pseudo-noun-incorporation)。Massam 认为动词的宾语可以在某些情况下不是 DP 而是光杆 NP。这样的结构马上就可以解释定冠词不能出现的 (17b)，因为冠词是典型的 D。从语义上说，NP 可以被看做是谓词 (这方面的文献太多，可以参考 Higginbotham (1985))，和动词直接结合后形成复式谓语 (complex predicate; 参见 Chomsky 1981)，把所谓的所有者当作自身的宾语。换句话说，(17b) 完全可能是假性名词合并，由“开心”和“孩子们”组成复式谓语，而所有者是这个复式谓语的论元。这样的论元以疑问词形式前移当然不是问题。

对匈牙利语的分析来自 Szabolcsi (1983, 1992) 并被 É. Kiss (2002) 所进一

^② 汉语中对应于所有者提升的是下类句式：

她把橘子剥了皮。

“把”通常引进宾语，在这里却和语义上的“皮”的所有者相连。如任何了解汉语语法的人所知，如何分析“橘子”的来源没有定论。比如 Huang 等 (2009) 就把“橘子”当作直接生成在宾语之外的“外宾语”。而这个分析跟 Tzotzil 语中的间接宾语标记是一致的。

步汇总。先看下面的语料:

- (18) a. (a) János könyv-e
定冠词 约翰 书-所有后缀
‘约翰的书’
- b. János-nak a könyv-e
约翰-与格 定冠词 书-所有后缀
‘约翰的书’
- c. János-nak ellopták a könyv-ét.
约翰-与格 偷-3 复数 定冠词 书-所有后缀-宾格
‘他们偷了约翰的书。’

所有者在定冠词之后不带格标记,在之前带与格。所有者和宾语分离后出现在句首也带与格, Szabolcsi 等人以此推论所有者首先上移到 DP 的限定语位置(见上文对英语的讨论),然后可进一步移出 DP 直达句首。但是 É. Kiss 还提到另一些例子:

- (19) a. az ő diák-jai-k
定冠词 他 学生-所有后缀-所有复数-3 复数
‘他们的学生们’
- b. ?* Csak (én) nekem a könyv-e-m veszett el.
只有 我-与格 定冠词 书-所有后缀-1 单数 丢失 虚词
应有义 ‘只有我的书丢了。’
- c. Nekik ellopták a labdá-j-uk-at.
他们-与格 偷-3 复数 定冠词 球-所有后缀-3 复数-宾格
‘他们的球被偷了。’

(19)和(18)的最关键区别在于用代词表达所有者时不能出现在定冠词左侧,而这个位置恰恰是 Szabolcsi 所说的所有者上移并接受与格的第一步。如果代词甚至不能出现在这个位置,那么句首的与格代词(见(19c))又来自何处?对此 Szabolcsi 没有解释,而 É. Kiss 的分析强行要求匈牙利语代词可以无格(细节见原著第168页)。我们认为,(19b)说明所有者在句首时所携带的与格另有来源,并非依赖于 DP 的限定语位置,其存在与否也跟该位置无关。也就是说,(18b-c)中的所有者都带与格,只是表面的巧合,并不说明两者同源。

总而言之,文献中见到的不同于英语(11)的情况,不但细究起来都伴随额外的特征,而且这些特征全都具备一个共性:标明句首的所有者可能另有来源,而不是从 DP 中移位的产物。这个结果基于不同语系的语言^③,不太可能是偶然的,反而进一步印证了所有者不能外移的普遍性。

第三个例外:在句子中结构位置高的英语副词都没有疑问形态。副词依照

^③ Tzotzil 属于玛雅语,Chamorro 属于南岛语,Hungarian 属于乌拉尔语(Uralic)。

语义和线性位置分类，并不是新闻^④。现有的分类接近三十类，不可能逐一讨论。本文选取有代表性的两类：评价 (evaluative) 副词和方式 (manner) 副词。下面的英语例子基于 Li 等 (2012，以下简称 LSL)：

- (20) a. Sarah luckily skillfully flew the airplane.
b. * Sarah skillfully luckily flew the airplane.
- (21) a. How skillfully did Sarah fly the airplane?
b. * How luckily did Sarah fly the airplane?

(20)表明，评价副词在句子中的位置高于方式副词(参见本文第1节关于高先低后的讨论)。如(21)所示，位置高的评价副词不能构成 wh 问句。这个现象并非偶然事件。LSL 用大量事实证明，1) 可否有 how-问句以副词类的位置决定，自频率 (frequentative) 副词以下的各类都可以，从高于频率副词类的主语指向 (subject-oriented) 副词往上都不可以；而且 2) 这个划分和语义无关：

- (22) a. How lucky of Sarah was it to fly the airplane?
b. How lucky an incident was it that Sarah flew the airplane?

评价副词 luckily 的形容词根完全可以组成 how-问句，甚至可以表达同样的意义。(22)里的两句，都可以作为(21b)的近义转述。换句话说，(21)所反映的对比，更可能是一个尚且不知道原因的句法现象。

本节所考察的这三类现象，在英语里都是疑问句规则的例外：主语疑问词拒绝使用 do，所有格疑问词必须拖带，位置高的副词甚至没有疑问形态。而三类似乎互不相关的情况，从另一个角度看却共享同一特征：

- (23) 成分移位的起点和终点不能都位于虚词组 (functional phrase) 之内。

在英语里主语是 TP 的限定语，T 是虚词。如果主语是疑问词，按照英语的问句法则要移动到 CP 的限定语位置，而 C 也是虚词。所有者位于 DP 的限定语位置，D 是虚词。从 DP 向 CP 移动同样违反 (23)。至于副词，频率类以下的副词在传统上统称为 VP 副词，现在仍旧被很多学者放在 VP/vP 之内^⑤。而比频率类位置高的副词从逻辑上推论应该在 VP/vP 之外，自然处于某个虚词短语 FP 之中，所以向 CP 进行 wh-移位也归于 (23)。现在的问题是，为什么 (23) 会存在？

因为 (23) 对移位的限制是首次被注意到，目前的句法理论尚且没有现成的答案。不仅如此，现有的理论工具也不足以推导出 (23)。以特征核查 (fea-

^④ 中文可以参考刘月华 (1983)；形式语法框架内对副词类的讨论可以追溯到 20 世纪 70 年代，近年有影响的是 Cinque (1999)。

^⑤ 从这个角度说，我们显然对 Cinque 的各种与副词相对应的虚词核心持审慎态度。另一个可能性是他的低端虚词核心其实都是动词的延伸 (Grimshaw 2000)，本质上仍旧是词汇性质的 (lexical categories)。

ture-checking) 为例, 所有者上行到 DP 的限定语位置或者主语上行到 TP 的限定语位置, 都是为了和 D 或者 T 建立临近关系。其推动力可以是格, 也可以是某种 EPP 特征。即使所有者是为了就 wh 性质和 D 达成一致(见上文关于(12-15)的讨论), 疑问词自身仍旧有待处理, 否则整个 DP 就不必被拖带到句首了。可见无论疑问词进入 DP 或 TP 的原因是什么, 并不能满足最终把它移入 CP 的要求。在这个位置上, 疑问词所面临的仍旧是“造句尚未成功, 自身仍需努力”。而(23)却表明, 这种努力恰恰在这个环境中会造成各种例外。我们不仅要问为什么(23)会存在, 而且还要问为什么主语、所有者和副词的行为同处于(23)却以完全不同的方式表现出来(比如为什么所有者不直接移至句首但拒绝使用 do)?

3. 热力学第二定律简介及应用

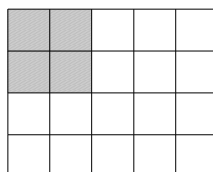
第二定律有多种表达方式, 本文使用 Boltzmann 对熵(entropy)的表述:

$$(24) S = k \ln \Omega$$

(24)中 S 代表熵, 定量一个热力学系统 R 内部(比如一间屋子)的能量(=热)的分布状况。 k 是 Boltzmann 常数, 在本文中看做 1。 Ω 可以理解为 R 中携带能量的分子在给定状态下存在多少种可能的组合方式。用一个常识性的例子来说明(24)的意义: 想象一个寒冷的屋子里放进一个烧得通红的铁块。随着时间推移, 屋子的整体温度会升高而铁块的温度会下降, 直到两者的温度等同。这背后的原理是, 携带能量的分子(姑且叫做热分子)在运动中碰撞到不携带能量的分子(冷分子)并把能量传递给后者, 而后者又会在运动中碰撞到其他冷分子。这个自然过程表现为热(即能量)在空间的传递, 也叫做能量的耗散(energy dissipation)。

能量耗散是一个自然过程, 因为热分子的随机运动不可避免, 与周边冷分子的碰撞也就不可避免。这个过程还可以从另一个角度来描写。当热铁块刚刚放进冷屋子的时候, 能量的分布表现为一种“秩序”或者说“结构”, 即有的地方集中的冷, 有的地方集中的热。能量耗散的最终结果, 是使携带能量的分子随机分布在整个屋子中, 从而使原来的系统达到一个新的状态, 其无序(chaos)程度上升, 即熵值增加。下面看一个量化的例子。

设“屋子”等于一个包含 $4 \times 5 = 20$ 个分子的两维空间, “热铁块”是其中的 $2 \times 2 = 4$ 个热分子, 屋子里的其余 16 个分子是冷的:



在这个初始状态，我们在屋子里摆放热铁块（即把 2×2 个热分子作为一个整体来安排）的方式只有 12 种（4 个横移位置乘以三个纵移位置），相当于 (24) 中的 $\Omega = 12$ ， $S = \ln 12 = 2.485$ 。经过一定时间之后，携带能量的热分子在自然耗散的作用下被随机地分布在空间，组合^⑥数目 $\Omega = 20!/4!(20-4)! = 20 \times 19 \times 18 \times 17/4! = 4845$ ， $S = \ln 4845 = 8.486$ 。换一个说法，在同一空间中的四个热分子充分耗散会有接近 5 千种可能的分布，其中碰巧以 2×2 为一个整体的分布只有 12 种可能。在自然条件下，后一类情况出现的概率小于全部可能性的 $1/400$ ，即便我们设定的热力学系统是如此之小。总之，能量的分布从有序向无序转换是“自动”的，在没有其他因素介入的情形下无序（熵值高的）分布的概率远远高于有序（熵值低的）分布。正因如此，暖气片虽然安装在一面墙上却可以温暖整个房间。

下面我们从能量耗散的角度考虑语言。造句需要能量，应该无需证明。关于第一语言习得方面的研究表明，幼儿使用实词的时间早于虚词 (Poeppe and Wexler 1993; Rizzi 1994; Haegeman 1995; Hyams 1996; Hoekstra and Hyams 1998)。从事语言进化研究的学者也基于各种原因而猜想，人类原始语言 (proto-language) 所使用的词汇中只包含相当于实词的“词项” (Bickerton 1990; Jackendoff 2002)。由此推断，涉及实词的神经活动对人类（以及其他物种）来说是更基础的。我们因此提出下述假设：

(25) 涉及实词的语法操作所需要的能量少于涉及虚词语法操作的能量。

此处所关心的毕竟是大脑的神经活动，所以 (25) 应该被理解为激活由 n 个语言运算单位集中组成的群体（比如某个皮层区域）所需要的能量。把 (25) 和句法中关于移位的拷贝理论 (2) 加在一起得出以下定理 (theorem)：

(26) 运算单位不等定理

设任何语法操作必须用能量激活包含 n 个语言运算单位 cu 的群体，则处理同一语法成分在实词环境中所需要的 cu 少于在虚词环境中所需要的 cu 。

具体来说，移位的疑问词和它的拷贝属于同一语法成分。在英语中，前者是 CP 的限定语，位于虚词环境中。而后者是移位的起点，可以在虚词环境中（比如主语在 TP 里）也可以在实词环境中（比如宾语在 VP 里）。我们先考虑后一种情况：

(27) [虚词环境 X ... [实词环境 ... X-拷贝 ...]]

可以证明，从 X 向 X-拷贝的造句顺序对能量的消耗永远小于从 X-拷贝向 X 的

^⑥ 是组合而不是排列，因为分子 3、9、11、18 为热分子和 9、11、3、18 为热分子，对于整个屋子的最终“温度”而言没有任何区别，对 W 来说是同一个分布态。注意，运算结果中也包含前面提到的 12 种分布。

顺序。

令 N_1 为处理 X 所需要激活的 cu 群的大小(即群体中的 cu 数量;下文同), N_2 为处理 X-拷贝所需要的 cu 群的大小, t_1 和 t_2 分别为处理 N_1 和 N_2 的时间点。从(26-27)得出:

- (28) a. $N_1 > N_2$
b. $|t_1 - t_2| = \Delta t > 0$

如果 t_1 小于(=早于) t_2 则先处理 X, t_2 小于 t_1 则先处理 X-拷贝。因为熵值自动增加(即能量自然耗散), cu 群中的 cu 数量在 Δt 之后肯定会减少^⑦。本节假设耗散率 r 为一个 $0 < r < 1$ 的相对于 cu 初始数量的常数(比如在 Δt 之后由于耗散而剩下 70%)。在证明开始之前,先用文字描述一下基本思路和结果:

如果从 X 开始生成句子,那么最初激活的 N_1 个 cu 在 Δt 之后仍旧处于激活状态的数量是 $N_1 \cdot r$; 如果这个数量不小于 N_2 , 则无需激活新的 cu 就可以处理 X-拷贝。这等于说一次激活 N_1 个 cu 就足以处理(27)中的共指关系。

设从 X-拷贝开始生成句子。 N_2 在 Δt 之后因为能量耗散剩下 $N_2 \cdot r$; 为了处理 X 所要激活的 cu 总量可分为两部分: 首先是 $N_1 - N_2$, 即弥补两个基本量本来的差别(见(28a)); 其次是 $N_2 - N_2 \cdot r$, 用来补足耗散所失去的数量。第一部分加上最初的 N_2 等于 N_1 , 再加上第二部分必定多于 N_1 。

可见由于能量耗散, 从高向低生成(27)比反向生成所消耗的能量更少。如果句法操作如 Chomsky(1995)所设想的那样永远偏爱低消耗运算, 那么高先低后就因为热力学第二定律而成为唯一的选择, 所以自然没有例外。下面是完整的证明。

由于耗散率 r 大于 0 小于 1, 所以 $N_1 > N_1 \cdot r > 0$, 且下述三种情况中必有一种成立:

- (29) a. $N_1 \cdot r > N_2$
b. $N_1 \cdot r = N_2$
c. $N_1 \cdot r < N_2$

对于前两种情况, 上文已经证明从 X 到 X-拷贝只需一次激活 N_1 个 cu 就足以处理(27), 而反向生成还要加上额外的 $N_2 - N_2 \cdot r$ 。 r 越大, 要补充激活的 cu 就越多。

如果是(29c), 则可以证明由高向低和由低向高所需要的 cu 数量分别为(30a-b):

- (30) a. 高 \rightarrow 低

^⑦ 这里的 Δt 最终需要更精确的定义, 是大脑处理 X 和 X-拷贝的时间间隔, 还是在实际造句中两个成分出现的时间间隔。但是这个区别对阐述我们的理论不产生影响。

$$N_1 + (N_2 - N_1 \cdot r)$$

b. 低 → 高

$$N_1 + (N_2 - N_2 \cdot r)$$

而且 (30b) 必定大于 (30a) :

$$(31) N_1 > N_2$$

$$\therefore N_1 \cdot r > N_2 \cdot r \quad (r > 0)$$

$$\therefore (N_2 - N_1 \cdot r) < (N_2 - N_2 \cdot r)$$

可见只要 (30) 成立, 则高先低后仍旧是消耗能量更少的选择。

对 (30a) 的证明: N_2 是处理 X-拷贝所需的 cu 数量, $N_1 \cdot r$ 是 Δt 之后仍旧处于激活态的 cu 且小于 N_2 (即 (29c))。因此两个数量之差是处理 X-拷贝所需要补足的 cu 的最小数量。也就是说, 如果 N_1 在 Δt 之后已经耗散到不足以处理 X-拷贝, 则需要补充激活的 cu 是 $N_2 - N_1 \cdot r$, 即 (30a)。再看 (30b)。首先需要激活 N_2 。在 Δt 之后剩下 $N_2 \cdot r$ 。为处理 X 显然必须补充激活更多的 cu, 其数量包含两部分:

$$(32) \text{ a. } N_1 - N_2$$

$$\text{ b. } N_2 - N_2 \cdot r$$

第一部分是两个基本量本身的差别, 第二部分是初始的 N_2 耗散掉的数量。两部分加在一起才足以补足处理 X 的 cu。因此前后激活的 cu 总量为 (33), 即 (30b) :

$$(33) N_2 + (N_1 - N_2) + (N_2 - N_2 \cdot r)$$

$$= N_1 + (N_2 - N_2 \cdot r)$$

有兴趣的读者不妨带入 (显然是虚拟的) 具体数量 (比如 $N_1 = 100$, $N_2 = 70$, $r = 0.6$) 来验证两个生成方向的能量消耗情况。

和 Kayne 对句子生成方向的理论相比, 这个建立在热力学第二定律之上的解释需要三条假设, 即实词比虚词需要的能量少、语法运算永远选择“节能”之路、以及耗散率 r 是一个 $0 < r < 1$ 的常量。第一个假设并非完全空穴来风 (参考了语言习得过程) 而且至少原则上是可以验证的。第二个假设已经贯穿于目前的句法理论 (见最简方案理论)。我们在后面将进一步考虑耗散率这个因素。

4. 关于 wh-移位

把移位的起点和终点跟实词虚词环境结合, 存在四种逻辑可能:

(34)	终点环境	起点环境
a.	虚词	实词
b.	虚词	虚词
c.	实词	虚词
d.	实词	实词

其中(34a)是上一节的重点。(34c)在目前的句法理论中,尤其是通过 Chomsky (2000, 2001)的探索与目标 (probe-goal) 理论部分回归到更加传统的句法操作之后,可以从根本上排除。比如移位的动机是为了就近处理某些特征,而已知的能影响句法操作的特征从本质上都是功能性的,也就是都跟虚词休戚相关。从这个角度上说,从虚词环境到实词环境的移位根本就不会发生。按照这个逻辑,(34d)也不可以。至少当移位成分是短语时,这个预测是正确的。最明显的例外是核心成分移位。始自 20 世纪 60-70 年代的生成语义学,进而得益于 Baker(1988)试图把词法纳入句法的理论,从名词向动词和从动词向动词的移位被广泛用来解释各种语言现象。但是如 Li(2005)指出,目前对核心词移位的使用不但有失控之嫌而且在一些关键语料方面也不能自圆其说,至少有很多问题尚待澄清,本文暂不涉及(但是见下文关于核心词移位距离的讨论)。在这一节中,我们集中考察(34b)。

如果自然语言在处理移位时如此介意对能量的有效使用,那么(34b)显然不属于理想情况。事实上,这样的移位注定不能如(34a)那样一箭双雕,用一次激活的 cu 处理移位的全过程:由于两头都是虚词环境,无论哪一个生成方向都在能量耗散的作用下注定需要补充激活更多的 cu。这不仅仅是低效,而且也让语言机制无法从两个选择中加以取舍。目前的研究表明,整个语言机制从根本上是建立在语言成分的非对称关系之上的(有关讨论见李亚非(2014))。但(34b)恰恰在最根本的能量消耗层面对称的。

如果上述逻辑成立,那么第 2 节中介绍的三种英语 wh-移位的异常情况就有了一个统一的解释:(23)等于(34b),三种异常的根本原因在于正常的 wh-移位在这里不能有效地使用能量。有意思的是,它们在细节上的不同可以从现有的理论中找到解释。

所有者虽然不能从 DP 内部移动到 CP 内部,却可以通过限定语-核心词一致性,使 D 携带 wh 特征,从而使整个 DP(比如 whose buddy)从语法角度等同于一个 wh 短语。而这个短语作为宾语,从 VP 环境中向 CP 移动属于(34a),是被容许的。这是“拖带”的根本原因,在第 2 节中已经详细论述。

句子的主语没有这个便利条件,因为在 TP 短语里,限定语(=主语)和核心词(=T)之间的一致是建立在所谓的 ϕ -特征(性、数、人称)之上的,和 wh 没有关系。既然不能通过一致性操作让 TP 携带 wh-特征,拖带也就无从谈起。这条路走不通,主语只好停在原地不动。而助动词 do 的使用,是英语继承了西日耳曼语的典型特点,用一个短语成分做 CP 的限定语,同时把就近的动词性成分放在 C 的位置,两项操作同时进行。如果主语不做 CP 的限定语,就没有必要在 C 放置任何成分,自然无需启用助动词 do。

位置高的副词没有 *wh* 形式，根本原因也是因为它们的初始位置是虚词环境。一个副词类是否有疑问形态，界限在主语指向类和频率类之间，从后者往下可以但从前者往上不可以 (LSL)。这个界限基本上对应于 VP 以内 (即传统上的 VP-adverbs) 和 VP 以外。虽然不同的学者关于这个区间的各种短语的理论、数量和名称不尽相同，但是基本的划分是一致的，即句子的核心部分是动词性的实词短语，而外部是各种功能性虚词短语。从这个角度上说，主语指向类往上，可以比较放心地被认为都位于虚词环境中，如果移位就会造成 (34b) 的局面。由此产生一个新问题：这些副词不能上移到 CP 里面，从道理上和主语不能移动是一样的，为什么后者可以通过原地不动来表达疑问 (见 (10a)) 而前者不行呢？或者说为什么不好的 (21b) 不能采用 (35) 的方法加以“挽救”？

(35) * Sarah how luckily flew the airplane?

为了回答这个问题，我们先看两个汉语的例子：

(36) a. 他 [在邻居修理什么的时候] 表达了不满？

b. 他 [在邻居为什么修理围墙的时候] 表达了不满？

汉语是疑问词不发生表面移位的语言。因为没有移位，所以“什么”在 (36a) 中对状语从句构成的岛不敏感。在作者本人的方言里，(36b) 的“为什么”有两种读法：重音在“什么”，答案可以是“…为面子…”；或者重音在“为”，直接对应于英语中的 *why*。(36b) 按照第一种读法可以接受，按照第二种读法不可以。自从 Huang (1982) 首次注意到这类对比之后^⑧，学者们提出过各种分析，有代表性的包括 Huang (1982)、Tsai (1994)、Reinhart (1997)、Hu (2002) (也见 Cheng (2003) 的总结和评估)。我们所关心的是 (36) 所代表的事实：论元的疑问形式可以直接出现在岛内，而同样的措施并不适用于某些状语，比如表 *why* 意义的“为什么”。回到英语，位于虚词环境里的疑问词不能再移动出来，从描写意义上说，它们所存在的虚词环境就是一个阻滞移位的“虚词岛”。虚词岛内部的疑问词论元 *who/what* 可以在原地表达，而状语则完全没有疑问形态，就像表达 *why* 意义的“为什么”不能出现在状语岛中一样。虽然其中的原因目前还不完全清楚^⑨，但是二者在描写和归纳层次上是一致的。

总之，起点和终点都在虚词环境的移位不可接受。可以通过拖带来挽救的是所有者；不能用到拖带的就退而求其次，采用原地发问的是主语；连原地发问都不被允许的副词只好连问都不能问，无论语义上是否说得通 (见 (22))。这里所看到的，是英语在语言机制所允许的范围内，穷尽各种手段完成疑问句

⑧ 最初的分界是论元和非论元，但事实证明汉语除了作为 *why* 来理解的“为什么”，其余所有的疑问词都可以出现在岛中。另外，使用 *wh*-岛的语料已经被 Liu (1986)、Xu (1990)、Hu (2002) 证明更有可能是回声问句 (echo question)，因而跟这里关注的现象无关。

⑨ 读者可以参考 Reinhart (1997)。

的交际功能，除非实在无路可走。

虽然上面这三种异常都是具体的语言现象，但是本文的解释却是建立在能量耗散和追求高效运算等基本因素之上的，因此其应用范围也不应该局限在一种语言或者一类现象之上。事实的确如此。比如 Lasnik 和 Saito (1992:110) 就注意到英语的话题化不会发生在主语身上。话题化在形式语法中是把有关成分移动到 TP 上边的某个位置。而这样的位置无论按照什么细节的句法分析，都肯定处于虚词环境之中。按照本文的思路，宾语从实词性的 VP 中话题化属于 (34a)，而主语的话题化则属于不可接受的 (34b)。另一个有趣的现象来自非洲的 Vata 语 (Rizzi 1990)：

- (37) a. àló * (ò) le saká la?
谁 他 吃 米饭 疑问标记
‘谁吃了米饭?’
- b. Yi Kòfí le (* mí) la?
什么 Kofi 吃 它 疑问标记
‘Kofi 吃了什么?’

这组例子的关键是所谓的恢复代词 (resumptive pronoun) 的使用。恢复代词的主要用途，是在移位造成某种语法问题或者理解困难时出现在本该发生移位的起点位置，比如汉语中的“你在大家都批评 (他) 的时候出来说话的那个同事”，如果本意是那个同事受到大家批评，在“批评”后面加进“他”反倒有助于理解。这个代词就是恢复代词。在 Vata 语里，当主语是疑问词时，恢复代词 ò ‘他’必须出现，而宾语疑问词却拒绝恢复代词 mí (相当于英语的 it)。Vata 的疑问句显然类似英语，疑问词上移的终点也应该是 CP。宾语上移可以，但主语上移则产生问题。我们已经看到英语对此的选择是让主语停留在 TP 里面。Vata 采用了语言机制所提供的另一个解决方式：把主语疑问词直接放在 CP 的限定语位置，然后在主语位置放入一个共指的恢复代词。整个句子没有移位，也就不需要满足来自移位的“节能”要求。

以上的例子形形色色，其实都是为了避开 (34b)。一个可能的反例是连续性循环移位 (successive cyclic movement)：

- (38) a. What did Jack think [_{CP} __ that Jill bought t]?
b. * What did Jack wonder [_{CP} why Jill bought t]?
- (39) It seems [_{CP} that Jill was believed [_{CP/TP} t to like tadpoles]].

为了解释 (38) 中的反差，疑问成分被认为需要经过最近的 CP 限定语位置才能走得更远。(38b) 中的间接问句已经有 why 做 CP 的限定语，所以初始位置在 t 的 what 被阻断了移位的路径。相对而言，(38a) 的宾语从句不是疑问句，that 之前的 CP 限定语位置空着 (标作 “__”)，允许 what 经由此地继续上行。但是

“_”是位于虚词环境里，而 what 的最终目的地是下一个 CP 的限定语位置。这恰恰是 (34b) 所禁止的。同样，(39) 是典型的提升结构 (raising)，不定式从句的主语 Jill 上移到临近的 (有主格的) 主语位置。然而这样移位的起点和终点都是 TP 限定语，是从虚词环境到虚词环境，应该不被接受。

这类“反例”有一个共同之处：移位的两端分散在不同句子里。我们认为这并非偶然。每个句子 (clause) 都是对我们所感知的一个独立事件的完整表达 (包括事件自身的内容和讲话者的出发点 (= aspect 和 tense) 及态度 (= modality))。因此可以想象，每个句子也伴随着新一轮的对 cu 的激活。按照这个假设，本文提出的造句方向理论应该以句子为计算界限，因为下一个句子需要激活新的 cu 群落，能量的消耗也重新算起。显然，这个假设尽管目前只能停留在“说得通”的阶段，但是具备实验层面的可证伪性 (falsifiability)，所以本质上不同于 Kayne 的 (6)。

另一类反例来自核心词移位。从 V 到 T 属于实词向虚词移位，不是问题。Baker (1988) 对动词合并 (verb-incorporation) 的分析，至少其最关键的部分也不是问题，因为动词合并的底层结构都是“双句” (bi-clausal)。虽然移位是从动词到动词，起点终点都是实词，却跨越两个句子，能量的耗费应该分开计算。真正构成问题的是在同一个句子中，从 V 向 v 的移动和从 T 向 C 的移动。前者的起始点都是实词性位置^⑩，后者都是虚词性位置。但是有一个现象值得注意：核心词移位和短语移位明显不同，其起点和终点必须在结构上紧密邻接 (Travis 1984)，也就是说距离近到不能再近。一个可能的解释是，距离近等于说 Δt 极小，因而首次激发的 cu 不会大量耗散，即使在 [虚词……虚词] 和 [实词……实词] 的情况下仍旧可以最大限度地降低需要补充激发的 cu 的数量。此外，虽然从 T 到 C 的核心词移位被广泛采用 (比如处理日耳曼语里的动词次位 (verb-second) 现象)，但是同时还有一种观点，认为 T 和 C 从根本上共享时态 (语料和理论动机详见 Chomsky (2004))。果真如此，那么把助动词从 T 上移到 C 反倒是多此一举了，如果需要时直接放到 C 上反倒更直接更符合运算经济性的要求。总之，核心词移位远非表面看起来那么不证自明，牵扯的问题不是一篇文章所能够解决的。我们只能在目前的理论框架内提出一些解决思路之后将其暂时搁置。

5. 关于能量耗散率

第 3 节的证明里假设被激活的 cu 依照特定的比率 r 发生能量耗散，且

^⑩ v 是实词而非虚词有多重原因。首先 v 的作用是提供施事论元 (Chomsky 1995; Kratzer 1996; 陈哲、李亚非 2016)，而在句法中，其他所有带论元的词项都是实词。另一个原因出于 Li (1990)。Li 注意到动词合并时的动词根永远不可能携带任何功能词缀，说明两者之间没有虚词短语。由此推论，v 只能是实词。详见原作。

$0 < r < 1$ 。这个假设里有两个细节并非不证自明。

第一，从逻辑上说 $r=0$ 是可能的，即首次激活的 cu 在 Δt 之后已经全部被耗散掉。在这种情况下，无论是从上至下还是从下往上生成句子，最后的结果都一样，即在总体上需要 N_1+N_2 个 cu 才能够处理 $X \cdots X$ -拷贝这个依赖关系。说到底， r 是否为零不是一个思辨性质的问题，而是需要并且迟早能够通过实验来确定的。我们在此只提出一个设想：语言之所以容许句法成分移位，正是因为在一个句子的范围内，大脑中运算单元 cu 的能量自然耗散率 r 大于 0。或者说，假使 $r=0$ ，那么现有的这种移位形式在人类语言中根本就不会存在，取而代之的将是其他方式（比如必须使用恢复代词）。

第二， r 可能不是一个相对于初始激活量的比率而是一个常数 k ，比如无论首次激活多少 cu ，每毫秒都有 5 个 cu 由于能量耗散而失去“活性”。如果 k 使所有被激活的 cu 在 Δt 之后耗散殆尽，那么我们回到上面的第一点，这里不再重复。排除这个情况，我们考虑耗散常数 k 对句子生成方向的影响。

已知 $N_1 > N_2$ ，必有以下三种情况（其中 $k \cdot \Delta t < N_1$ ）：

(40) a. $N_1 - k \cdot \Delta t > N_2$ (参见(29))

b. $N_1 - k \cdot \Delta t = N_2$

c. $N_1 - k \cdot \Delta t < N_2$

从基本逻辑上说， $k \cdot \Delta t$ 作为一个常量（在任何给定的 $X \cdots X$ -拷贝关系中， Δt 是已知的）相当于 N_2 比 N_1 在单位时间内耗散得更快。举例说明：令 $N_1 = 100$ ， $N_2 = 60$ 。如果按照耗散比率 $r = 50\%$ ，则 N_1 剩下 50， N_2 剩下 30。而按照耗散常数，使 $k \cdot \Delta t = 50cu$ 的话， N_1 仍旧剩下 50， N_2 却只剩下 10。也就是说如果耗散速度是一个常数 k ，那么由低向高生成句子所需要的 cu 总量比用 r 来计算会更大。所以第 3 节证明的最后结论至少对于 (40a-b) 来说不但仍旧成立，而且是以更显著的效果成立。关于 (40c) 的证明如下。

设先处理 X ，需要激活的 cu 总量是：

$$(41) \quad N_1 + (N_2 - (N_1 - k \cdot \Delta t)) \\ = N_2 + k \cdot \Delta t$$

其中 $(N_1 - k \cdot \Delta t)$ 是 Δt 之后仍旧处于激活状态的 cu 。这个数量和 N_2 之差就是处理后者所需要补足的 cu ，再加上最初的 N_1 得出 cu 的总量。如果先处理 X -拷贝， Δt 之后需要补充激活的 cu 是以下两项之和：

(42) a. $N_1 - N_2$

b. $N_2 - (N_2 - k \cdot \Delta t)$

(42a) 是两个基本量的差。(42b) 里的 $(N_2 - k \cdot \Delta t)$ 为耗散后的剩余 cu ，其与 N_2 的差是需要先补足 N_2 的数量（作为 (42a) 的基础）。因此这个生成方向的 cu 总

量为初始量加(42):

$$(43) N_2 + (N_1 - N_2) + (N_2 - (N_2 - k \cdot \Delta t))$$

由于我们设定 $k \cdot \Delta t < N_1$, 所以 $(N_2 - k \cdot \Delta t)$ 可以大于零, 也可能等于零(不考虑小于零的情况, 因为没有处于激活态的 cu 数量不会是负数)。如果大于零, (43) 化简如下:

$$(44) (43) = N_1 + (N_2 - N_2 + k \cdot \Delta t) = N_1 + k \cdot \Delta t$$

已知 $N_1 > N_2$, 所以(44) 大于(41), 即从 X-拷贝到 X 更耗费能量。如果 $(N_2 - k \cdot \Delta t)$ 等于零, 则:

$$(45) (43) = N_2 + (N_1 - N_2) + (N_2 - 0) = N_1 + N_2$$

鉴于 $k \cdot \Delta t < N_1$, 所以 $(N_1 + N_2) > (N_2 + k \cdot \Delta t)$, 即(45) 同样大于(41)。

6. 结语

本文试图从能量自然耗散的角度来解释两类现象。一是(按照树形图表达习惯)所有的语言都遵循高先低后的生成循序, 二是 wh-移位的起点终点不能都在虚词环境之中。如第 2、4 节所示, 这个限制同样适用于其他移位(比如话题化)和其他语言(比如 Vata, Tzotzil, Chamorro, Hungarian)。同时, 该理论也和广泛使用的核心词移位发生部分冲突。因为核心词移位自身仍有待进一步澄清, 目前尚不清楚冲突的真正原因。

References [引用文献]

- Aissen, Judith. 1987. *Tzotzil Clause Structure*. Dordrecht: Reidel.
———. 1996. Pied-piping, abstract agreement, and functional projections in Tzotzil. *Natural Language and Linguistic Theory* 14, 447-91.
Authier, J.-Marc. 1991. Null subjects in the DP and inalienable construal in French. In Aaron L. Halpern, ed., *The Proceedings of the 9th WCCFL*. Center for the Study of Language and Information, Stanford, CA. Pp.15-28.
Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago, IL: Chicago University Press.
———. 2002. Building and merging, not checking. *Linguistic Inquiry* 33, 321-8.
Bickerton, Derek. 1990. *Language and Species*. Chicago, IL: University of Chicago Press.
Chen, Zhe (陈哲) and Yafei Li (李亚非). 2016. Looking at vP from the Chinese perspective. *Foreign Language Teaching and Research* 4:483-97. [2016, 从汉语的角度看 vP. 《外语教学与研究》第 4 期, 483-97 页。]
Cheng, L.-S. Lisa (郑礼珊). 1991. On the typology of wh-questions. Ph.D. diss., MIT, Cambridge, MA.
———. 2003. Wh-in-situ. *GLoT International* 7, 103-9 and 129-37.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
———. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.
———. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels and Juan Uriagereka, eds., *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, MA: The MIT Press. Pp.89-155.

- . 2001. Derivation by phase. In Michael Kenstowicz, ed., *Ken Hale: A Life in Language*. Cambridge, MA: The MIT Press. Pp.1–52.
- . 2004. Beyond explanatory adequacy. In Adrianna Belletti, ed., *The Cartography of Syntactic Structures: Structures and Beyond*. Oxford: Oxford University Press. Pp.104–31.
- Chung, Sandra. 1991. Functional heads and proper government in Chamorro. *Lingua* 85 85–134.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective*. Oxford: Oxford University Press.
- É. Kiss, Catalin. 2002. *The Syntax of Hungarian*. Cambridge: Cambridge University Press.
- Gavruseva, Elena. 2000. On the syntax of possessor extraction. *Lingua* 110 743–72.
- Grimshaw, Jane. 2000. Extended projection and locality. In Peter Coopmans, Martin Everaert and Jane Grimshaw, eds., *Lexical Specification and Insertion*. Amsterdam: John Benjamins. Pp.115–33.
- Guéron, Jacqueline. 1985. Inalienable possession, PRO-inclusion and lexical chains. In Jacqueline Guéron, Hans G. Obenauer and Jean-Yves Pollock, eds., *Grammatical Representation*. Dordrecht: Foris. Pp.43–86.
- . 2006. Inalienable possession. In Martin Everaert and Henk van Riemsdijk, eds., *The Blackwell Companion to Syntax*. Malden, MA: Blackwell. Chapter 35 and Pp.589–638.
- Haegeman, Liliane. 1995. Root infinitives, tense and truncated structures. *Language Acquisition* 4, 205–55.
- Higginbotham, James. 1985. On semantics. *Linguistic Inquiry* 16 547–93.
- Hoekstra, Teun and Nina Hyams. 1998. Agreement and finiteness of V2: Evidence from child language. *Proceedings of BUCLD 22* 240–52.
- Hole, Daniel. 2005. Reconciling “possessor” datives and “beneficiary” datives – towards a unified voice account of dative binding in German. In C. Maienborn and A. Wöllstein, eds., *Event Arguments: Foundations and Applications*. Tübingen: Niemeyer. Pp.213–42.
- Hu, Jianhua (胡建华). 2002. Prominence and locality in grammar: The syntax and semantics of *wh*-questions and reflexives. Ph.D. diss., City University of Hong Kong.
- Huang, C.-T. James (黄正德). 1982. Logical relations in Chinese and the theory of grammar. Ph.D. diss., MIT, Cambridge, MA.
- Huang, C.-T. James, Y.-H. Audrey Li (李艳惠) and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge: Cambridge University Press.
- Hyams, Nina. 1996. The underspecification of functional categories in early grammar. In Harald Clahsen, ed., *Generative Perspectives on Language Acquisition*. Amsterdam: John Benjamins. Pp. 91–128.
- Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Julien, Marit. 2000. Syntactic heads and word formation. Ph.D. diss., University of Tromsø, Norway.
- Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.
- Kempchinsky, Paula. 1992. The Spanish possessive dative construction: θ -role assignment and proper government. In Paul Hirschbühler and E. F. Konrad Koerner, eds., *Romance Languages and Modern Linguistic Theory*. Amsterdam: John Benjamins. Pp.135–49.
- Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck and Laurie Zaring, eds., *Phrase Structure and the Lexicon*. Dordrecht: Kluwer. Pp.109–37.
- Lasnik, Howard and Mamoru Saito. 1992. *Move-alpha: Conditions on Its Application and Output*. Cambridge, MA: The MIT Press.
- Li, Yafei. 1990. X^0 -binding and verb incorporation. *Linguistic Inquiry* 21 399–426.
- . 2005. *X^0 : A Theory of the Morphology-Syntax Interface*. Cambridge, MA: The MIT Press.
- . 2014. Formal syntax, iconicity, and studies of Chinese. *Studies of the Chinese Language* 6: 521–30. [2014, 形式句法、象似性与汉语研究 《中国语文》第6期, 521–30页。]
- Li, Yafei, Rebecca Shields, and Vivian Lin. 2012. Adverb classes and the nature of minimality. *Natural Language and Linguistic Theory* 30 217–60.

- Liu, Feng-his (刘凤樾). 1986. On topic-traces in Chinese. In Mary Dalrymple, Jeffrey Goldberg, Kristin Hanson, Michael Inman, Christopher Piñon, and Stephen Wechsler, eds., *Proceedings of the 5th WCCFL*. Stanford, CA: CSLI Publications. Pp.142-53.
- Liu, Yuehua (刘月华). 1983. Classification of adverbials and the order of multiple adverbials. *Research and Explorations in Grammar 1*. Beijing: The Commercial Press. Pp.32-56. [状语的分类和多项状语的顺序。见《中国语文》杂志社编,《语法研究和探索(一)》。北京:商务印书馆。32-56页。]
- Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19, 153-97.
- Poeppl, David and Kenneth Wexler. 1993. The full competence hypothesis of clause structure in early German. *Language* 69, 1-33.
- Reinhart, Tanya. 1997. Quantificational scope: How labor is divided between QR and choice-functions. *Linguistics and Philosophy* 20, 335-97.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: The MIT Press.
- . 1994. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3:371-93.
- , ed. 2004. *The Structure of CP and IP: The Cartography of Syntactic Structures*, Vol. 2. New York, NY: Oxford University Press.
- Szabolcsi, Anna. 1983. The possessor that ran away from home. *The Linguistic Review* 3, 89-102.
- . 1992. Subject suppression or lexical PRO? The case of derived nominals in Hungarian. *Lingua* 86, 149-76.
- Tomlin, Russell. 1986. *Basic Word Order: Functional Principles*. London: Groom Helm.
- Travis, Lisa. 1984. Parameters and effects of word order variation. Ph.D. diss., MIT, Cambridge, MA.
- Tsai, Wei-tien (蔡维天). 1994. On economizing A-bar dependencies. Ph.D. diss., MIT, Cambridge, MA.
- Vergnaud, Jean-Roger and Maria L. Zubizarreta. 1992. The definite determiner and the inalienable constructions in French and in English. *Linguistic Inquiry* 23, 595-652.
- Xu, Liejiong (徐烈炯). 1990. Remarks on LF movement in Chinese questions. *Linguistics* 28, 355-82.
- Yoon, James. 1990. Theta theory and the grammar of inalienable possession constructions. In *Proceedings of NELS 20*. GLSA, Amherst, MA. Pp.502-16.

作者简介

李亚非,男,博士,美国威斯康星大学语言学系教授。研究兴趣:句法、词法-句法界面、语言学方法论。代表作: X^0 : *A Theory of the Morphology-Syntax Interface*, *The Syntax of Chinese*(合著)和 *Structural head and aspectuality*。电子邮件: yafeili@wisc.edu

LI Yafei, male, Ph.D., is a professor of linguistics at the University of Wisconsin-Madison. His research interest includes syntax, morphology-syntax interface and research methodology. His representative publications are: X^0 : *A Theory of the Morphology-Syntax Interface*, *The Syntax of Chinese* (co-author), and "Structural head and aspectuality". E-mail: address: yafeili@wisc.edu

作者单位及通信地址: 南京大学文学院/美国威斯康星大学语言学系

Affiliation and Contact:

School of Liberal Arts, Nanjing University

Department of Linguistics, University of Wisconsin-Madison

1158 Van Hise Hall, University of Wisconsin, Madison, WI 53706, USA