

广义话题结构理论视角下话题自足句成句性研究*

尚英¹, 宋柔², 卢达威²

1.北京语言大学 预科教育学院, 北京, 100083

2.北京语言大学 语言信息处理研究所, 北京, 100083

摘要: 话题自足句是在广义话题结构理论的基础上定义的。话题自足句的成句性是广义话题结构的重要性质之一。本文在 38 万字不同语体的广义话题结构语料库中对话题自足句的成句性进行了实证性调查, 发现有少量话题自足句不成句, 对不成句现象进行了分析、分类, 并提出了使其成句的办法。这将进一步完善广义话题结构理论, 并能提高使用话题自足句的应用系统的性能。

关键词: 广义话题结构; 话题自足句; 成句性

中图分类号: TP391

文献标识码: A

Study on the Sentence-Formability of Topic Sufficient Sentence under the Perspective of Generalized Topic Structure Theory

Shang Ying¹, Song Rou², Lu Dawei²

1. College of Preparatory Education, Beijing Language and Culture University, Beijing, 100083, China

2. Center of Language Information Processing, Beijing Language and Culture University, Beijing, 100083, China

Abstract: The Topic Sufficient Sentence (TSS) is defined on the basis of Generalized Topic Structure (GTS) theory. The sentence-formability of TSS is one of the important features of GTS. Based on the corpus of large scale and different language styles, this paper focuses on the sentence-formability of TSS. We find that a small amount of TSS can't form sentences. By analyzing and classifying those sentences, we present some methods to ensure the sentence-formability of TSS. This not only provides a way to make GTS theory perfect, but also improves the performance of application systems that use TSS.

Keywords: Generalized Topic Structure, Topic Sufficient Sentence, Sentence-Formability

1. 引言

话题是当前汉语本体研究和语言信息处理研究中的热点。赵元任(1979)指出“把主语、谓语当做话题和说明来看待, 较为合适。”朱德熙(1982)继承了这种观点。Li & Thompson(1976)从类型学的角度指出汉语属话题凸显的语言, 后又对句子的话题和主语做了系统对比。徐烈炯, 刘丹青(1998)认为汉语中话题被高度语法化, 成了一种句法成分。曹逢甫(1979, 1990)在汉语篇章层面上提出了“主题链”概念; 沈开木(1996)、屈承熹(1998)、许余龙(2004)等也在篇章层面对话题进行了研究。吕叔湘(1979)、范继淹(1986)、沈家煊(2012)发展了赵元任的零句说, 关注了汉语流水句现象, 同时对汉语话题也有深入研究; 陈平(1987)从零形回指的角度探讨了汉语话题推进情况。

话题研究在理论探索过程中取得了丰富的成果, 综观这些成果, 可以看到: 目前关于汉语话题的研究在理论上进行了多方面探讨, 由于不以应用为直接目标, 所得结论和规则还不能直接用于信息处理等领域。比如关于“话题”的论述中的最基本的单位——汉语的句子, 有的人进行了定义, 而有的人没有。依据现有的关于汉语“句子”的定义, 在实际语料中仍然难以清晰地把握“句子”的边界。此外, 上述成果在观点论证过程中多采用举例的方法, 少见大规模真实语料的系统性调查分析。目前研究多引口语和叙述体小说中的实例, 这在一定

***基金项目:** 国家自然科学基金资助项目(项目编号 61171129), 北京语言大学校级科研项目(中央高校基本科研业务专项资金资助)(项目编号 13YBG30)。

程度上束缚了研究范围，将视野放至多语体将更有助于汉语话题及话题结构的深入研究。

宋柔（2013）面向语言工程的需要，以标点句为基本单位，提出了广义话题结构的观念和表示方法，归纳出了广义话题结构遵从的堆栈模型并拓展为流水模型，指出广义话题结构具有两个重要性质：话题的不可穿越性和话题自足句的成句性。本文在 38 万字不同语体的广义话题结构语料库中，基于人工标注对话题自足句的成句性进行了实证性调查和研究。

2. 话题自足句及其成句性

话题自足句是在广义话题结构理论的基础上定义的。汉语篇章中的每个标点句都对应一个话题自足句。如果标点句不共享其他标点句中成分，那么它的话题自足句就是它本身。否则，设标点句 c 在广义话题结构 S 中，S 是根据流水模型理论表示的，把 c 在 S 中的话题逐个地按原顺序补充到 c 的左端，把 c 在 S 中的说明或说明尾部逐个地按原顺序补充到 c 的右端，最后得到的就是 c 在 S 中的话题自足句。“自足”的意思是，c 在字面上能找到的话题都找到了，都集中在这个“句子”中。如：

例 1

普通八哥较常见，通体黑色，嘴基上羽额耸立，形成羽幘；头顶、颊、枕和耳羽具绿色金属光泽；

我们用下表显示例中标点句与话题自足句之间的对应关系，左栏是标点句换行缩进图示，右栏是每个标点句对应的话题自足句，其中标下划线的部分是补上的话题。

表 1 例 1 的标点句与话题自足句的对应

| 标点句 | 话题自足句 |
|----------------------------------|---|
| c ₁ 普通八哥较常见， | S ₁ <u>普通八哥较常见，</u> |
| c ₂ 通体黑色， | S ₂ <u>普通八哥通体黑色，</u> |
| c ₃ 嘴基上羽额耸立， | S ₃ <u>普通八哥嘴基上羽额耸立，</u> |
| c ₄ 形成羽幘； | S ₄ <u>普通八哥嘴基上羽额形成羽幘；</u> |
| c ₅ 头顶、颊、枕和耳羽具绿色金属光泽； | S ₅ <u>普通八哥头顶、颊、枕和耳羽具绿色金属光泽；</u> |

话题自足句具有成句性，这是从话题和篇章的角度说的。任何自然语言中“句子”的成句性都是无法严格定义的，汉语缺少形式标记，更无法严格定义，区分是否成句的操作基础只能是语感。这种语感的大致的意思是：语法通顺并且语义清楚正确。反过来说，不成句的话题自足句或者是语法不通，或者是语义不清甚至语义错乱。

赵元任、朱德熙、沈家煊等人指出，汉语的主谓关系就是话题-说明关系。因此可以说汉语中结构完整的“成句”的单位，具有话题-说明关系。那么，反过来，一对具有话题-说明关系的词语串，中间不加标点，是否能连成句呢？

从理论上说，这个问题的答案是否定的。因为我们谈的是广义话题结构，包括状中关系、述宾关系等结构，它们句法成分不完整，通常会导致语义不清。当然，从实际调查中看，广义话题结构中大多数的话题-说明关系，从句法结构的角度看就是主谓关系，这对于成句性是一种正面的保障。

语料库的调查说明，绝大多数话题自足句都是成句的。上例的话题结构中用堆栈模型生成话题自足句 S₁—S₅ 都是成句的。再看下面的例子：

例 2

爱因斯坦的主要著作已编入 || 《爱因斯坦文集》，

中文本共 3 卷，

由许良英、赵中立、范岱年、李宝恒、张宣三编译，

这个话题结构用节栈模型生成话题自足句 S₁—S₃ 如下，都是成句的：

S₁: 爱因斯坦的主要著作已编入《爱因斯坦文集》，

S₂: 《爱因斯坦文集》中文本共 3 卷，

S₃:《爱因斯坦文集》中文本由许良英、赵中立、范岱年、李宝恒、张宣三编译,

例 3

『制定一个跨世纪的宏伟纲领,
继续把改革开放和社会主义现代化建设事业推向前进,』

是全国人民关心的大事。

这是汇流模型的话题结构,标记『』括起来的标点句,其说明部分在括号后面。用汇流模型生成话题自足句如下:

S₁: 制定一个跨世纪的宏伟纲领是全国人民关心的大事。

S₂: 继续把改革开放和社会主义现代化建设事业推向前进是全国人民关心的大事。

S₁、S₂ 都是成句的。

例 4

吃完晚饭,
邢老汉叼着烟锅想说点什么,

这是话题后置的话题结构,用逆向堆栈模型生成话题自足句:

S₁: 邢老汉吃完晚饭,

S₂: 邢老汉叼着烟锅想说点什么,

S₁、S₂ 都是成句的。

由上文话题自足句的生成过程可以看出,每个话题自足句只带一个说明。如果其中的话题在标点句序列中还带有其他说明,那些说明就被割除了。毕竟少了信息,且去除了标点句间的停顿,于是可能发生语法不通、语义不清甚至错乱的情况,使得少数话题自足句不成句。

3. 话题自足句及其成句性研究的意义

研究话题自足句及其成句性有重要意义。

首先,汉语中没有天然的“句子”。在现实的汉语篇章中,只能看到标点句。基于广义话题理论的话题自足句可看作抽象的“句子”在汉语中存在的现实形态。这种句子同其他语言中的“句子”处于同一语法层面。汉语学界对于理想的“单句”已经进行了长期的研究,积累了大量成果。于是,把话题自足句当作单句,当作从标点句到篇章的中间层次,是一种理论上和实践上都比较理想的篇章语法体系构建方法。

第二,对于语言信息处理来说,把话题自足句当作从标点句到篇章的中间层次也是比较适当的。计算语言学对于单句的分析已经比较成熟。话题自足句相当于单句,其中积聚了比较完整的信息,对于机器翻译、信息提取、自动文摘等应用具有现实的价值。话题自足句的成句性则保证了单句分析的成果能够运用到话题自足句上来。

第三,汉语篇章中约一半的标点句需要从广义话题结构中共享话题信息。听者/读者在听/读的过程中,接收的只是标点句。把标点句所缺少的信息补充完整,应当是人的认知过程不可或缺的一步。话题自足句具有如下性质:信息比较完整;保全了标点句的词序,而词序对于汉语的重要性是大家的共识;生成过程足够简单,可以适应人脑实时处理标点句的需要。由于话题自足句的这些性质,我们虽然不能说人在这一步认知过程结束时,大脑中存储的认知结果的形式一定是一个话题自足句,但可以猜想这种认知结果的形式同话题自足句具有相当大的相似性。因此,话题自足句的研究对于探索人的语言认知过程具有重要意义。

第四,我们特别关注话题自足句中少数不成句的现象,详尽地调查和发现这类实例,归纳这种现象的特征,研究使其成句的方法。这样做的目的有四个。一是通过话题自足句成句性的视角,发掘汉语中过去未曾注意到的语言现象,推进汉语的研究。二是基于这些特征,想办法通过完善广义话题结构的模型,改进话题自足句的构造方法,提高话题自足句的成句率。三是基于这些特征,剔除不成句的话题自足句,从而提高使用话题自足句的应用系统的性能。四是发现话题自足句与人的认知结果在形式上的差异,进一步探索人的认知机制。

4. 话题自足句不成句现象的分类研究

4.1. 句法不通

4.1.1. 篇章功能成分

有的篇章功能成分用在话题与说明之间，导致话题与说明相连以后不成句。可分四类：

4.1.1.1. 句间连词

有些句间连词如“但是”、“并且”“然后”等，可以用在用作说明的独立标点句之首，可以用在两个具有相应逻辑关系的说明之间，但不能用在话题与说明之间。当用作说明的标点句句首有这种成分时，话题与说明相连所形成的话题自足句就不成句。

例 5

他房子里只有一铺炕和两个旧得发黑的木板箱，

但收拾得倒很干净。

“但”是表示转折的句间连词，它可以出现在标点句之首，包括话题前和说明前（如本例那样），可以出现在两个具有转折关系的说明之间，如“他数学很好但语文很糟”，但不能出现在话题与说明之间。当它出现在用作说明的标点句的首部时，如果把话题和说明部分连成一个话题自足句，这个问题就暴露出来了。“他房子里但收拾得倒很干净。”中“但”处在句子主谓关系的话题与说明之间，导致了话题自足句句法不通。为解决这个问题，可以修改广义话题结构理论中话题自足句的处理方法，在对话题自足句作句法分析时临时把连词“但”删除，待进行话题自足句之间的逻辑关系分析时，再考虑“但”的功能。又如：

例 6

1976年4月，

他再次失去一切职务，

并受到错误批判。

“并”是表示递进的连词，它可以出现在标点句之首，也可以用在两个具有递进意义的说明之间，但不能出现在话题与说明之间。本例中话题和说明部分连成一个话题自足句“1976年4月他并受到错误批判。”，“并”处在句子主谓关系的话题与说明之间，导致了话题自足句句法不通。

这类连词是一个封闭的集合，在话题自足句生成过程中如果遇到这些连词，正常顺序下不成句，可以通过机械的办法将其调整成句。

4.1.1.2. 代词

有些代词可以离开上下文单独使用，如“你、我、他”等，有的代词离开上文语境便不能独立使用，如“其、其余、前者”等。它们也是可以出现在标点句首，但不能出现在话题后面。否则，一旦把话题和说明连成话题自足句，就会导致语法不通。

例 7

大黄鱼是鲈形目石首鱼科黄鱼属的1种。

……

体长椭圆形，

侧扁。

尾柄细长，

其长为其高的3倍多；

例中最后一个标点句，话题与说明连接成话题自足句“大黄鱼尾柄其长为其高的3倍多”句法不通，原因是代词“其”虽然能作主语，但不能出现在它指代的名词的后面紧邻处，因

而生成话题自足句时，不能出现在话题和说明之间。

还有一些篇章成分，由于在语料中出现次数较少，在此略去。

4.1.2. 标点句中宾语成为新支话题

标点句中的宾语被后面的标点句说明形成新支话题结构，此时新支话题作为原标点句中的宾语，往往是不定指的，但是在生成的话题自足句中是主语，通常要求定指，否则句法错误。为此，要将不定指变成定指，办法是加定指代词“这、这些”等。如果原来新支话题以“一”加量词的数量结构开始，其中的数词“一”往往要删除。

例 8

那都是 || 些愣头愣脑的小伙子。

前天把一个卖瓜子的捆了一绳子，
昨天又把一个木匠的家伙收了，
害得人连哭带嚎。

例中新支话题是“些愣头愣脑的小伙子”，与说明“前天把一个卖瓜子的捆了绳子，”“昨天又把一个木匠的家伙收了，”连接成话题自足句时，须在话题前加定指代词“这”，使得话题自足句成为“这些愣头愣脑的小伙子前天把一个卖瓜子的捆了一绳子，”“这些愣头愣脑的小伙子昨天又把一个木匠的家伙收了，”“这些愣头愣脑的小伙子昨天又害得人连哭带嚎。”

例 9

那老汉思忖了几锅子烟的工夫，

思忖出了 || 一个主意，

就是给狗求得一个官方保护。

上例中的新支话题“一个主意”，是不定指的，与说明“就是给狗求得一个官方保护”连接成话题自足句“一个主意就是给狗求得一个官方保护”时，所表示的语义成了“主意之一”。如果要保持原意又要语法通顺，就须在话题前加定指代词“这”，同时需要删除数词“一”，成为“这个主意就是给狗求得一个官方保护。”

语料中我们还发现两例新支话题结构因文言而导致话题与说明连接不成句的情况，此处不详述。

4.2. 语义模糊不清或语义错乱

首先我们将话题自足句语义问题分为三大类：一类是首部语义缺失，归为主语省略句；第二类话题自足句主谓齐全，但语义不清楚；第三类是话题自足句主谓齐全，能够独立表达清楚的语义，但是独立表达的语义与其在标点句序列中的语义不一致。

主语省略句情况复杂，受篇幅所限，需另文详述，这里只讨论其它两类。

4.2.1. 语义不清

话题所在句(准确地说是话题所在的标点句所对应的话题自足句。这里和下面的例子中，为了避免行文繁杂，就直接说话题所在句)中的旧对象，与说明标点句中的新对象，在标点句序列中的关系是清晰的，当中间的标点句被隔离，话题与说明直接相连成为话题自足句时，这种成分之间的语义关系可能模糊不清，需要添加虚性词显性地表示它们的关系。从语料调查中的情况看，引起语义不清的成分，其语义角色主要有以下几种：

4.2.1.1. 时间

例 10

慈云桂是中国计算机专家。

……

中华人民共和国成立后，

任副教授、教授、系主任、计算机研究所所长、……，
长期从事教学和研究工作。

例中“中华人民共和国成立后”一句是事件性时间，同时它带有两个说明需要形成内层话题结构，是外层话题“慈云桂”的说明，直接生成话题自足句为：

S₁：慈云桂中华人民共和国成立后任副教授、教授、系主任……，

S₂：慈云桂中华人民共和国成立后长期从事教学和研究工作，

S₁、S₂的语义结构关系是{主体₁+时间+行为}。

主体和行为之间插入时间成分时，时间成分通常是简单的时间名词、时间名词短语，如：

我昨天晚上急得一夜没睡着。

马西莫夫2006年出任哈政府副总理。

他们黄昏前到达了另一个村镇。

当时间成分为事件性时间时，通常该事件行为与时间成分后面的行为均由同一主体发出或者是模糊的主体，如：

他毕业后回了上海。

我们开会前先收拾一下会议室。

事件性时间“毕业后”、“开会前”中的行为“毕业”、“开会”的主体有两种可能，一种是与时间成分后面的行为主体统一，前一例都是“他”，后一例是“我们”；另一种可能是“毕业”、“开会”的主体是模糊的，不特别指明。

事件性时间{主体₂+行为+时间方位词}中出现了新的主体即主体₂，则两个主体之间需要插入显性的标记，表示主体₂是时间成分的开始，整个话题自足句可以分析为{主体₁+{主体₂+行为+时间方位词}+行为}；否则主体₁与主体₂关系不清楚，导致整个句子语义模糊。上例中的S₁、S₂即是这种情况。

这种情况下，需要在主体₂前加形式标记，本例中可以加介词“在”：

S₁：慈云桂在中华人民共和国成立后任副教授、教授、系主任……，

S₂：慈云桂在中华人民共和国成立后长期从事教学和研究工作，

4.2.1.2. 处所

新旧对象之间存在主体与处所、处所与状态的关系，需要通过加词的方式将处所成分显性表示出来。

例 11

c₁:斑鲛是鲱形目鲱科斑鲛属的 1 种，

……

c₂: 分布于印度到东印度群岛和朝鲜及日本南部，

c₃: 中国沿海均有分布。

上例话题为“斑鲛”为已有对象，与标点句 c₃中“中国沿海”是主体与处所的关系，可以在处所成分前加“在”，对应的话题自足句是：

斑鲛在中国沿海均有分布。

4.2.1.3. 存在物、部件、可用设施或拥有物

话题与跨标点句的说明中的对象是主体与存在物、部件、可用设施、拥有物等关系时，常需要在对象前加“有”表示主体和对象之间的关系。

例 12

c₁:带鱼是鲈形目带鱼科带鱼属的 1 种（见图带鱼外形）。

……

c₂: 卵径约为 1.4 ~ 2.2 毫米，

c₃: 油球一个，

c₄: 油球径为 0.4 ~ 0.5 毫米；

标点句 c₃中的“油球”是卵的一个部件，需要在前面插入“有”，得到话题自足句“带

鱼卵有油球一个，”使其关系明确，语义清楚。

4.2.1.4. 类属与特征

话题与跨标点说明中的新对象是种属或主系关系，需要在种属、主系间加系词“为/是”。

例 13

角鮟鱇亚目无腹鳍，
无假鳃，
副蝶骨与额骨不连接，
体侧扁，
深海鱼类，

最后一个标点句“深海鱼类”是话题“角鮟鱇亚目”的父类，如果直接生成话题自足句为“角鮟鱇亚目深海鱼类”，不自然，中间加上“为/是”生成话题自足句“角鮟鱇亚目为(是)深海鱼类，”。

4.2.1.5. 动词论元

汉语中二价动词，如果施事 S 和受事 O 同时出现，则可能的顺序有 3 种：SVO，SOV，OSV。如：

- | | |
|---------|----------|
| ①我喝牛奶了。 | ④小华认识小刚。 |
| ②我牛奶喝了。 | ⑤小华小刚认识。 |
| ③牛奶我喝了。 | ⑥小刚小华认识。 |

①—③三个句子，尽管语序不同，但表达了同样的语义关系，两个论元中，“我”是“喝”的施事，“牛奶”是“喝”的受事。我们认为之所以上例中 O 与 S 位置可以灵活出现，是因为二者的词汇语义决定了与 V 的语义关系，否则它们的顺序调整会受限，往往需要其他手段予以辅助，以便理解。如：

④—⑥三个句子中只有④的语义是清晰的，⑤⑥的语义不清楚，因为“小刚”与“小华”对于“认识”的语义关系不确定，如果将二者放在一起，尽管从语序上有前后之分，但还是难以理解。

在广义话题结构中，有时候话题与跨标点句说明连接成话题自足句时，有些 OSV 或者 SOV 的语序不能表达原有的语义或者存在歧义，这时我们为了保证语义的清晰，需要对话题自足句中的相关成分进行调序。

例 14

白朗军长驱入陕，
攻占商南等地，
一度逼近西安。
又沿渭水西进，
由陇县进入甘肃省境。
陕西一带哥老会纷纷归附，

在上例标点句序列中，最后一个标点句表达的语义是“陕西一带哥老会纷纷归附白朗军”。但在堆栈模型下生成的话题自足句为“白朗军陕西一带哥老会纷纷归附”，这个句子本身可接受度差；另外如果没有足够的历史知识，可能会将“陕西一带哥老会”与“白朗军”当作隶属关系，可见这种表达还会影响到语义的理解，因此需要调序。

调序超出了流水模型现有的功能范围，所以这类现象列在现有体系无法覆盖的范围内。

4.2.1.6. 复杂关系要素不全

复杂关系句所述的关系涉及多个要素，其中有些要素在上文提到，下文往往省略。话题自足句脱离了广义话题结构的语境，是孤立的，在被省略成分不能作为话题被共享的情况下，话题自足句可能发生语义不清的问题。

例 15

- c₁: 2 0 0 4 年,
c₂: 国内生产总值达到 1 3 . 6 5 万亿元,
c₃: 比上年增长 9 . 5 %;
c₄: 财政收入 2 . 6 3 万亿元,
c₅: 增长 2 1 . 4 %;
c₆: 社会消费品零售总额 5 . 4 万亿元,
c₇: 增长 1 3 . 3 %;
c₈: 进出口贸易总额 1 . 1 5 万亿美元,
c₉: 增长 3 5 . 7 %,
c₁₀: 由上年居世界第四位上升为第三位;

c₅、c₇、c₉ 对应的话题自足句为:

- S₅:2004 年财政收入增长 21.4%;
S₇:2004 年社会消费品零售总额增长 13.3%;
S₉:2004 年进出口贸易总额增长 35.7%;

“增长”关系需要说明增长的主体、增长的基点、增长量。在标点句序列中三个标点句所说的对象增长率都是比较于基点“上年”而说的，但是在话题结构中，由于省略了比较对象，生成的话题自足句 S₅、S₇、S₉ 缺失了增长基点，从而导致语义表达不清楚。因此需要将缺失的关联对象“**比上年**”补出来。这一对象的补足可通过模式匹配的方式从上文中获取。

还有个别更为复杂的情况，限于篇幅此处略去。

4.2.2. 语义错乱

跨标点句的说明如果关联到话题所在句中的几个对象，其中有的是话题，另一些是其他关联要素，它们并不相邻，用现有流水模型不能把它们都提取出来，话题自足句由于关联对象缺失而致语义失解。此时需要在适当位置插入话题所在句中的非话题的关联对象，并且常常需要添加介词、结构助词等虚词，使关联对象之间的关系得以清晰表达。这种处理方式超出了现有流水模型的功能，只能归入现有流水模型无法覆盖的范围。

例 16

- c₁: 安徽省……
c₂: 涝灾则以淮河流域较多,
c₃: 主要出现在 7 ~ 8 月;
c₄: 次为长江流域,
c₅: 多出现在 6 ~ 7 月。

例中标点句 c₃、c₅ 对应的话题自足句分别是:

- S₃:安徽省涝灾主要出现在 7-8 月,
S₅:安徽省涝灾多出现在 6-7 月,

原文要表达的语义是“安徽省涝灾在淮河流域主要出现在 7-8 月”，“安徽省涝灾在长江流域多出现在 6-7 月”，而由于“淮河流域”和“长江流域”无法作为话题被共享，话题自足句 S₃、S₅ 表达的语义错误，背离了原文要表达的语义。需要在此类情况下补足必要的成分，有时还需要另加虚词，以保证话题自足句的通顺合法，因此例中需要在“淮河流域”和“长江流域前”加介词“在”，如此处理后生成的话题自足句如下：

- S₃[']:安徽省涝灾在淮河流域主要出现在 7-8 月,
S₅[']:安徽省涝灾在长江流域多出现在 6-7 月,

这样的处理方法超出了流水模型的功能范围。

还有些现象涉及到对象的动态变化，限于篇幅本文不加讨论。

5. 结语

本文界定了话题自足句及其成句性，阐述了研究话题自足句及其成句性的意义，列举并分析了语料中话题自足句不成句的主要情况，并做了详细分类。主要分为两大类：句法不通和语义不清或错乱。其中句法不通的情况主要是篇章连接成分导致话题自足句不成句和新支话题不定指导致话题自足句不成句。篇章连接成分是一个封闭的集合且出现的位置固定在标点句首，新支话题不定指的情况具有较好的形式特征，能够用机械化的办法将不定指变为定指，因此总体上可以通过改进流水模型中话题自足句的生成方法，将其不成句的问题解决掉。

据统计，语料中共有 30963 个标点句，对应 30963 个话题自足句，其中有 1963 个不成句的话题自足句。不成句的话题自足句根据不成句原因的分类情况如下表所示：

表 2 话题自足句不成句现象的分类情况

| 不成句原因分类 | 频次 | 频次/不成句话题自足句总数 | 频次/话题自足句总数 |
|-------------|------|---------------|------------|
| 篇章成分导致 | 869 | 44.3% | 2.8% |
| 新支话题不定指导致 | 251 | 12.8% | 0.8% |
| 主语省略句 | 547 | 27.8% | 1.7% |
| 其他语义不清或语义错乱 | 296 | 15.1% | 1.0% |
| 合计 | 1963 | 100.0% | 6.3% |

表 2 中其他语义不清或语义错乱类的话题自足句不成句的语言现象比较复杂，在现有流水模型的机制中无法提供改进机制使得话题自足句成句。这些现象实例虽然分了多个类别，但是从整个语料来看，所占比例很少。这些现象的处理方法并非完全无迹可循，比如对于复杂关系句引入模式匹配的方法就是一种可以探索的途径。还有相当多的部分可以通过加入介词等功能词来解决。这些可能的解决途径不仅为广义话题结构流水模型的改进提供了可能的方向，而且对于探索说汉语者在听/读标点句序列过程中的话题信息认知机制，提供了十分有益的启示。

受篇幅限制，本文对上表中的主语省略句未进行详细说明，其他各类中也有些出现数量少的细类无法详细示例解释，相关内容可参见笔者的博士学位论文。

本文是广义话题结构理论实证性研究的一部分，说明话题自足句是隐藏在汉语篇章中的一种客观存在的、通常具备成句性的语言单位，少数不成句的则有若干系统性的原因。实证性研究对于科学理论的可靠性、可用性具有重要意义，但只是相关理论研究的一部分。汉语篇章广义话题结构的研究刚刚开始，其中最具实质性的课题之一是如何分析判断标点句之间的话题-说明关系。这一工作具有极大的挑战性，是我们今后工作的目标，需要在实证性研究工作的基础上，经过长期努力来解决。

参考文献

- [1] 赵元任.汉语口语语法(吕叔湘译)[M].北京:商务印书馆,1979
- [2] 朱德熙.语法讲义[M].北京:商务印书馆,1982
- [3] 朱德熙.语法答问[M].北京:商务印书馆,1985
- [4] Li, C. & Thompson, S. Subject and Topic: A New Typology of Language [C]. Charles N. Li (ed.). Subject and Topic. New York: Academic Press, 1976: 457-489
- [5] 徐烈炯、刘丹青.话题的结构与功能[M].上海:上海教育出版社,1998
- [6] 曹逢甫.主题在汉语中的功能研究:迈向语段分析的第一步 [M].北京:语文出版社,1979
- [7] 曹逢甫.汉语的句子与子句结构 [M].北京:北京语言大学出版社,1990
- [8] 沈开木.现代汉语话语语言学[M].北京:商务印书馆,1996
- [9] 屈承熹.汉语篇章语法[M].北京:北京语言大学出版社,2006
- [10] 许余龙.篇章回指的功能语用探索[M].上海:上海外语教育出版社,2004
- [11] 吕叔湘.汉语语法分析问题[M].北京:商务印书馆,1979

- [12] 范继淹.汉语句段研究 [C], 范继淹语言学论文集, 北京: 语文出版社, 1986:252-270
- [13] 沈家焯. “零句”和“流水句”——为赵元任先生诞辰 120 周年而作 [J]. 中国语文, 2012(5): 403-415
- [14] 陈平.汉语零形回指的话语分析 [J].中国语文, 1987(5):363-378
- [15] 宋柔.汉语篇章广义话题结构的流水模型[J]. 中国语文, 2013(6):483-494
- [16] 蒋玉茹, 宋柔.基于广义话题理论的话题句识别 [J], 中文信息学报, 2012(5): 114-128
- [17] 尚英.汉语篇章广义话题结构理论的实证性研究 [D], 北京: 北京语言大学, 2014
- [18] 尚英, 宋柔.基于广义话题结构语料库的语体对比研究——以报告体与小说体为例 [J].计算机工程与应用, 2014(11): 21-27
- [19] 张瑞朋.现代汉语书面语中跨标点句句法关系约束条件的研究 [D], 北京: 北京语言大学, 2007