

说过去就没了

同样一门语言，三岁孩子不用人教就能用起来，写下来却能复杂到没人听得懂。四种成熟解释各有实验撑着，却互相取消——输入究竟是穷是富、重点在知识还是在时间、因果箭头指向听的人还是说的人。本文认为四家漏掉了同一件事：一段话说过去就没了，你没法请它重来。

王德生 + Claude · 约 1.7 万字 · 18 页 · 三种阅读方式 · 发表于2026年8月1日

有一种句子在纸上很常见：把主语说出来，插进去一整段限定，再回来给它谓语。跨七种欧洲语言的语料清点显示，这种「中间套」在书面语里最深只到三层，而在口语里连两层都罕见到近乎不存在；更要命的是，它是随着书面语的出现才被造出来的。与此同时，一路往后接的长句在口语里活得好好的，能接到三四层。限制挑的不是难度，是形状——被筛掉的偏偏是那种要求你把一个成分悬在半空、越过一段材料再回来接住的结构。本文由此提出：塑形语言的不是输入的贫乏、不是社会互动的丰富、也不是记忆的容量，而是信道让不让你回头，以及回头的代价由谁承担；而这个量与记忆容量、与呈现速度、与看还是听，三者都正交——给足记忆而不许回头，同一批结构照样长不出来。文末给十处跨领域读数、两处反过来迫使它让步的地方、一个用现成范式一个下午就能跑的四组实验（三家解释在这一个实验里给出三种不同的分组），以及一条写死到 2031 年的赌注。

引言

这篇文章从一个几乎不被当成问题的事实出发：你没法让刚才那句话原样再来一遍。

在面对面说话时，你可以请对方重说——但那已经是另一次说了，而且你得开口。在广播、讲座、一次性通知里，连这个都没有。而在一页纸上，你可以把同一句读七遍，不惊动任何人。

这件事之所以值得现在问，有两个理由。其一，关于语言被什么塑成，四种主要解释争了半个世纪，各有实验、各有阵地，而它们对同一批现象给出的归因互相取消；一个反复被争论却争不出结果的地方，通常说明大家争的不是主变量。其二，人类第一次大规模地拥有了反常的信道：可以随意倒回去的口语（录音与录像），以及不许回看的文字（限时屏幕阅读、逐词呈现）。这两样东西把原本永远绑在一起的两个变量拆开了，于是一个此前无法被提出的问题，今天可以被直接实验。

本文立的判断只有一条：塑形语言的是信道的可回头性，而不是输入的贫乏、互动的丰富或记忆的容量。

路线图如下：第二到第四章逐一交代四家解释说到了哪一步、又在哪里停下；第五章把矛盾写死；第六章给出变量的定义与三条读数；第七章加上第二根轴，得到四格与两条方向相反的排序；第八章给四格各找一个真实标本；第九章给出一条已经躺在语料里的读数；第十章说明四家为什么各自推不出这一条；第十一章在十个别的领域里兑现；第十二章报告两处迫使命题让步的地方；第十三章与最接近的六种说法逐条分界；第十四章交出不适用的范围；第十五章写清按它设计干预会怎样出事；第十六章给出证伪清单；第十七章回答“为什么这件事一直没被看见”；第十八章处理本文自己也是标本这件事；末章是一条写死日期的赌注。

一、三家在争同一件事，而它们的处方互相取消

先摆一个不需要论证的事实：世界上所有的语言，都能让一个三岁的孩子在没人教语法的情况下用起来；而同样这些语言，写下来的时候可以复杂到没有人能听懂。

这两件事摆在一起就是个麻烦。语言到底是被什么塑成现在这个样子的？三条成熟的解释各有阵地，也各有实验，而它们互相拆台。

第一条来自生成语法。 它的支点是一个论证：孩子听到的语言输入太少、太乱、太不带纠错，不足以支持他最终掌握的那套规则；能解释这个落差的只有一件事——那套规则的骨架不是学来的，是长出来的。语言之所以在世界各地长成相似的形状，是因为它必须长在同一副骨架上。这条解释的处方很明确：要解释语言的形态，去看那副先天的骨架，输入只负责设定几个开关。

第二条来自发展心理学与灵长类认知。 它反过来说：先天骨架是个不必要的假设。孩子学会语言，靠的是一套更基础的社会认知能力——跟着大人的目光看同一样东西、理解对方在意图什么、从大量具体的用例里一点点抽出模式。语言不是被激活的，是被用出来的。这条解释的处方是：要解释语言的形态，去看人怎么用它，尤其是两个人怎么在一件共同关心的事上把注意力对齐。

第三条来自认知科学与语言演化，而且它把前两条一起绕开了。 它说，你们都在争“知识从哪儿来”，可真正卡住语言的不是知识，是时间。人的记忆转瞬即逝，新进来的材料飞快地覆盖旧的；一句话说过去就没了，来不及原样存住。所以语言系统必须一边听一边压缩、一边递交，把音压成词、词压成块、块压成意思——压不完就永远丢了。这条解释的处方是：语言的形态是被这个当下就得处理完的瓶颈雕出来的，从处理、习得到历史演变，都是这一个约束的下游。

三条并排放着，麻烦就出来了。第一条说输入太穷，第二条说输入其实足够丰富只要你会看，第三条说输入丰不丰富根本不是重点、重点是它不等人。

而这一支自己内部还有一场没打完的仗。同一年的评论里，有人直接顶了回去：**记忆限制与组块的能力是因人而异、因任务而异的，它解释不了语言结构那么整齐的规律；真正塑形结构的更可能在产出那一端——说话的人觉得哪种说法顺口，顺口的说法在语言里变多，多了之后听的人也就更好懂。这一条把因果箭头从“听不过来”掉头成了“说不顺口”。**

于是我们有四种说法在同一个现象上给出四种不相容的归因：先天骨架、社会互动、记忆瓶颈、产出难易。

本文要主张的是：**这四种说法漏掉了同一件事，而且漏得非常一致——它们都默认了，一段话只能听一遍。**

准确地说，它们都把“信息进来了就过去了”当作背景条件，而不是当作变量。可这件事恰恰是可以取不同值的：有的信道允许你回头再看一遍，有的不允许；有的回头只要你自己动一下眼睛，有的回头必须开口请人重说。

本文的承重判断是一句话：

塑形语言的不是输入的贫乏，不是社会互动的丰富，也不是记忆的容量，而是信道的不可回头性——一段话能不能被请求重播，以及请求重播的代价由谁承担。

这句话与记忆容量是正交的：给你无限的记忆，只要不许回头，同一批结构照样长不出来；反过来，把记忆压到很小但允许你反复回看，那些结构又会回来。第七章会把这一点做成一个一下午就能跑的实验。

二、先天骨架那一条说到哪一步

第一条解释的核心论证有一个流传很广的名字：刺激的贫乏。

它的形状是这样的：孩子最终掌握的规则里，有一些是他不可能从听到的话里归纳出来的——因为要归纳出那条规则，他必须听到某一类反例被明确否定，而现实中没有人会去否定它，他也几乎听不到那种句子。既然结论超出了证据，那么多出来的那部分就必须来自别处；这个"别处"就是先天的结构。

这条论证极其锋利，也因此挨了三十年的检验。批评主要落在两处：一是"孩子听不到"这件事被高估了，大规模语料出来之后，很多号称零频的结构其实并非零频；二是归纳所需要的东西被低估了，统计学习加上一点点偏置能走得比原先设想的远得多。这些争论本文不重复，因为它们打的是"输入到底够不够"，而本文要指出的是这条解释的**结构**里有一个从未被当作变量的零件。

那个零件是：**输入的次数**。

刺激贫乏论证里，"孩子听到了什么"总是被记成一个集合——哪些句子出现过、哪些没出现过。集合是没有时间的：一个句子出现过一次，和出现过十次，在集合里是一回事；一个句子听过之后能不能倒回去再听一遍，在集合里根本没有位置。

可只要把时间放回去，问题的形状立刻变了。孩子面对的从来不是一个可以反复查阅的语料库，而是一条**一次性流过去的溪水**。他不能说"请把刚才那句从第三个词开始重放一遍"。他能做的只有两件事：当场处理掉，或者永远失去。

这不是在替刺激贫乏论证补一个细节，它可能改变论证的走向。因为"证据不足以支持结论"这句话里的"证据"，如果指的是**可以被反复检视的证据**，那么任何一条不可回头的信道都会自动造成贫乏——不是因为世界给得少，而是因为拿到手的東西留不住。而世界给得多不多，是一个关于环境的问题；留不留得住，是一个关于**信道**的问题。两者可以独立变化。

先天骨架这条解释说到了一件真的事：**最终掌握的东西超出了任何一次输入所能提供的**。它没说到的是：这个超出，有多少是因为输入本身稀少，有多少是因为输入不可被重访。前一个原因要求你去找一副骨架，后一个原因只要求你去看那条信道。

三、基于用法那一条说到哪一步

第二条解释把重心从骨架移到了互动。

它的证据链相当扎实。孩子在会说话之前很久，就已经会跟着大人的目光看同一樣东西、会用手指把别人的注意力引到自己关心的地方、会在对方拿不到东西时递过去。这一整套"我们此刻在共同关心同一件事"的能力，是灵长类里人独有的那一段。语言就长在这套能力上：先有共同注意，才有指着某物说出一个词；先能猜出对方想干什么，才能把一串声音理解成一个意图。至于语法，它是从大量具体用例里逐步抽出来的——先记住整块的说法，再慢慢发现哪些位置可以换、哪些不能。

这条解释的力量在于它把"学会语言"放回了一个可以观察的场景里：两个人，一件东西，一次对齐。它也解释了为什么语言在互动最密集的地方长得最快。

但它同样有一个未被当作变量的零件，而且这个零件就藏在它最得意的那个场景里。

面对面的互动，是所有信道里唯一可以当场要求重来的那一种。

孩子没听清，可以皱眉；大人看见他皱眉，会自动放慢、拆开、换个说法再说一遍。这个"再说一遍"不是修辞，它是一次真实的重播请求，而且这个请求几乎不用付出什么——一个疑惑的表情就够了。研究者们把这一族现象叫作修补，通常放在"互动如何支持学习"的名下讨论。

本文要指出的是：**这些修补之所以能发生，前提是这条信道允许回头；而这个前提本身，从来没有被当成可以变化的东西。**

如果把互动的丰富度固定住——同样多的共同注意、同样多的意图理解、同样多的用例——只把"能不能要求重说"这一件事拿掉，会发生什么？基于用法的解释没有位置容纳这个问题，因为在它的图景里，互动天然就带着修补，两者是一体的。

而现实里它们是可以被分开的。听广播长大的孩子、只能看不能问的课堂、字幕一闪而过的视频、一次性播报的通知——这些场合有输入、有意图、甚至有共同关注的对象，唯独没有"再说一遍"。

基于用法这条解释说到了一件真的事：**语言长在共同注意上**。它没说到的：共同注意能不能被反复重建，取决于信道让不让你回去。

四、瓶颈那一条说到哪一步，以及它内部那场没打完的仗

第三条解释离本文最近，所以说得最细。

它的表述是：人的记忆是转瞬即逝的，新材料迅速覆盖旧材料；因此语言系统必须"急切地"重新编码与压缩输入，把声音压成音节、音节压成词、词压成短语、短语压成意思，一层压完立刻交给下一层；而由于这个瓶颈在每一层都重现，语言表征必须是多层的；又由于压缩必须在下一批材料到来之前完成，系统必须把一切可用的信息都用来做预测。这一整套约束被概括成一句口号式的名字：**当下，或者永不**。

它的解释力确实很宽：处理、习得、历时演变、乃至语言为什么长成这个结构，都被收进同一条约束的下游。本文接受这条约束的存在，而且后面还要用它。

争议在于它够不够。同期就有一条正面的反驳：**记忆的限制与组块的能力是高度可变的**——不同的人差别很大，同一个人在不同任务上差别也很大，而语言结构的规律却相当整齐；用一个可变的量去解释一个不变的规律，中间少了一步。这条反驳给出的替代路线是：结构更多来自**产出端**——说的人觉得哪种排列容易组织，容易的排列在语言中出现得多，出现得多之后听的人处理起来也就更快。因果箭头由此从"听不过来"掉头成了"说不顺口"。

两边都有道理，而本文认为它们在同一处失手。

"记忆转瞬即逝"与"这段话不能重来"，是两件不同的事，而它们在这场辩论里一直被当成同一件。

记忆是关于我这一侧的容量：我能同时抓住多少东西、抓多久。可重来是关于信道的：那段材料还在不在外面，我能不能再取一次。这两件事在面对面说话时恰好绑在一起——声音说过去就没了，所以既超出记忆也无法重取；正因为它们在最常见的场合里总是同时出现，才没有人把它们拆开。

一旦拆开，立刻能看出它们不是一回事：

- 一个记忆极差的人，坐在一段可以随意倒回去的文字前，能读懂任何复杂的句子——他只要多回去几次。
- 一个记忆极好的人，坐在一段只放一遍的广播前，遇到那种必须把前半句悬着、等很久才落地的结构，照样会崩。

给足记忆并不能替代允许回头。 这就是本文与瓶颈那一条的分界线所在，而且它可以被实验分开——第十六章给出的第一个实验，用的就是这条缝。

至于产出端那条反驳，本文同样要与它分开。它的变量在说的人那一侧，本文的变量在信道上；而信道可以在说的人完全不变的情况下改变。第十三章会把这条分离线做成一个具体的对照。

五、四条不能同时为真

第一对：输入是穷还是富。

先天骨架说输入穷到无法支持结论，因此必须假设骨架；基于用法说输入其实很富，只要把社会认知那一层算进去。两条对同一批材料给出相反的读数，而且不能靠"程度不同"调和——一条要求缺口存在，另一条要求缺口不存在。

第二对：重点在知识还是在时间。

前两条都在争"孩子凭什么知道"，瓶颈那一条说这个争论问错了：无论知识从哪儿来，它都必须在几百毫秒之内被处理掉，否则谈不上知道。这不是补充，是改换赛道——它把两条解释共同的追问对象取消掉了。

第三对：因果箭头朝哪边。

瓶颈说是听的人处理不过来，于是结构被压浅；产出端的反驳说是说的人组织不顺，于是某些排列变多。两条给出相反的因果方向，而且各自都能解释同一批语料。

三对矛盾指向同一个空缺。四种说法各自盯着一样东西——骨架、互动、记忆、产出——却都把同一件事当成了不需要说明的背景：一段话说过去就没了，你没法请它重来。

把这件背景拿出来当变量，三对矛盾同时松开：

- 输入之所以显得穷，不必是因为世界给得少。一条不可回头的信道会自动制造贫乏：给一百次也只等于一百个一次，因为每一次都在你还没处理完的时候就流走了。缺口是真的，但它的原因在信道上，不在环境里，也不必去假设一副骨架来填。
- 互动之所以显得富，很大一部分来自一件极少被单列的事：面对面是唯一可以低成本要求重播的信道。孩子一个疑惑的表情就能换来"再说一遍"。把互动的其余部分全部保留、只拿掉这一件，富就会塌下去一大半。
- 而听与说之所以吵不出结果，是因为两边都在同一层里找原因。信道的可回头性既不在听的人那一侧，也不在说的人那一侧，它在两者之间；它同时决定了听的人能不能重取，和说的人必须预先垫进去多少冗余。第七章的第二根轴处理的正是这一点。

还有一件更要紧的事，四条解释都没有位置容纳。

如果限制来自记忆容量或处理速度，那么被压掉的应当是**复杂的结构**——越复杂越难，难到某个点就消失。可真实语料里被压掉的不是"复杂"，是一种**特定形状的复杂**：那种要求你把一个还没说完的东西悬在半空、等一段材料过去之后再回来接上的结构。而另一种同样复杂、甚至更长的结构——一路往后接、不用回头的那种——在口语里活得好好的。

限制不挑复杂度，它挑的是"要不要回头"。第九章给出这条读数的现成证据。

六、换一个变量：能不能回头

可回头性指这样一件事：接收的一方，在材料流过去之后，还能不能让同一段材料原样地再来一次。

四点澄清，缺一点它就会被读回旧概念。

第一，它不是记忆。 记忆是"我这一侧还抓着多少"，可回头性是"那段东西还在不在外面、我能不能再取一次"。两者可以独立取值：记忆差而可回头，读得懂任何句子；记忆好而不可回头，遇到需要悬置的结构照样崩。这条正交性是本文全部主张的支点，也是它最容易被推翻的地方。

第二，它不是速度。 一段话说得慢，不等于可以回头；一段文字读得快，也不妨碍随时回退。真正区分的不是每秒进来多少，而是**流过去的那部分还能不能被重新取用**。第十六章的第一个实验专门把速度控住，只变回退。

第三，它不是媒介。 说与写是两种媒介，可回头性是一种属性，而属性并不跟着媒介走：一段录像可以随意倒回去，它是"说"的；一场没有草稿、只念一遍的口头通知，和一份限时不许翻页的书面测验，在这个属性上是同一格。把变量定在媒介上，会把这些格子搅成一团——第十三章处理这个混淆。

第四，它是程度，不是有无。 面对面对话是可以回头的，只是要开口；广播不可回头，除非你自己录了下来。真实的信道分布在一条连续谱上，而决定它落在哪儿的，往往不是技术，是**代价**。

怎么测——三条互相独立的读数。

其一，**重播代价**。让接收者要回刚才那一段，需要付出什么：一次眼动、一次按键、一句请求、还是不可能。前两者近乎零，第三者带社会成本，第四者是无穷。这条可以直接按场景编码，不需要仪器。

其二，**悬置容忍度**。给定一种要求把成分悬在半空、等一段材料之后再接上的结构，接收者在多深的层次上仍能正确理解。这条有现成的实验范式，也有现成的语料读数。

其三，**冗余率**。同样的意思，在这条信道里被重复了多少次、用了多少提示性的套语与平行结构。这一条与前一条方向相反——本文预测它们在四个格子里呈镜像分布，而这个镜像正是最容易被证伪的地方。

有了这个变量，本文的承重命题可以写成一句：

语言在一条信道里能长成什么形状，取决于那条信道允许接收者以多大代价请求重播；而被这个约束筛掉的不是复杂结构，是需要回头才能完成的结构。

机制只有两步，都不需要假设任何人不聪明：

第一步，悬置的风险。 每当一个成分被说出而它的归宿还没到，接收者就欠着一笔账。信道可回头时，这笔账最坏的下场是"再看一眼"；不可回头时，这笔账的下场是永久损失一整句的意思。

第二步，说者的预付。 说的人知道这一点，于是在不可回头的信道里，他会预先把回头的活儿替听者做掉——重复、拆短、把要悬置的东西换成一路往后接的排列、用固定套语让人不必现场分析。这些做法长期沉淀下来，就成了这条信道里的"语体"。

两步走完，语言的形状就分岔了，而全程没有出现任何关于骨架、意图或记忆容量的假设。

七、第二根轴：回头的代价由谁承担

只有一根轴的东西不是维度，是名字。可回头性必须与另一根独立的轴交叉，才能把现实切成可分辨的格子。

第二根轴是：**回头这件事的代价，落在接收者身上还是发出者身上。**

它与第一根轴独立，因为"能不能回头"与"回头由谁买单"可以自由组合：

- 我自己倒回去重读一遍——代价是我的几秒钟，无人知道，发出者不受影响。
- 我开口说"你刚才说什么"——代价落在发出者身上（他得重说），而且这个请求是**公开的**，我要为自己没听懂付出一点面子。
- 我没法回头，只能自己在脑子里补——代价全在我，且补不回来。
- 我没法回头，但发出者事先就把话说了两遍、押了韵、用了人人都熟的套语——**他预付了这笔代价**。

两根轴交叉，得到四格。这张表是全文的骨架。

	可回头	不可回头
代价由接收者承担	格Ⅰ·自取重播 纸面文字、可倒带的录音与录像。回头无声、无限次、不惊动任何人	格Ⅲ·独自受苦 讲座、广播、一次性通知。听不懂只能自己扛，且扛不住也没人知道
代价由发出者承担	格Ⅱ·请求重说 面对面对话、可以打断的教学。回头要开口，带社会成本，但换来的是为你定制的重说	格Ⅳ·预先垫付 口传程式、仪式套语、飞行与航海的标准通话。发出者用重复、套语、平行结构把回头的活儿事先做掉

本文的可裁决排序有两条，方向相反，这是它最要紧的一处：

可容许的悬置深度：格Ⅰ > 格Ⅱ > 格Ⅲ > 格Ⅳ。冗余率：格Ⅳ > 格Ⅲ > 格Ⅱ > 格Ⅰ。

两条排序必须同时成立，否则本文错。若某个格子既允许深悬置又高冗余，或既不允许悬置又低冗余，那说明这两个量不是同一个约束的两面，本文的机制陈述要重写。

这个排序与四家给出的都不同：

- 先天骨架不预测任何按信道分布的差异——骨架是同一副。
- 基于用法会把格Ⅱ排在最好一端（互动最丰富），本文把格Ⅰ排在最好一端。
- 瓶颈那一条按处理负荷排序，会把格Ⅰ与格Ⅱ排在一起（都不那么急），本文把它们分开，且预测它们的差别恰恰体现在**冗余**上而不是在深度上。
- 产出端那一条不预测接收侧的操纵会改变结构，而本文的第一个实验正是只动接收侧。

还有一件必须先说清的事，免得这张表被读成“文字优于口语”。四个格子在**能承载什么**上不同，不在**好不好**上不同。格Ⅳ的口传程式能做到格Ⅰ做不到的事：在没有任何外部存储的条件下，让一整套知识跨代传下去且几乎不走样。那不是简陋，那是把冗余当存储用。

八、四个格子的现实标本

格Ⅰ·自取重播。 一页纸、一份文件、一段可以拖动进度条的录像。它的特征不是“用眼睛看”，是**回头无声**：你可以把同一句读七遍，没有人知道，也没有人被打扰。

于是这一格里长出了别处长不出的东西：把一个主语说出来、插进去一整段限定、再回来给它谓语；用一个代词指向三十行之前的某物；把整段论证的结论压到最后一句。这些做法有一个共同点——**它们都要求读者在某个时刻回去一次**，而在这一格里，回去几乎不要钱。

顺带说一件常被忽略的事：这一格的语体是**被训练出来的**。会写这种句子的人，几乎都是先读了大量这种句子的人。这不是天赋分布，是信道分布。

格Ⅱ·请求重说。 面对面对话。这一格的独特之处不在“有互动”，而在于**回头的按钮握在接收者手里，但按下去要发出声音**。

代价不是零，这一点很要紧。正因为要开口、要承认自己没跟上，人会在很多时候选择不按——于是说话的人也知道对方可能不会按，从而把话说得比对着一份文件时更碎、更短、更爱重复。这一格的语体是两种力量的平衡：可以要求重说，但不好意思老要求。

一个可查的细节：成人对幼儿说话时，重复率与拆解率显著高于成人之间的对话。通常的解释是“迁就理解水平”。本文给一个更省的解释：**幼儿按那个按钮的能力最差**（他甚至不知道自己没懂），于是成人自动把代价接了过去——这已经是格Ⅱ在向格Ⅳ滑动。

格III·独自受苦。 讲座、广播、一次性通知、没有字幕的直播。有说的人，有听的人，但没有那个按钮。

这一格最能说明本文的变量不是媒介：讲座与对话都是"说"，但一个有按钮一个没有，而它们的语体差别之大，超过讲座与文章之间的差别。一个有经验的演讲者做的全部功课，几乎都可以概括成一句话——**替听众把回头这件事先做掉**：先说结论、报数字前先说"三点"、每讲完一段回头总结一次。这些技巧在写文章时不但不必要，还显得啰嗦。

格IV·预先垫付。 口传史诗、仪式经文、行业里的标准通话。

飞行与航海的标准通话是最干净的标本：话必须简短、必须复诵、必须用固定的词表、必须避免任何需要悬置的结构。这不是文化，是被信道逼出来的设计——无线电里回头的代价接近无穷，于是一切都被预付。

口传程式则把这件事做到了极致：格律、套语、平行句、成对的固定搭配。这些东西在文章里是修辞，在这一格里是**存储介质**——它们让听者不必现场分析就能接住，也让传者不必逐字记忆就能复述。

四格摆出来之后，一件事变得清楚：**我们通常说的"口语比书面语简单"，其实混了三个格子（II、III、IV），而这三格之间的差别，不比它们与格I的差别小。**

九、一条已经躺在语料里的读数

本文最要紧的一条预测——**被筛掉的不是复杂结构，而是需要回头的结构**——不必等新实验，它在语料里已经有了读数，而且这个读数出自一位并不打算论证本文的语言学家。

先把两种"复杂"分开。

第一种叫**往后接**：一个句子说完主干，再往后挂一个从句，从句说完再往后挂一个。"我忘了她知道我是谁"就是这种。它可以接得很长，但接收者永远不必回头——每一段接上来的时候，前面那一段已经落地了。

第二种叫**中间套**：把一个从句塞进另一个句子的当中，让主干的后半截被推到很远的地方去。"那只被那只被那条狗追的猫杀死的老鼠吃了麦芽"就是这种。它要求接收者把"老鼠"这个主语悬在半空，穿过两层材料，再回到最后那个动词上把它接住。

按记忆容量或处理速度来解释，这两种复杂都该随深度变难，只是难的速率不同。按本文的变量，它们应当在**不可回头的信道里分道扬镳**：往后接不要求回头，所以口语里应当活得好好的；中间套要求回头，所以口语里应当近乎消失。

这正是语料给出的结果。一项跨七种欧洲语言、对真实语料中多重中间套的系统清点显示：**书面语里中间套的最大深度是三**，而且深度三的真实例子在整个调查中只有十三个；**口语里，多重中间套即使只到深度二也罕见到近乎不存在**。同一项工作还指出，往后接的情形完全不同——它在口语里可以到深度三、在书面语里更深，是真正能产的那一种。

还有一个细节更值得停一下：这项调查追踪了多重中间套在欧洲语言里的来历，结论是**它随着书面语的出现而出现**，成型于古典时期那种讲究"周期句"的书面文体。也就是说，这不是一种一直存在、只是口语懒得用的结构；它是在**信道变得可以回头之后才被造出来的**。

三条读数合起来，正好是本文的机制陈述：

1. 不可回头的信道里，中间套近乎为零，而往后接照常繁茂——**限制挑的是形状，不是难度**。
2. 中间套的上限（三）与往后接的上限（更深）不是同一个数——若限制来自同一个容量，这两个上限没有理由分得这么开。
3. 中间套的历史起点与可回头信道的历史起点重合——**结构跟着信道走**。

必须说清楚一件事，免得这一节被当成结论：这项调查的作者并没有主张本文的命题。他给出的是语料事实与对"无限递归"这一假设的质疑；把这些事实读成"信道可回头性的效应"，是本文的推论，不是他的主张。而正因为它是推论，它可以错——第十六章给出的第一个实验，就是用来判它对错的。

十、四家各自为什么到不了这一条

先天骨架那一条到不了，因为它的证据是集合，不是流。

刺激贫乏论证把孩子听到的东西记成一个句子集合，而集合没有时间也没有次数。要它把"能不能重取"当变量，等于要它把证据的形态从集合改成流——而这一改，"贫乏"这个词的意思本身就变了：贫乏可能不是世界给得少，而是给过的留不住。

基于用法那一条到不了，因为它把互动与修补捆成了一体。

在它的图景里，共同注意天然带着"没懂再说一遍"。于是它推不出这样一个问题：把共同注意、意图理解、用例数量全部保持不变，只拿掉"再说一遍"这一件，会怎样。这个问题在它的框架里不成立，而在现实里到处都是——广播、讲座、字幕、一次性播报。

瓶颈那一条到不了，因为它把"记忆转瞬即逝"与"材料不可重取"当成了同一件事。

这不是它的疏忽：在面对面说话这个最常见的场合里，两者确实同时出现。但它们可以被分开，而且分开之后给出不同的预测——给足记忆、不许回头，本文预测中间套照样长不出来，瓶颈那一条预测约束应当随记忆容量放松而放松。这一条差别是可实验的，第十六章第一项就是它。

产出端那条反驳到不了，因为它的变量在说的人那一侧。

它主张结构来自"哪种排列容易组织"。而本文的第一个实验只改接收侧——同一批句子、同一批说话人、同一份材料，只变读者能不能回退。若结构效应在这种操纵下出现，产出端那条解释无法容纳它。

最后一句，关于四家共同的默认。

四家都默认信道是给定的，这在各自的场地里都合理：生成语法研究的是能力而非使用条件，发展心理学研究的是面对面这一种场景，瓶颈研究的是在线处理，产出研究的是说话人。本文所做的，只是把这个在四个场地里都被当作背景的东西挪到前台，并且发现——一旦它成为变量，四条互相取消的预测同时被安顿，而且它们各自的适用范围恰好对应四个格子中的一个。

十一、它在别处也兑现

一个判断如果只在它出生的领域里成立，那是一条经验规律，不是一个辨别维度。这一章把它放到别处，每一处都要求它给出一件具体的、可被查证的预测。

其一，法律文本。 合同与法条是全部信道里回头代价最低的一格——不但可以重读，还可以对照、检索、逐条援引。本文预测：同一法域里，供当庭宣读的部分（判词主文、告知义务）在悬置深度上应显著低于供检索援引的部分（条文与释义），而这两者出自同一批起草者、同一套术语。这条差别若不存在，说明起草者的习惯压过了信道，本文在这一处失效。

其二，飞行与航海通话。 无线电里回头的代价接近无穷（噪声、时机、频率占用），于是标准通话把一切预付：固定词表、强制复诵、禁止长句。本文预测：在引入可回放记录与文字数据链之后，同一岗位的口头通话会缓慢变长、变复杂，因为回头的代价降了。这条有现成的历史材料可查。

其三，字幕与配音。字幕可以随录像一起倒回去，现场同传不能。本文预测：同一部作品的字幕译文，其悬置深度高于同一位译者做的现场口译，而冗余率相反。译者是同一个人，语言是同一门，只有信道变了。

其四，教学录像与直播课。二者内容可以完全相同。本文预测：允许回放的课程里，教师使用需要悬置的表述的比例更高，而"我再说一遍"类的显式重复更低；且这个差别在教师知道课程会被录下来时就已经出现——他不必等学生真的回放。

其五，儿童指导语。前面提过：成人对幼儿说话时的高重复率，通常被解释为迁就理解水平。本文给一个可分开的预测：在幼儿获得了可靠的"没听懂就问"能力之后，成人的重复率应当下降，而这个下降的时点应当跟着"会问"而不是跟着"听得懂"。这两条曲线在发展数据里可以分开。

其六，手语。手语是"看"的，但它同样一过就没了——除非录下来。本文因此预测：手语与口语在悬置深度上应当同格，而与书面语不同格；若手语表现得像书面语，说明本文把变量定错了地方（错在信道的感官通道而不是可回头性）。这是本文最容易被现成数据打掉的一条。

其七，代码与配置。代码是可回头的极致，于是它长出了深度极高的嵌套。而告警消息、日志行、命令行提示是不可回头的（滚过去就滚过去了），它们的语体恰恰是电报式的、平行的、高冗余的。本文预测：同一个团队写的代码嵌套深度与告警文案的嵌套深度差异，大于不同团队之间的代码差异。

其八，会议纪要与会议本身。同一件事，说的时候必须线性推进、边说边确认，写成纪要之后立刻出现层级、编号、"综上"与前后指代。如果这只是"书面更正式"，那么把纪要念出来应当没问题——而实际上念纪要几乎总要重新组织语言，这正是信道在起作用。

其九，流式语音识别与逐字输出的模型。一个必须边听边定稿、不许回改的系统，与一个可以看完整段再输出的系统，面对同一段输入会给出不同的结构选择。本文预测：前者会系统性地偏向往后接、避免中间套——不是因为它更笨，而是因为它落在格Ⅲ。这条可以在现成系统上直接测，而且是本文最便宜的一次跨域检验。

其十，考试。允许翻回去的阅读题与只播一遍的听力题，考的往往被认为是同一种理解力。本文预测：两者的难度差异不是均匀的，而是集中在需要悬置的题型上；如果把听力材料改成允许重听两次，那一类题的正确率提升应显著高于其他题型。这条一个学期就能测完，且教育测量领域本来就有现成数据。

十二、两处反过来加约束的地方

只会附和的例证是装饰。真正说明一个判断有承重能力的，是它在某处遇到阻力并被迫让步。本文遇到两处。

第一处：口语并不简单，它只是走另一种复杂。

一种流行的读法是"口语简单、书面复杂"，而语体研究早就指出这是错的。口语在句法的绵密上并不输——一句话里挂着好几个小句，一路接下去，结构相当繁复；书面语的复杂不在小句多，而在把大量内容压进名词性的成分里，用名物化、修饰语、定语堆叠把一件事塞进一个短语。两者是两种不同的复杂，不是多与少。

这一条逼本文让步两次。

其一，本文不能主张"不可回头的信道产生简单的语言"，那是错的，而且有充分的反证。能主张的只有更窄的一句：不可回头的信道排斥需要回头才能完成的结构。绵密的小句链完全可以在口语里活得很好——因为它们一路往后接，不要求任何回头。

其二，也是更有价值的一次修正：既然书面语的复杂主要长在名词性堆叠上，那么本文的预测就应当分开——不可回头的信道压制的是"需要悬置的句法"，而不是"信息密度"。一个高密度但一路往后接的表达，在口

语里应当照样可行；反过来，一个低密度但要求悬置的表达，在口语里照样出不来。这个预测比原先那条更细，也更容易被打掉：只要找到一批高悬置、低密度的口语结构，本文就错了。

第二处：自定步速的阅读实验说明"能停"不等于"能退"。

心理语言学里有一族标准做法：让被试自己按键，一次显示一个词或一个短语，按一下出下一个——速度完全由他控制，但**已经过去的看不见了**。这是一个既非格 I 也非格 III 的中间态：时间任你花，材料却不可重取。

它对本文是麻烦，也是礼物。麻烦在于：如果"能停"就够了，那么本文强调的"能退"就是多余的；礼物在于：这个范式的存在，让"停"与"退"可以被干净地分开测——而据本文所知，**几乎没有人把它当成一个关于信道的操纵来用**，它一直被当作测量阅读时间的工具。

由此本文被迫把命题再收窄一格：**起作用的不是接收者对时间的控制权，而是材料的可重取性**。并且给出一个立刻可做的判据：在自定步速且不可回退的条件下，中间套的理解应当接近听觉条件，而不是接近自由阅读。若结果相反——自定步速就足以恢复中间套的可理解性——那么本文的变量应当从"可回头"改成"可控速"，整篇的第二根轴要跟着重画。

这两处让步都不是补丁。它们把本文从"信道决定语言复杂度"这个过宽的说法，推到了两个更窄、也更可测的量上：**需要悬置的结构，与材料的可重取性**。

十三、与最接近的几种说法划清界线

一个判断最危险的时刻不是被反驳，是被读者顺手归到某个他早就知道的名目底下。这一章逐个处理，每一处都必须落到"同一个案例，那个说法判 A，本文判非 A"。

与"口语性与书写性"这一支。 这是本文最近的邻居，必须先处理。那一支的核心洞见是：声音是转瞬即逝的，它存在的方式就是正在消失；因此以口传为主的文化在思维与表达上呈现出可辨认的特征——爱用并列而不是从属、爱用套语、贴近生活世界、依赖记忆术。凡本文谈的那些差别，几乎都能在那一支里找到一句现成的话。

分离线在于：**那一支的变量是媒介与文化（口传的 vs 有文字的），本文的变量是信道的可回头性，而后者横切前者**。判决性对照有两个方向：其一，取一段**可回头的"口语"**——一段可以随意倒回去的录音——本文预测它在悬置深度上向格 I 靠拢，那一支预测它仍是口语；其二，取一段**不可回头的"文字"**——限时不许回看的屏幕阅读——本文预测它向格 III 靠拢，那一支预测它仍是文字。这两个对照都便宜，且预测方向相反。

与"当下或永不"那一条。 分离线已在第四章给出：记忆容量与材料可重取性正交。判决性对照：控住呈现速度、只变能否回退。那一条预测差异应随记忆负荷变化而变化，本文预测差异随可退与否而变化，且在低负荷条件下依然存在。

与产出端的解释。 它的变量在说的人那一侧。本文的第一个实验只动接收侧：同一批材料、同一个作者、同一种呈现速度，只变读者能不能回退。若结构效应在这种操纵下出现，产出端解释无法容纳；若不出现，本文这条命题被削掉一半。

与语体研究里的"绵密与密度"。 上一章已经把它当作约束吸收进来了，这里只补分离线：那一支描述的是两种复杂在两种媒介里的分布，本文给的是一个**跨媒介的判据**——要不要回头。可裁决之处在于，本文预测存在"不可回头的书面语"这一格，且它的语体应当更像讲座而不是更像文章；那一支不预测这一格的存在。

与口传程式研究。 关于史诗如何靠套语与格律被记住与被创作，有一整套成熟的研究，它占住了本文格 IV 的全部内容。分离线：那一支解释的是**歌手怎么做到的**（创作与记忆的技术），本文解释的是**为什么这条信道非要求这样不可**（回头代价无穷时，冗余必须被预付）。两者不冲突，但可分：本文预测**在任何回头代价接**

近无穷的信道里都应当出现同族的预付手段，包括与史诗毫无文化关系的现代场合——飞行通话、急救口令、军事口令。那一支不作这个预测。

与识字的认知后果研究。 有一支文献主张读写训练本身改变了思维方式与语言意识。分离线：那一支的自变量是**训练**（会不会读写），本文的自变量是**当下这条信道**。可裁决：同一批完全识字的成年人，在可回头与不可回头两种条件下应当表现出本文预测的差别——训练是同一份，差别只能来自信道。

最后指认最近的那一位：**口语性与书写性那一支**。本文之所以仍不能被它吸收，只靠一件事——它的分界线画在媒介上，本文的画在可回头性上，而这两条线在“可回放的口语”与“不可回看的书面”这两格里正面分开。这两格是否真的表现出本文预测的样子，是本文最该被攻击的地方。

十四、它不适用在哪

其一，语言的最小单位那一层。 音系与形态的规律（哪些音可以拼在一起、词怎么变形）在本文的射程之外。可回头性作用在“要不要把一个成分悬着”这个层面上，而音节内部没有这种悬置。任何把本文推到音变上的做法都超出了它的证据。

其二，只有一个人的语言使用。 自言自语、写给自己看的笔记、内心的思考——这些场合没有一个独立的接收者，“回头的代价由谁承担”这根轴退化。本文对这一类没有说法，而这一类恰恰是很多语言与思维关系研究的主战场。

其三，接收者本来就无法悬置的场合。 儿童在某个年龄之前、失语症的某些类型、极端疲劳与分心状态——这些情况下，即使信道完全可回头，需要悬置的结构照样出不来。此时限制来自接收者，不来自信道，本文的变量不解释它。

其四，高度约定化的书面场合。 有些书面语体的复杂度更多来自行业惯例与法律责任，而不是来自可回头性——保险条款写成那样，一半是因为每一个逗号都被诉讼检验过。本文只主张可回头性是**必要条件**（不可回头就长不出来），不主张它是**充分条件**（可回头不一定就会长出来）。

其五，我说不清代价的量纲。 第七章那根“代价由谁承担”的轴，在两端是清楚的（一次眼动 vs 开口请求），中间是含糊的：一次微表情算多少代价？一次“嗯？”算多少？本文给不出刻度，只能给排序。这是它目前最钝的地方。

其六，极短的时间尺度。 本文关于语体与结构分布的读数以语料为单位，需要成千上万句。任何用几个句子的直觉来检验它的做法都会得到噪声——包括读者此刻在心里造的那几个反例。

把六条合起来，本文实际主张的范围要比标题窄得多：**在有独立接收者、接收者具备悬置能力、且我们关心的是句法层面的结构分布时，信道的可回头性是“需要悬置的结构能否存在”的必要条件。** 超出这个范围的部分，我不认账。

十五、按它设计干预，会怎样出事

第一种：把“可回头”当成好东西，于是到处装回放。

如果悬置深度是好事，那么给一切信道装上回放按钮就是进步。这个推论是错的，而且错得很具体：**回放按钮会把回头的代价从发出者身上挪回接收者身上**。在格Ⅱ里，我没听懂可以让你重说一遍——重说的是你，而且你会为我定制；在格Ⅰ里，我只能自己回去看，看不懂就是我的事。装上回放的那一刻，发出者的预付义务被静静地取消了。

课堂录像化是这条反噬最现成的例子：有了回放，教师不必再替听众把回头的活儿做掉，可以讲得更密、更连贯、更像文章；而真正需要那份预付的学生，恰恰是最不会去回放的那一批。**技术把平均体验改善了，把分布拉开了。**

第二种：把"预付冗余"当成啰嗦，一律砍掉。

格IV的那些套语、复诵、平行结构，在书面标准下全是冗余，很容易被当作低效率清理掉。清理的后果不会立刻出现——因为在平时，接收者的注意力足以补上。它会在**最坏的那一天**出现：噪声大、压力高、人疲惫，而那正是这些冗余当初被造出来的场合。

任何要精简标准通话、精简急救口令、精简交接班话术的改革，都应当先回答一个问题：**我打算把回头的代价挪到谁身上。**

第三种：用信道去替代教学。

如果不可回头的信道压制某些结构，那么一个很自然的教育推论是：多让孩子听、少让孩子读，或者反过来。两个方向都太快了。本文只说信道决定**什么结构能在这条信道里被接住**，没说它决定一个人**能长出什么能力**——这两件事之间还隔着大量本文没有处理的东西。

第四种：把它做成一条可以检查的规定。

"重要通知必须使用不超过一层悬置的句式"——这类规定会立刻长出来，然后被形式化地执行：句子被机械地切短，而信息被塞进更长的名词堆里，悬置没减，密度上去了。这正是第十二章那处让步的实践版本：**盯着句法层的指标去管，会把负担赶到不被管的那一层。**

第五种：本文自己会被优化。 留到第十八章，因为那不是别人的问题。

十六、怎样算它错

这一节是全文最要紧的部分。前面的论证如果没有它，就只是一个说得通的故事。

其一，决定性实验（现成范式，一个下午）。 用一批含多重中间套与等长往后接的句子，设四组，全部逐词呈现、速度由被试自己按键控制：

- 甲组，已呈现的词保留在屏幕上，可自由回视（格I）；
- 乙组，逐词呈现、看过即隐去，速度自控（"能停不能退"）；
- 丙组，听觉呈现、一次播完，不可重听（格III）；
- 丁组，听觉呈现，但可以按键要求重播任意片段。

结果变量：中间套与往后接两类句子的理解正确率之差。

- 本文预测：**甲 ≈ 丁 > 乙 ≈ 丙**，即分界线落在"能不能重取"上，而不落在"看还是听"上，也不落在"速度归谁"上。
- 瓶颈那一条预测：分界线随处理负荷走，乙组因速度自控应显著优于丙组。
- 媒介那一支预测：分界线落在看与听之间，即甲乙一侧、丙丁一侧。

三种预测把四个组切成三种不同的分法，这一个实验就能把三家分开。**若乙组显著优于丙组、且接近甲组，本文的核心变量选错了，第六、七章要重写。**

其二，只动接收侧（打产出端解释）。 同一批文本、同一位作者、同一种呈现速度，只变读者能否回退，测理解与偏好。产出端解释预测无差异（说的人没变）；本文预测中间套那一类出现显著差异。

其三，历史与语料侧的正向读数。 本文为真时应当出现一件可测的事：一种语言在获得可回头信道（书写系统普及、印刷、录音）之后，其正式语体中的中间套深度上限应当上升，且上升发生在书写普及之后而不是之前。这一条要求的是时间顺序，不只是相关。

其四，手语那一条（最容易打掉本文）。 若手语（不可回头，但走视觉）在悬置深度上表现得像书面语，那么本文把变量定在可回头性上就是错的，真正起作用的可能是感官通道。这条数据未必现成，但可做。

其五，会让本文尴尬的一条。 第十一章预测：飞行通话在引入可回放记录与文字数据链之后会缓慢变长变复杂。若历史材料显示它反而更简短、更程式化（因为规程收紧压过了信道松动），说明制度的力量在这一处压过了本文的变量，第八章那个"被信道逼出来的设计"应当改写为"被信道与规程共同定出来的设计"。

其六，两个明写交出去的洞。

第一个洞：我给不出"回头代价"的刻度。两端清楚，中间含糊，而第七章那根轴需要它。

第二个洞：我说不清可回头信道普及之后的那一步。中间套随书写出现，可这条链条上还有一环没接——是先有了可回头的信道，还是先有了想说那种话的人。本文默认前者，但没有证据能排除后者。

十七、为什么这件事一直没被看见

一个变量如果这么要紧，为什么四家都漏了？把这个问题答清楚，本文才算完整——否则最可能的解释是：它没那么要紧。

答案有三层，三层都不是疏忽。

第一层：在最常研究的那个场合里，两个变量恰好绑在一起。

面对面说话是语言研究的默认场景。而在这个场景里，"记忆转瞬即逝"与"材料不可重取"总是同时出现：声音说过去就没了，既超出记忆也拿不回来。两个原因永远同时在场，就不会有人想到把它们分开——这不是懒惰，这是共变。要把它们分开，你需要一个反常的场合：可回放的口语，或不可回看的文字。而这两样东西在人类历史上极晚才大量出现。

第二层：可回头性不是语言的属性，所以它不在语言学的变量表上。

一门学科的变量表由它的对象决定。语言学的对象是语言：音、词、句、篇。而"这段材料还能不能再取一次"不是语言的性质，是语言此刻所在的那条管子的性质。它更像纸的厚度、无线电的带宽、屏幕的刷新率——听上去像工程参数，不像语言学问题。

于是它落在了所有人的责任田之间：语言学家觉得那是技术条件，工程师觉得那是内容问题，心理学家把它折算成"呈现方式"这个控制变量——折算成控制变量的意思，就是把它固定住不去动它。而一个被所有人当作控制变量的东西，永远不会被当作自变量来测。

第三层，也是最要紧的：不可回头留下的痕迹，长得像"没有"。

如果一条信道压制了某种结构，那么在这条信道的语料里，你看到的不是"被压制的结构"，而是它压根不出现。零出现是一种极难解释的证据——它可以来自压制，也可以来自不需要、不习惯、不流行、样本不够大。

而一旦某个结构长期不出现，后来的人就会把它的不出现在读成"这门语言/这种文体本来就不用它"，进而读成一种风格特征，最后读成一种文化性格。从"信道不允许"到"我们文化就这样"，中间只隔着两三代人，而且每一步都不需要任何人说谎。

这一层解释还给出一条可查的推论：在那些同一群人同时使用可回头与不可回头两种信道的场合里，这件事应当早就被从业者知道。确实如此——播音编辑知道稿子要"写得能听懂"，法庭口译知道有些句子必须拆开

重排，飞行调度知道话要短到什么程度。他们全都知道，也全都说不出为什么，因为他们知道的是**做法**，而做法从来不会自己长成变量。

十八、本文自己也是标本

如果这条判断成立，它首先应当适用于你正在读的这一篇。

第一件要认的账：这篇文章重度依赖它自己所描述的那个特权。

数一数它对你提出的要求：第七章那张表要你在四个格子之间来回对照；第九章要你把"往后接"与"中间套"两个刚定义的东西记住并用到后面；第十三章要你同时握住六种别人的说法。这些要求只有在**你可以随时回去看一眼**的前提下才是合理的。

也就是说，本文是一篇**只能存在于格 I 的文章**。把它念出来——哪怕一个字不改、语速放到最慢——它的相当一部分会当场失效。而这不是它写得不好；如果它改到能被听懂，它就必须放弃第九章那种需要把两个概念悬着对照的写法，而那恰恰是它论证的关键一步。

这一点本身就是本文的一次自我检验，而且是可做的：把本文录成音频，找两组人，一组听、一组读，测他们对第九章那条推论的复述准确率。本文预测差距显著大于对第一章的复述差距。若差距相同，说明本文高估了自己所依赖的悬置。

第二件要认的账：我在写它的时候，一直在享受回头。

我可以把一句话写完再回去改前半句，可以把第十二章写完再回头把第六章的措辞收紧，可以把整篇读三遍。**任何一个只能一次成型的人都写不出这样一篇东西**——不是因为他不如我聪明，而是因为他所在的格子不允许这种结构。本文关于"信道决定什么能被造出来"的全部主张，在它自己的生产过程里已经演了一遍。

第三件，也是最不舒服的一件。

本文批评四家把信道当成背景，而本文自己也有一个被当成背景的东西：**它默认接收者是一个人**。今天大量的语言并不是给人读的——它被写进接口、喂给模型、由程序解析。对这些接收者来说，"可回头"是一个完全不同量级的东西：它们可以把整段材料原样存住、任意重取、并行比对。按本文的逻辑，**为这类接收者写的语言应当允许远比人类书面语更深的悬置**，而这件事正在发生，我却在这篇里几乎没有处理它。

我不打算假装这里有什么深意。我只是和我所描述的那四家一样，把此刻对我最方便的那个条件，当成了不需要说明的背景。

十九、结论，与一条写死日期的赌注

先把结论收拢成三句。

第一句：**关于语言被什么塑成，四种成熟解释对同一批现象给出互相取消的归因**——先天骨架、社会互动、记忆瓶颈、产出难易。四条都不是错的，但四条都把同一件事当成了背景：一段话说过去就没了，你没法请它重来。

第二句：**主变量是信道的可回头性，以及回头的代价由谁承担**。被这个约束筛掉的不是复杂结构，是**需要回头才能完成的结构**——这就是为什么一路往后接的长句在口语里活得好好的，而必须把成分悬在半空的中间套在口语里近乎不存在，且它随书面语的出现才被造出来。

第三句：**因此"口语简单、书面复杂"这个说法应当被拆掉。** 它把三个不同的格子（请求重说、独自受苦、预先垫付）混成了一个。真正可动的量有两个，都是被别的学科逼出来的：**需要悬置的结构，与材料的可重取性**——后者与记忆容量、与呈现速度、与感官通道，都是正交的。

剩下的事，是让它可以被杀死。

赌注：到 2031 年 12 月 31 日。

条件一：在此之前，第十六章第一项那个四组实验（同一批句子，看/听 × 可重取/不可重取交叉，速度均由被试自控）已被至少两个独立实验室做过。

条件二：结果可从公开渠道取得。

若两个条件成立，而**"逐词呈现、看过即隐、速度自控"那一组在中间套上的成绩显著优于听觉一次性呈现组、并接近可自由回视组——本文作废。** 不是需要修正，是作废：因为那意味着起作用的是接收者对时间的控制权，而不是材料的可重取性，第六、七章的整个骨架都建在错的地方。

再加一条更快到期的：**到 2028 年 12 月 31 日**，第十一章第九项那个最便宜的跨域检验——比较一个必须边听边定稿、不许回改的系统与一个可以看完完整段再输出的系统，在中间套与往后接上的选择差异——应当已经有人做出来。它不需要被试、不需要伦理审批、不需要经费，只需要两个现成的系统和一批句子。

如果三年之内连这个都没有人做，那不是本文的错，但也不是本文的功劳。它只说明这条判断还没有进入任何人真正会去检验的问题清单——而按本文自己的机制，这是可以解释的：**一个问题要被提出来，先得有人在它所在的那条管子外面站一会儿。而我们几乎所有关于语言的直觉，都是在管子里形成的。**

附：English abstract

Once It's Said, It's Gone: language is shaped not by poverty of the stimulus, social interaction, or memory capacity, but by whether the channel lets you go back

Abstract. Four mature accounts of what shapes language — innate structure, usage-based social cognition, the now-or-never processing bottleneck, and production-side distributional effects — issue mutually cancelling attributions on the same phenomena. This paper argues that all four treat as background what should be the variable: **the reversibility of the channel**, i.e. at what cost a receiver can request a replay. Crucially, reversibility is orthogonal to memory capacity, to presentation rate, and to sensory modality. The claim is selective rather than global: irreversible channels do not suppress complexity as such, but specifically those structures that require going back — which is why right-branching chains thrive in speech while multiple centre-embedding is virtually absent from it and appears historically only with writing. Crossing reversibility with who bears the cost of going back yields a 2×2 (self-served replay, requested repetition, suffering alone, prepaid redundancy) and two opposed orderings — suspension depth decreasing, redundancy increasing — that must both hold. Ten cross-domain redemptions are given; two domains force the claim to concede ground. A four-group experiment using an off-the-shelf paradigm separates this account from both the bottleneck and the medium account in a single run. A dated bet closes the paper.

Keywords reversibility of the channel; replay cost; suspended structure; centre-embedding; channel and register; irretrievability

参考文献

1. Chomsky, N. (1980). **Rules and Representations**. Columbia University Press. (刺激的贫乏)
2. Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. **The Linguistic Review**, 19(1-2), 9-50.
3. Tomasello, M. (2003). **Constructing a Language: A Usage-Based Theory of Language Acquisition**. Harvard University Press.
4. Tomasello, M. (1999). **The Cultural Origins of Human Cognition**. Harvard University Press. (共同注意与意图理解)
5. Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: a fundamental constraint on language. **Behavioral and Brain Sciences**, 39, e62.
6. MacDonald, M. C. (2016). Memory limitations and chunking are variable and cannot explain language structure. **Behavioral and Brain Sciences**, 39, e84. (同期评论)
7. MacDonald, M. C. (2013). How language production shapes language form and comprehension. **Frontiers in Psychology**, 4, 226. (PDC)
8. Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. **Journal of Linguistics**, 43(2), 365-392. (书面语最大深度为三；口语中多重中间套近乎不存在；随书面语出现)
9. Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. **Cognition**, 68(1), 1-76.
10. Ong, W. J. (1982). **Orality and Literacy: The Technologizing of the Word**. Methuen.
11. Halliday, M. A. K. (1985). **Spoken and Written Language**. Deakin University Press. (语法绵密度与词汇密度之分)
12. Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (ed.), **Spoken and Written Language**. Ablex.
13. Lord, A. B. (1960). **The Singer of Tales**. Harvard University Press. (口传程式)
14. Goody, J., & Watt, I. (1963). The consequences of literacy. **Comparative Studies in Society and History**, 5(3), 304-345.
15. Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. **Journal of Experimental Psychology: General**, 111(2), 228-238. (自定步速阅读范式)
16. Potter, M. C. (1984). Rapid serial visual presentation (RSVP): a method for studying language processing. In D. Kieras & M. Just (eds.), **New Methods in Reading Comprehension Research**. Erlbaum.

文献说明：本文对上述文献均为义释，不使用直接引文；卷期页码须按所用版本核校。第九章所用的语料读数（书面语中多重中间套最大深度为三、口语中近乎不存在、该结构随书面语出现）取自第 8 条；**把这些事实读成“信道可回头性的效应”是本文的推论，不是该文献的主张**，正文已注明。本文的核心实验**均未执行**，交出的是命题、判据与设计。

字数：正文约 1.75 万字 · 版本 v1.0 · 2026-08-01

人机分工声明：本篇由 Claude 依王德生的方法体系执行三家源材料的选取、冲突判定、占位检查、变量替换与全部行文；文中实证读数均引自公开文献，未做新的数据采集。

附：与本栏另外两篇的关系

本栏另有一篇《谁说了这一次到此为止》，讲的是练习里的同一类事——「一次」的边界由谁划，决定了什么东西被当作可比较的两次。那一篇管**边界**，本篇管**能不能回去**：一个说这段活动到哪儿算完，一个说完了之后还能不能再取一遍。另有《可预期的尺子》讲判据能否被提前推断如何决定一个领域还提得出什么问题——三篇合起来是同一个形状的三处：**决定长出什么的，往往不是内容，是内容所在的那条管子的性质。**

这一篇撞的三家，与可核对的出处

本篇撞的是站外三家理论，不是站内学员论文（同本栏之八、之九、之十的口径）。三家都在解释同一件事——孩子凭什么学会、语言又为什么长成这个形状——而给出的归因互相取消：一家说输入太穷所以必须有先天骨架，一家说输入其实足够只要把社会认知算进去，第三家说穷不穷根本不是重点、重点是它不等人。第三家内部还有一场没打完的仗，同期评论直接顶回去：记忆限制是可变的，解释不了结构那么整齐的规律。下面三条给出可直接点开核对的原始出处。正文第二至第四章逐一交代四家说到了哪一步，第五章把矛盾写死，第九章给出一条已经躺在语料里、四家都不预测其形状的读数，第十三章再与另外五种最容易被混为一谈的说法逐条分界——**其中第一位是本文承认最可能吸收掉自己的那一个。**

生成语法·乔姆斯基一脉

刺激的贫乏：孩子听到的话不足以支持他最终掌握的规则，多出来的那部分必须是先天的

论证的形状是：结论超出了证据，因此差额来自骨架。**与本文的分离线**：它把孩子听到的东西记成一个**集合**，而集合没有时间也没有次数——一句话听过之后能不能倒回去再听一遍，在集合里根本没有位置。

发展心理学与灵长类认知·托马塞洛

基于用法的习得：语言长在共同注意与意图理解上，语法是从大量具体用例里抽出来的

先有「我们此刻在共同关心同一件事」，才有指着某物说出一个词。**与本文的分离线**：它把互动与「没懂就再说一遍」捆成了一体，于是推不出这样一个问题——把共同注意与用例数量全部保持不变，只拿掉「再说一遍」，会怎样。

认知科学与语言演化·克里斯蒂安森与查特（2016）

The Now-or-Never Bottleneck：记忆转瞬即逝，语言必须当场被压缩与递交，否则永远失去

这条约束被推得极宽：处理、习得、历时演变乃至结构本身，都是它的下游。**与本文的分离线**：它把「记忆转瞬即逝」与「材料不可重取」当成了同一件事——在面对面说话时二者确实同时出现，但可以分开：给足记忆而不许回头，同一批结构照样长不出来。