基于大语言模型的
汉语主宾可逆句语义与施事程度评估

发言人：刘柱 （计算语言学方向）
导　师： 刘颖 教授

概要

- 背景（研究对象、性质、研究动机和研究贡献）
- 方法（数据收集、评估指标）
- 实验及分析
- 结论

问题背景

- 主宾可逆句：主语和宾语位置的成分交换，语义基本保持不变。
- 主宾不可逆句：交换之后，语义发生了（反向）变化。
- $s = AVO \quad s' = OVA \quad m$ 表示语义映射
- 可逆句I：$I = \{s \in S | m(s) = m(s')\}$

例如：十个人吃一顿饭 ⇒ 一顿饭吃十个人

- 不可逆句U: $U = \{s \in S | m(s) \neq m(s')\}$

例如：大鱼吃小鱼 ⇏ 小鱼吃大鱼

性质

- 对称性

  互逆操作前后的句子仍保持（不）可逆性，即仍处于同一个空间

$$\{s'|s \in S\} = S, where\ S \in \{I, U\}$$

- 方向性

  考虑到不可逆句中A具有明显的施事特征，O具有明显的受事特征，我们规定互为可逆的源句子：按句子的线性顺序，论元的施事性减弱

十个人吃一顿饭 ⇒ 一顿饭吃十个人

- 语言类型学和认知语言学的研究发现：A一般编码施事，O一般编码受事，当A和O的施事性或受事性发生变化的时候，A和O也会发生格配置变动

动机

- 大语言模型是从海量数据和超大规模中参数学习到的浓缩知识库，并在很多语言理解和生成任务上产生很好的效果
- 针对上述现象，评估大语言模型是否可以反映如下知识：

（1）大语言模型是否可以区分出可逆句和不可逆句的语义变化？

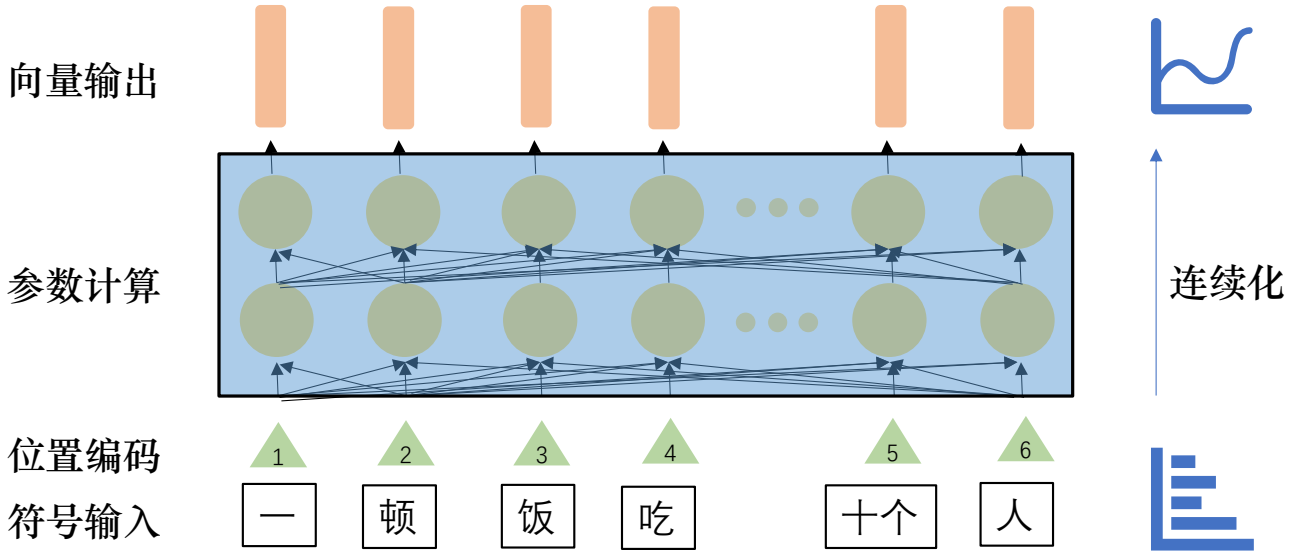
（2）大语言模型是否可以感知到不同情况的施事性程度差异？

研究贡献

- 本文形式化识别并定义两类互为对照的与汉语语序相关的两类句式，研究了它们的性质，以及定义了施事度评价指标。
- 收集了相关对比语料
- 在最新的计算语言模型上对它们的语义变化以及施事度进行了评估

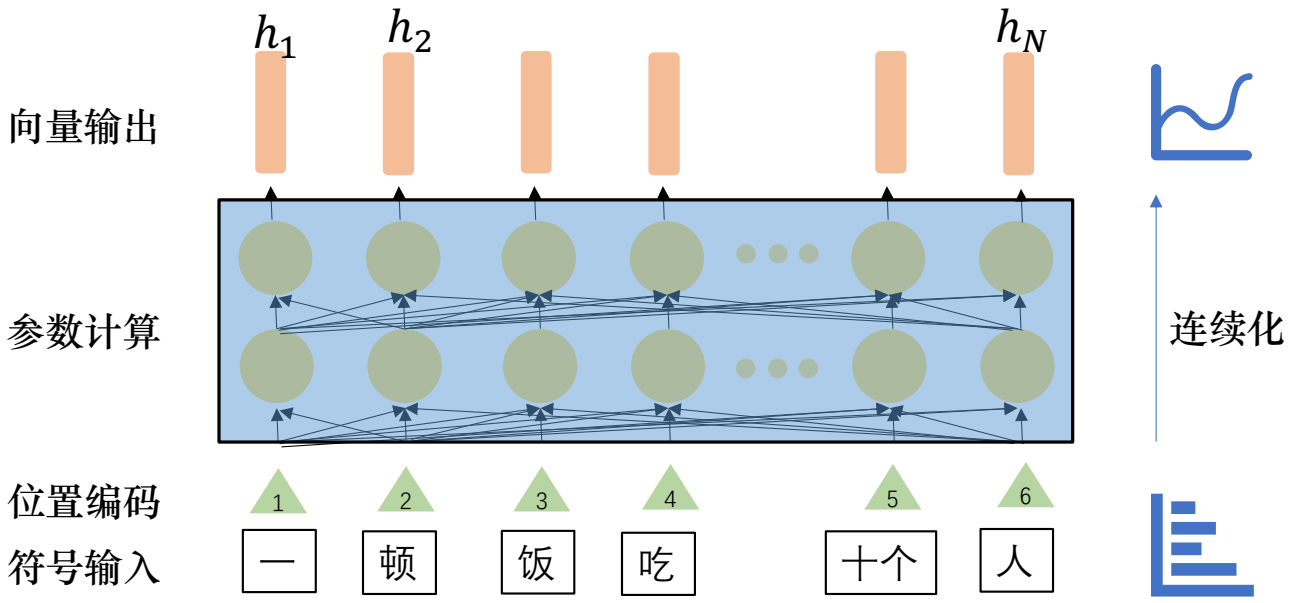
语言模型

- 目前性能最优的语言计算模型基于Transformer架构，句中的每一个单元（token）被映射为一个实值向量，该向量用来预测任务或者生成下一个单元。



语言模型 – 句向量表示

- 通常做法：词向量取平均 $\frac{1}{N}\sum_1^N h_i$
- 生成模型：最后一个词向量 h_N



数据收集

- 可逆句集*I*

参考已有研究[1]，从动词语义为切入点，共收集六种类型的可逆句，共41句

语义类型	可逆的源句
混合义	小葱拌豆腐
依附义	上面放好苹果
供给义	三个人住一个屋子
笼罩义	白雪覆盖着大地
进入义	三个女生又插进了我们班
充满义	乌云布满了天空

数据收集

- 不可逆句集U

选择词频较高的动词，内省造出20条。
为排除语法干扰，仅考虑语义，要求逆转之后仍合乎语法。
例如：一匹马抬两个人；老师打了学生；赵国游说秦国。
排除：我踩地板（*地板踩我）

- 对照集：

句内随机：句内字随机打乱
句间随机：不同句子进行配对

评估指标1：语义相似性

- 可逆句之间的相似性应该大于不可逆句的。

$$\frac{1}{N}\sum_{s\in I} Sim(e_s, e_s') > \frac{1}{M}\sum_{s\in U} Sim(e_s, e_s')$$

- 相似度评估指标：

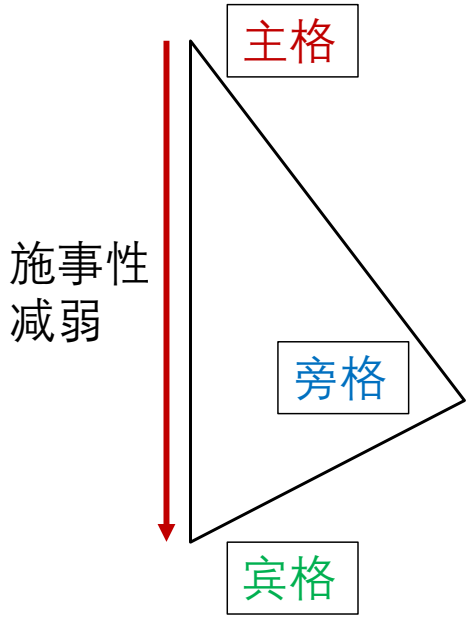
（1）作用于向量的余弦相似度

（2）标注可逆句相似度为1，不可逆句相似度为0，计算排序相关性。

评估指标2: 施事性

	施事	受事
自主性	+	-
引发性	+	-
受影响性	-	+

施事和受事的最大语义区分 (Naess,2007)



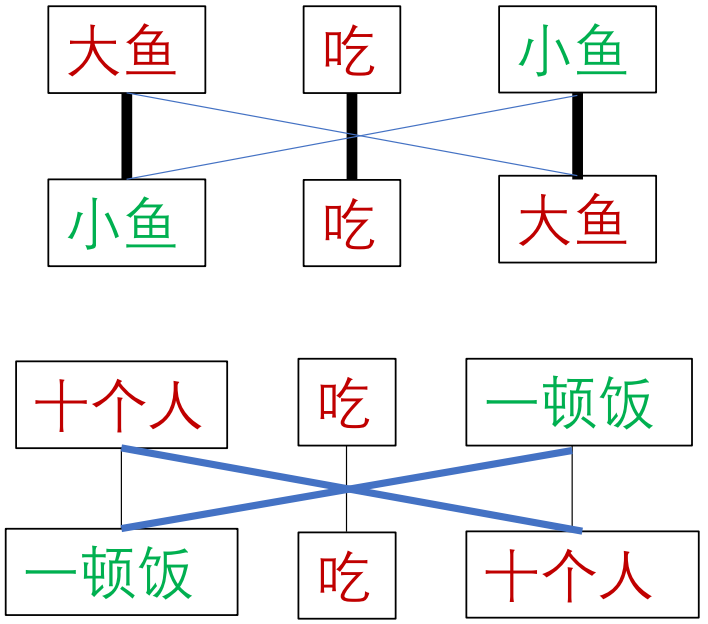
- 汉语的语序编码不同的语义角色，在典型的不可逆句中，前一个论元（主格，N1）表示施事，后一个论元（宾格, N2）表示受事

  例如：大鱼吃小鱼

在可逆句中，由于施受事不明显，主格也可能表示工具、或者受事；宾格可能表示施事。例如：一顿饭吃十个人；北屋住人

评估指标2: 施事性

- 断言 – 相比可逆句，不可逆句（从施事性角度）：
  对应位置（左左，右右，中）的论元相似度更高；
  相反位置（左右，右左）的论元相似度更低



评估指标2: 施事性

类型	不可逆成分	可逆成分	不可逆>可逆
左左	(大鱼，**小鱼**)	(一顿饭，**十个人**)	是
右右	(小鱼，**大鱼**)	(十个人，**一顿饭**)	是
左右	(大鱼，**大鱼**)	(十个人，**十个人**)	否
右左	(小鱼，**小鱼**)	(一顿饭，**一顿饭**)	否
中	(吃，**吃**)	(吃，**吃**)	是

Table 1: 不同位置类型在是否可逆的两种配置下的相似性程度差异以及举例说明

$$(-1)^{\mathbb{1}_{\{i\neq j\}}}\cdot[\frac{1}{N}\sum_{s\in I}Sim(\mathbf{e}_s^i,\mathbf{e}_s'^j)-\frac{1}{M}\sum_{s\in U}Sim(\mathbf{e}_s^i,\mathbf{e}_s'^j)]>0, i,j\in\{左,右,中\}$$

（施事度评分，值越大越好，I表示可逆句，U表示不可逆句）

实验设置

- 语言模型：

中文大语言模型：CPM-Bee；100亿参数；超万亿（trillion）高质量语料上进行预训练，目前在各大中文任务中表现优异；自回归

对比模型:

SBERT: 基于更小规模的双向BERT模型

CoSENT：针对句子相似任务做对比训练的高阶SBERT

- 数据：可逆句集、不可逆句集、随机对照集

实验一

- 目标：可逆集变化前后的语义是否更相似？
- 结论：(1) 使用通用的方式获得的句向量没办法区分这两种情况☹
  （2）大模型在不可逆句的相似性反而更高（获取方式？更看重形式？）
  （3）传统模型对于句内的顺序更加不敏感（句内随机）☹

模型	可逆句	不可逆句	句内随机	句间随机	相关性	其他数据结果
SBERT	93.2	92.0	88.5	71.2	43.0	-
CoSENT	96.0	93.8	91.7	56.2	**51.1**	63.1
CPM(M)	79.4	91.3	58.9	26.6	41.1	-
CPM(L)	46.4	60.0	41.3	25.0	26.3	-

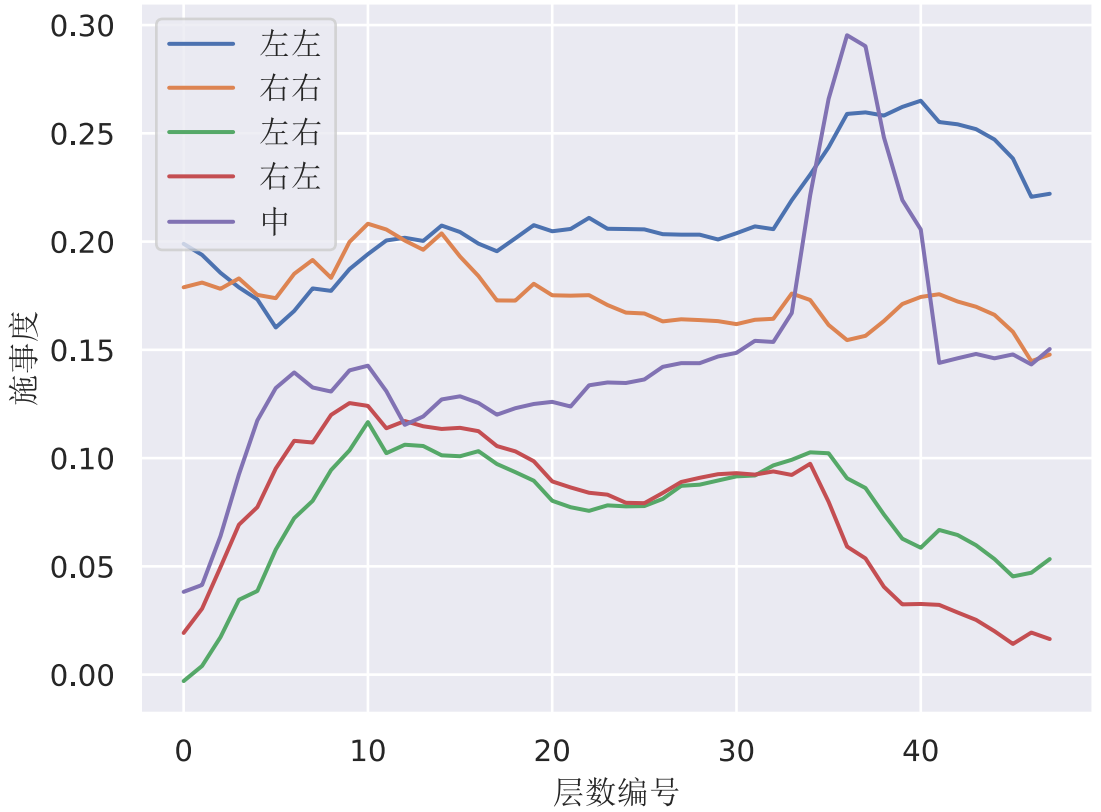
Table 2: 不同模型在不同类型的数据下句子对之间的平均相似度

实验二

- 不同位置类型下，施事性在两组句子内的表现？

由于大语言模型可能受到层数变化，先找到每一层上的最佳体现

- 施事度都大于0，表明各层都可以反映施事性这一特征。
- 变化并不单调，不一定最后一层最好

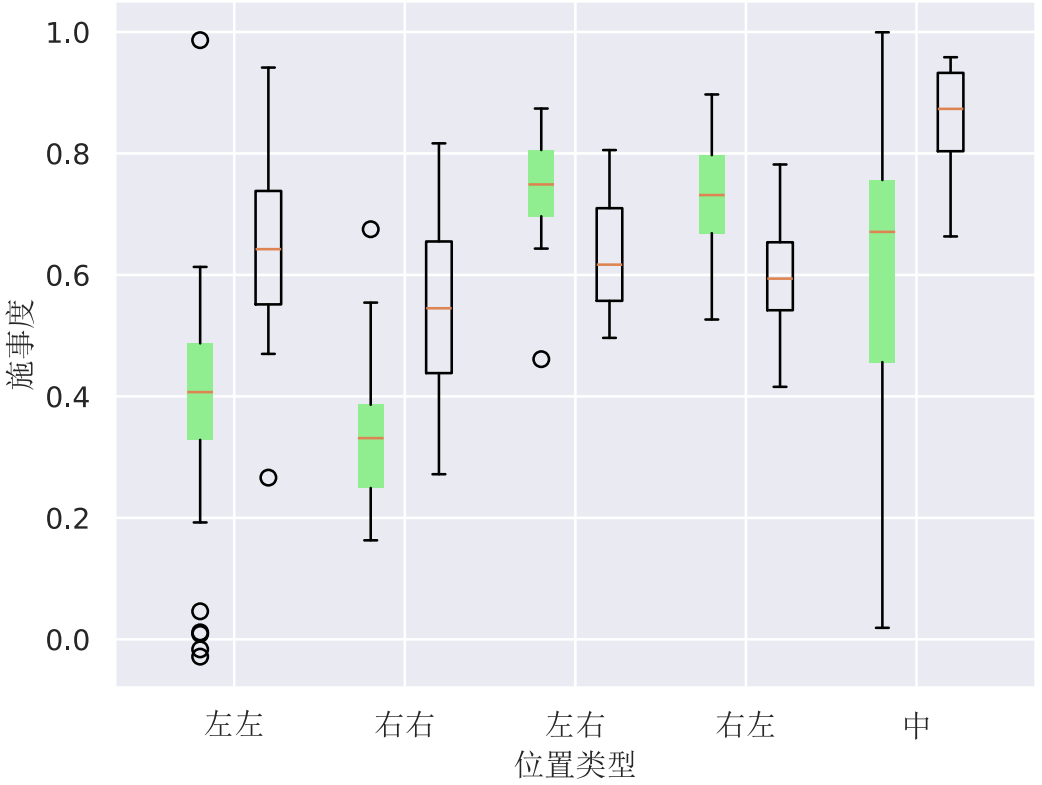


实验三

- 每种类型在最佳情况下，施事度的分布是怎么样的？
- 均值具有显著差异
- 模型可以很好反映不同位置的施事度

位置类型(%)	可逆句	不可逆句	施事度	p值
左左	38.0	64.5	26.5	5.54e-6
右右	33.5	54.3	20.8	1.86e-6
左右	74.6	63.0	11.6	1.03e-5
右左	72.5	59.9	12.6	1.84e-5
中	56.7	86.2	29.5	3e-4

Table 3: 五种位置类型下两种句型的施事度比较



（绿色表示可逆句，白色表示不可逆句）

案例分析

以动词分析为例

- 动词性施事性强的是一些情感词汇（不可逆句），这些词汇的受事一般不能做主语；
- 动词施事性弱的是一些“覆盖义”动词，这些词汇往往不能表达较强的施受性（对比：“吃”）

不可逆句	可逆句
我想念我的朋友们	桌子蒙着一块花布
总理爱人民	太阳晒着稻草堆
下属害怕领导	鲜花开遍了原野
小蝌蚪找妈妈	空格里签名字
学生听老师	茄子炒肉

Table 4: 动词相似性前五（不可逆）和后五（可逆）的源句

结论

- 本文分析了大语言模型对于可逆句和不可逆句的语义变化的不同感知，发现传统提取句向量的方式不能很好反映它们的语义差异。
- 之后进一步分析大模型对于汉语不同语序位置的施事度的反映，发现它可以捕捉到这一性质，可能具备了区分二者的潜能。

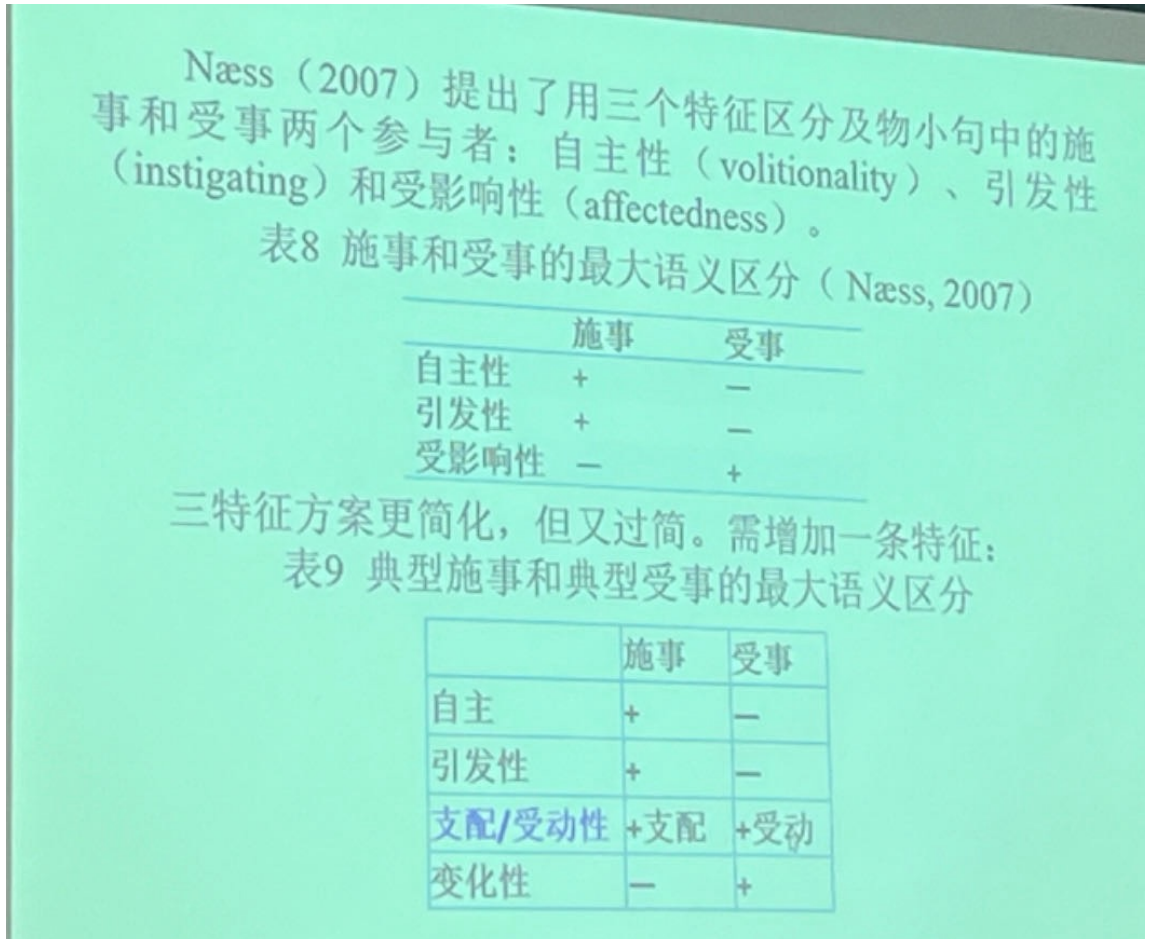
未来工作：

- 收集更多数据、设计更多的任务来全面反映大模型对它们的理解。
- 研究更多其他语言相似的情况（格配置变动）。

Q & A

谢谢大家聆听！

施事性和受事性强弱的构成特征



Næss（2007）提出了用三个特征区分及物小句中的施事和受事两个参与者：自主性（volitionality）、引发性（instigating）和受影响性（affectedness）。

表8 施事和受事的最大语义区分（Næss, 2007）

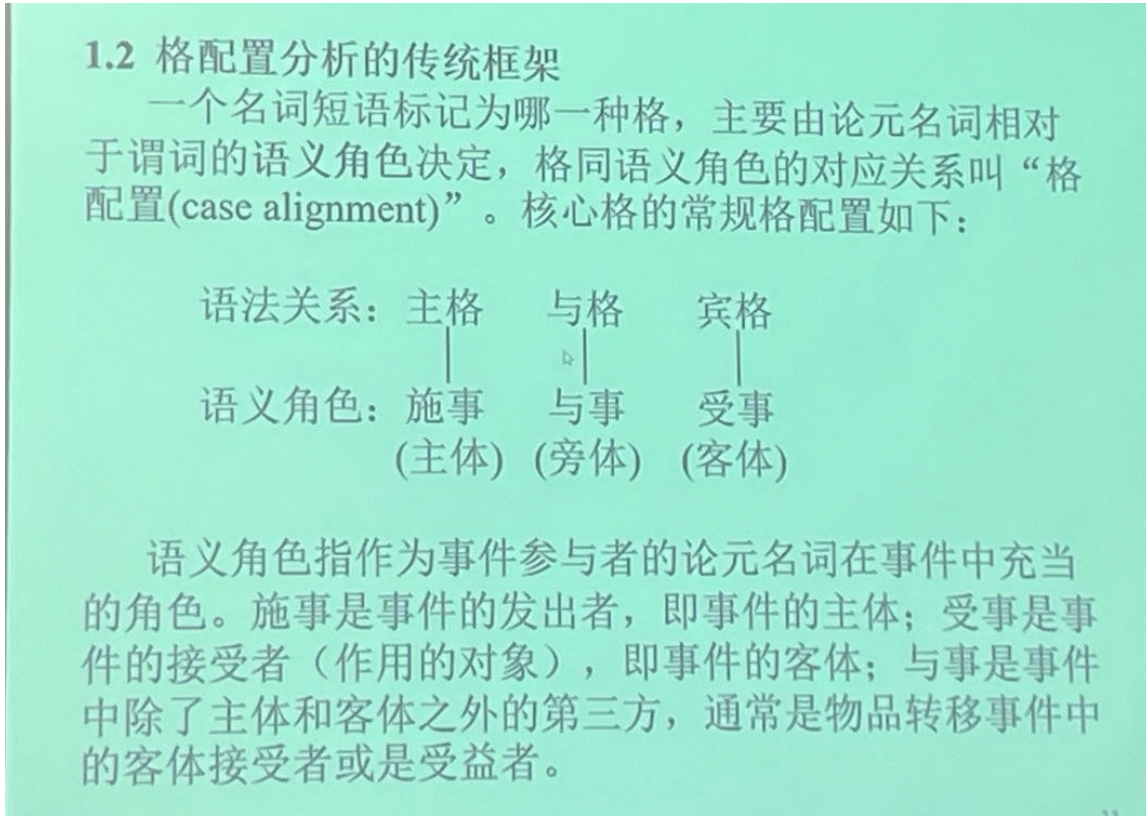
	施事	受事
自主性	+	—
引发性	+	—
受影响性	—	+

三特征方案更简化，但又过简。需增加一条特征：

表9 典型施事和典型受事的最大语义区分

	施事	受事
自主	+	—
引发性	+	—
支配/受动性	+支配	+受动
变化性	—	+

格配置的常规分析框架



1.2 格配置分析的传统框架

一个名词短语标记为哪一种格，主要由论元名词相对于谓词的语义角色决定，格同语义角色的对应关系叫“格配置(case alignment)”。核心格的常规格配置如下：

语法关系：	主格	与格	宾格
语义角色：	施事	与事	受事
	(主体)	(旁体)	(客体)

语义角色指作为事件参与者的论元名词在事件中充当的角色。施事是事件的发出者，即事件的主体；受事是事件的接受者（作用的对象），即事件的客体；与事是事件中除了主体和客体之外的第三方，通常是物品转移事件中的客体接受者或是受益者。

可逆句	不可逆句
两份水泥配一份沙子	一匹马抬两个人
鱼头炖豆腐	两个人搬一张沙发
白菜熬粉条	两个人抬一张桌子
鸡蛋炒黄瓜	总理爱人民
小葱拌豆腐	张三追累了李四
茄子炒肉	秦国打败了燕国
好苹果放上面	老师打了学生
菜摆桌子上	那个女生笑话他
粮食堆仓库里	学生听老师
像片贴右上角	赵国游说秦国
名字签空格里	大鱼吃小鱼
口袋缝左边	我想念我的朋友们
玉米种前院	下属害怕领导
地瓜种后院	他刚失去了她
三个人住一个屋子	后排的人猛推前面的人
两个人骑一匹马	他拉了后面的人
五个人坐一条板凳	小蝌蚪找妈妈
两个人睡一张床	同桌踢了他
七个人吃一顿饭	哥哥保护弟弟
好几个人洗一盆水	家长批评孩子
大楼笼罩着晨雾	
大地覆盖着白雪	
山谷弥漫着烟雾	
稻草堆晒着太阳	
汽车盖着油布	
桌子蒙着一块花布	
天空布满了乌云	
原野开遍了鲜花	
前沿布满了地雷	
大地洒满雪花	
天空飞满了树叶	
街头聚满了人群	
商场挤满了顾客	
屋里已经进了不少水	
我们班又插进了三个女生	
游行队伍里夹进了一个便衣警察	
暗房里透进一线光亮儿	
他的血管里输进了二百毫升的人造血浆	
人造血浆输进了他的血管里	

Table 5: 可逆句和不可逆句示例