

时间信息提取的几个问题*

陈振宇¹, 陈振宁²

(1 复旦大学 中国语言文学系, 上海 200433

2 文华学院, 湖北 武汉 430023)

摘要: 怎样将汉语文本中的时间信息提取出来, 是当代计算语言学的一个难点。本方案将时间体和事件归为“时间单位”, 并从五个阶段以及论元之间的关系对它们进行逻辑描写。利用突显阶段和对齐规则, 可以推出两个单位之间的对齐关系。再运用一些与上下文无关的规则和原则, 对结论进行修正, 以便把蕴涵在文本中的信息更为详细地描写出来。在对齐和修正过程的同时, 功能性成分也会实施某种改写操作, 以便更准确地反映说话者的意图。

关键词: 时间信息提取; 时间单位; 突显阶段; 对齐

中图分类号: TP391 **文献标识码:**

Several Issues about Temporal Information Extraction

Chen Zhenyu¹, Chen Zhenning²

(1 Dept. of Chinese Language and Literature, Fudan University, Shanghai 200433, China

2. Wenhua College, Wuhan, Hubei 430074, China)

Abstract: For computational linguistics, it is difficult to extract temporal information from a given Chinese text. In this program, the temporal entities and events were ranged into *temporal units*, which should be logically described as matrixes with five phases and relations among arguments. Alining relation between two units was deduced with the help of profiled phases and alining laws. Then some of the context-independent laws and principles were used to amend the conclusion in order to provide more details about the information entailed in the text. All the process of alining and amending went with some kinds of rewriting operation realized by functional components which reflected the intention of speaker more accurately.

Key words: temporal information extraction; temporal unit; profiled phase; alining process

* **收稿日期:** **定稿日期:**

基金项目: 教育部人文社会科学规划基金项目《现代汉语语法与语义计算研究》(批准号: 13YJA740005)。

作者简介: 陈振宇(1968—), 男, 副教授, 主要研究方向为信息抽取、语义知识库、现代汉语语法; 陈振宁(1977—), 女, 教师, 主要研究方向为信息抽取和语义知识库。

1 已往的研究以及本文的框架

对自然语言的语句和篇章进行形式化, 将其中关于事件的时间进程和事件之间的时间关系的语义信息转化为数据, 以供计算机储存与分析, 这就是所谓“时间信息提取”(extraction of temporal information)。它并不仅仅是计算领域的事, 实际上也是语言本体研究中必不可少的内容。时间信息提取研究受到了语法学较大的影响, 但是, 它与时(tense)、体(aspect)等语法范畴毕竟有所不同:

1) 研究目的不同。语法学主要是为了获得有关时间标记的语法意义、适用性、语法化以及类型学特征等方面的知识; 而在时间信息提取中, 需把语句所表达的所有事件和时间体, 都“拉”到一条一维单向的时间轴上, 并把其前、后、伴随等时间对应关系一一标明。

2) 研究视角不同。语法学可以只关注于少数的热议问题, 如汉语的时、体系统, 以此“以小见大”, 但这些问题在信息提取中并不一定占据中心地位。有的信息研究比语法学考察得更为全面, 如要求阐释“前后结构”和“上下结构”所包括的时间关系, 有的则比语法学更为简化, 如时、体成分虽属不同的语法范畴, 但都表示时序关系, 所以完全可以用统一的形式方法加以处理。

3) 研究方法不同, 语法学采用归纳、总结的方法, 重视对语料的观察; 而只有为时间系统整体建立数学模型, 才能进行演绎和运算。

早在 90 年代, 郭宏蕾、姚天顺^[1]就提出, 要将时间描述映射到时间轴上, 从而准确刻画文本中各事件的时间相关性的任务; 不过文中虽然提出了有关时间区间、事件区间、事件状态等一系列语义分解目标, 但实现这些目标的具体语言分析手段却不大明确。这时, 汉语学界的时间研究才刚开始不久, 最早论述这方面知识的专著如龚千炎^[2]、戴耀晶^[3]等才出版, 学界普遍缺乏对汉语时间系统的整体了解。自那以后, 汉语时间信息提取研究有了长足的进展, 国内发表的有关论文大致可分为以下五类:

1) 对时间体的语义内容进行详细的形式化分解的, 如马红妹、王挺等^[4], 吴平博、陈群秀等^[5], 以及林静、曹德芳等^[6]和贺瑞芳、秦兵等^[7]所涉及的 TIMEX2 自动标注系统。

2) 仿照汉语学界的一般研究, 从时(又称为时制、时态)、体(又称为时态、体貌)、情状(又称为时相、事件时间结构)入手寻找信息提取的语言学手段的, 如马红妹、齐璇等^[8], 陈振宇、陈振宁^[9], 陈振宇^[10]。

3) 从时态(tense)逻辑入手进行时间关系计算的, 如麻志毅、林鸿飞等^[11], 其实质是确定一个命题为真的时间。

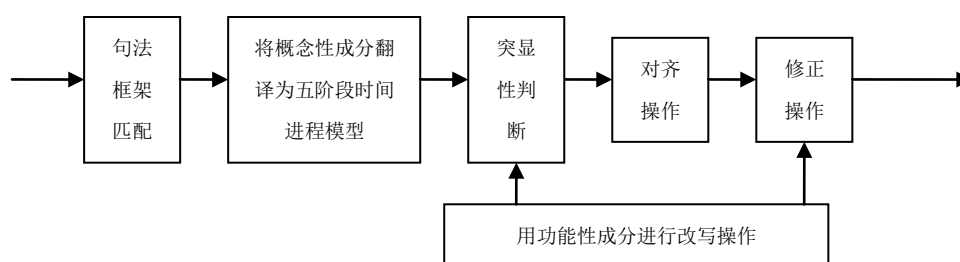
4) 从 Allen^[12]的 13 种时间关系入手提取时间信息的, 如林闯、刘婷等^[13]和刘婷、林闯等^[14], 徐永东、徐志明等^{[15]、[16]}和徐永东、王亚东等^[17]。有的也提到 Reichenbach^[18]的三个时间。

5) 除此之外, 还有在有关理论上采用机器学习方法的, 如王昀、苑春法等^[19], 贺瑞芳、秦兵等^[7]。

今天大多数研究者都认识到, 时序关系在汉语中主要采用语义和语用来表达, 而没有严格的时、体标记。一些研究者最初从时、体标记入手来寻求提取方案, 但发现不能完全依靠它们, 无标记现象不但普遍存在, 而且更为基础。如郭宏蕾、姚天顺^[1]就提到了复句的时间语义结构, 并认为主、从句事件间有时间依存性; 马红妹、齐璇等^[8]在关于事件时意义的判断中, 认为当事件没有时间短语限定时, 需要通过上文事件来进行判断, 并运用事件的语义角色(他们称为“概念信息体关网络”)来加强时间短语的识别; 林静、曹德芳等^[6]在时间短语意义内容不够明确时, 考察它与“最近叙述时间”、“报道时间”和“文档时间”之间的关系; 麻志毅、林鸿飞等^[11]也提到文本中若干句子是否划归为同一事件过程的问题; 陈振宇^[10]则讨论了汉语时、体标记“了”的不准确性, 提出了一个以语义规则为主, 时间标记为辅的时序关系计算方案(书中的“第二方案”)(331-378、416-439)。

另一些研究者则试图把无标记模型与有标记作用合并使用。最为典型的是脱胎于Allen^[12]的理论系统的若干研究，在对时间关系进行分类后，一方面试图做进一步的推导，以发掘蕴含的时间信息，另一方面试图分析事件和时间体、事件和事件之间的关系的语言表现手段。如林静、苑春法^[20]既借助于有标记的方位词（前、后等）、时间副词、时相动词（开始、完成等）、时间介词、连词（反映并列、顺承、因果、假设等具有时间性质的关系）等，也提到了一些较为复杂的关系，如“他说，前天……”与“前天，他说……”之间的区别；徐永东、王亚东等^[17]则把这些称为偏移算子，另加上偏移量（“立刻、马上”等）以及时态算子（“已经、着、了、过”等）。

汉语的实际情况要求我们建立一个以“无标记的时间关系”作为主干，以“有标记的时间关系”作为补充的理论体系，才能较好地解决时间信息提取问题。我们认为，时间信息提取中各个模块的分工如下图所示：



图（1）：时间信息提取的总框架

自然语言语句由若干句法成分所组成，分为“概念性成分”（又称为“实词性成分”、“独立的句法成分”等）和“功能性成分”（又称为“虚词性成分”、“非句法成分”等）。前者代表事物，可以翻译为独立的基础表达式；后者不代表事物，没有内涵和外延，不能翻译为独立的表达式，但是它们支配或依附于概念性成分，会对后者的表达式进行改变，这种改变称为“改写”（rewrite）。先有基础表达式，再有改写操作，所以“功能性成分”既不能忽视，但也不能喧宾夺主。

流程图（1）显示，当没有相关的功能性语符时，可以从概念性成分的句法语义内容出发，通过突显、对齐、修正三个步骤，推导出语句的时间信息；不过有了功能性成分，它便可以在各个操作步骤中参与进去，更为严格地控制语句的时间信息。

本文包括以下几个问题：时间单位与其突显性、时序关系及其对齐方法、修正操作，以及运用特定语符的改写操作。由于功能性成分是前人研究中的重点，限于篇幅，本文将重点介绍我们在无标记的时间关系方面的研究成果。

2 无标记的时间关系

2.1 时间单位及其突显性

概念性成分表示时间单位和实体单位。时间单位（temporal unit，简记为“T”）包括时间体（temporal entity）和事件（event），实体单位则指填入事件中的那些论元（argument）。时间体指对时间的专门表述，如“去年、上个月、十九世纪”等，它们通常是时间状语或补语，而事件由句子的核心谓词及其论元结构来表达。

在语法学上，时间体一般由体词性成分表达，而事件一般由谓词性成分表达；但其语义信息是相通的：一个事件可以像时间体那样充当其它事件的时间状语、补语或其它关联成分。如“他走进来（时），看见小玉”中事件“他走进来”充当后一事件“他看见小玉”的时间状语，即背景时间。在信息提取中，如把它们作为两个子系统来处理，就会繁琐、冗余，

如 TIMEX2 标注对“去年十一月底以前”“他读书以前”的处理。现在的理论越来越趋向于使用同样的描写模型和认知原则来处理它们。

任何一个时间单位，不论是时间体还是事件，在理想的认知模型中，都具有“五阶段（phase）模板”的时间结构：Ss（state before start，起始前状态）、s（start，起始）、Sc（state of continue，延续状态）、e（end，终结）、Se（state after end，终结后状态）。其中三个是状态（state），即相对平稳的阶段，即使有变化，也是相对平稳的量变；而 s 和 e 是边界（boundary），即发生质变的阶段；一般来说，状态是时段；边界是时点。下面例（1）为事件“转让”的逻辑表达式：

- (1) x（以 n）转让 y（给 z）
 x（以 n）将 y 转让（给）z
 x（以 n）将 y 对 z 转让
 x 转让 z y
 y（x）（以 n）转让（给）z
 ……
 \Rightarrow Ss 转让[possess（x，y） \wedge \neg possess（z，y） \wedge \neg possess（x，n） \wedge possess（z，n）] s Sc 转让[cause（x，change（possess（x，y） \wedge \neg possess（z，y） \wedge \neg possess（x，n） \wedge possess（z，n）））] e/Se 转让[\neg possess（x，y） \wedge possess（z，y） \wedge possess（x，n） \wedge \neg possess（z，n）]

语法学研究多效法 Vendler^[21]，将动词划分为几个简单类型，但是对于时间信息的提取来说，它们都过于简单，没有达到计算所要求的最少信息量，至少需要做以下几点改进：

1) 在语言中，最基本的时间结构不是由谓词，而是由谓词及其论元所组成的句法框架来表达的，如“x（以 n）转让 y（给 z）”“x（以 n）将 y 对 z 转让”等等，它围绕核心“转让”组织起来，有若干空位 x、y、z、n，是这一事件中的不同语义角色施事、受事、与事、等价物在句中的位置。

2) “ \Rightarrow ”后面是有关框架的基础表达式，至少应该包括三个部分：a 事件的阶段，b 每一个阶段的语义内涵，c 阶段的突显性。如“Ss 转让[…]”指转让事件有开始前状态 Ss；“[]”内为这一阶段的语义内涵，“possess”指拥有关系，指在转让开始之前（用 Ss 表示），x 拥有 y，z 不拥有 y，但 x 不拥有 n，z 拥有 n；这一关系在转让终结和终结之后发生了改变。这样一来，事件模型即是以事件诸阶段为横轴，以每一阶段的论元关系为纵轴所构成的意义的矩阵。

公式中加在阶段 Sc 和 e 上的方框，表示在转让事件中延续和终结阶段最为突显，最易引起双方的重视。要实现时间信息提取，需要一些技术手段。其中最为重要的是在无标记的情况下，必须从时间单位的时间结构中抓出一些要紧的地方，作为对齐操作的抓手，这一抓手就是阶段的突显性。

根据突显性的强弱，单一时间单位中的阶段分为三种（若 p 指某一特定阶段）：默认的突显性阶段，一般都会受到重视，记为“p”；背景性阶段，一般不会被注意到，记为“p”（加下划线斜体）；普通阶段，不被忽视但也不会受重视，用正常字母书写。陈振宇对怎样判断突显性的问题有详细讨论^{[10]、[22]}。显然，不同的事件有不同的突显性分布，如：

- (2) x 拥有 y \Rightarrow Ss 拥有[\neg possess（x，y）] s/Sc 拥有[possess（x，y）] e/Se 拥有[\neg possess（x，y）]
 (3) x（从 z）得 y \Rightarrow Ss 得[possess（z，y） \wedge \neg possess（x，y）] s/Sc 得[cause（？，change（possess（z，y） \wedge \neg possess（x，y）））] e/Se 得[\neg possess（z，y） \wedge possess（x，y）]

“拥有”一般指拥有的延续阶段，所以突显延续 Sc 阶段，有时也指从不拥有变为拥有，所以也可以突显开始 s 阶段，但拥有状态的结束则不被关心，必须有“拥有过”等特殊标记

才会使人注意到它。而“得、获得”等事件是所谓终结事件，不关心开始，只突显事件的结果 e。根据我们的研究，在最基础的层次上，形容词框架共有 8 种突显性类型，动词框架则有 19 种，它们分为五大典型类及若干非典型的类。这一数字并非定论，不过刻画得越仔细，越有利于计算。

“五阶段模板”不仅适用于事件，同样适用于时间体，例如“春节”：下面“locate”意为位于，T 指其他时间单位，可以是其他时间体，也可以是事件。根据语境颗粒度的大小，有时把“春节”视为时段，此时突显节日开始和持续的 s 和 Sc 阶段：

(4) 春节 \Rightarrow Ss 春节[\neg locate (T, 春节)] $\boxed{s/Sc}$ 春节[locate (T, 春节)] e/Se 春节[\neg locate (T, 春节)]

有时把“春节”视为时点，此时它相当于一个瞬间事件，即不考虑开始终结的过程，仅看成一个“s=e”的点，这时突显的就是这个点：

(5) 春节 \Rightarrow Ss 春节[\neg locate (T, 春节)] $\boxed{s=e}$ 春节[locate (T, 春节)] Se 春节[\neg locate (T, 春节)]

填入事件的论元，称为实体单位，有一些特殊的实体单位，必须以一定的事件为其背景，在翻译时必须把事件也写出来，例如“股息”，这是因为一个人拥有股票，从而获得了一笔钱，故需要把一个因果事件描写出来。有关复杂实体单位的问题，我们将另文论述。

2. 2 前景事件、背景事件、对齐规则

两个时间单位之间的时序关系，Allen 及其追随者们归纳了 13 种之多，但有的仅七种，如林闯、刘婷等^[13]；有的仅三种，如徐永东、徐志明等^{[15]、[16]}。这是因为有的时序关系是可以由其他时序关系和事件的性质合成的，所以可以简化掉。另外，Allen 等讲的是事件或时间单位之间的时序关系，例如“他抽着烟走进来”，它们的关系当然是 contain，即在他抽烟的持续过程中，发生他走进来的事，为“他抽完烟走进来”，当然是 after，即他走进来发生在他抽完烟之后，可见在时间单位层次上，时序关系不能太少。

换一个角度看则不然，仍以“他抽着烟走进来”为例，可以认为“走进来”整个事件（包括开始、持续、终结在内）向“抽烟”的持续阶段对齐，写为公式即“Sc 抽烟 \geq s/Sc/e 走进来”；而“他抽完烟走进来”，是“走进来”整个事件向“抽烟”的终结后状态对齐，写为公式即“Se 抽烟 \geq s/Sc/e 走进来”。这样一来，所有的时序关系都可以归结为一个概念，即“对齐”（aline），它大致相当于前人所说的“包含”（contain），但有细微的差异。其中“ \geq ”表示是右边的阶段向左边对齐，当然左边的时间长度也大于或等于右边。

aline 本是指两个（或多个）相互独立的机器部件，其互相平行或共同运动的侧面排到正确的位置或方向，从而保证既能各自独立运动，又能在正确的位置上互动。我们用这一术语，是指两个独立的时间单位，将处于同一时间位置的阶段在时轴上排到正确的位置上。显然，时间单位层次的所有时序关系，都可以分解为阶段层次的对齐操作。

如果暂时抛开细微末节以及不规则的地方，那么任意两个时间单位之间，存在两种可能的语言结构关系：前后结构和主从结构，它们各自有不同的对齐操作，但在认知本质上则是相通的，即在信息链中，任何两个句法相联、语义相关且不是平等并列的事件，都有“背景”（background）与“前景”（foreground）的相互关系。前景事件在当下的言语活动中是突出的，是信息的焦点所在，是言说者最想说明的东西；而背景事件则是旧信息，作用是提供相关的背景性知识，以衬托前景事件。背景与前景的关系被形象地比喻为窗口与窗中的风景。

例如连续的语句会导致背景与前景的更迭转换，如下例中有四个时间单位，：

(6) 昨天 (T₁)，小王把股票转让给小李 (T₂)，他得到 10000 块钱 (T₃)，都用来支付税款了 (T₄)。

按语句出场次序，前一个时间单位为后一个提供时间背景，后者则相对于前者为前景信息。一个时间单位 T_i 为它后面的时间单位提供的背景或“窗口”，称为“时间论域”（temporal

scope), 简称“时域”。这样一来, 事件的对齐操作就转化为下面两个问题: 1) 背景事件会提供什么样的时域? 2) 前景事件在时域中如何呈现?

先看第一个问题, 特定时间单位所提供的时域由它自身的突显性质决定:

a、当 T_i 突显某一状态时, T_i 提供的时域就是这一状态本身。如 (6) 中, T_1 “昨天”在被看成时段时, 它提供的时域就是其突显的持续状态 “Sc 昨天”。

b、当 T_i 突显某一边界时, 一般来说, 以该边界之后的那个状态为时域, 如 (6) 中, T_3 “他得到 10000 块钱”突显终结阶段 e, 所以它提供的时域便是终结之后的状态 “Se 得到”, 后面的事件 “支付税款” 显然是在得到之后才发生的。

再来看第二个问题, 一个特定的时间单位在时域中呈现的样子也由它自身的突显性质决定, 即前景事件 T_{i+1} 的突显阶段在背景事件 T_i 所提供的时域之中呈现。如 (6) 中, T_4 突显支付的过程 Sc 或结果 e, 所以对齐关系 “Se 得到 \geq Sc/e 支付”, 意为 “在得到事件终结之后, 有支付事件发生”。

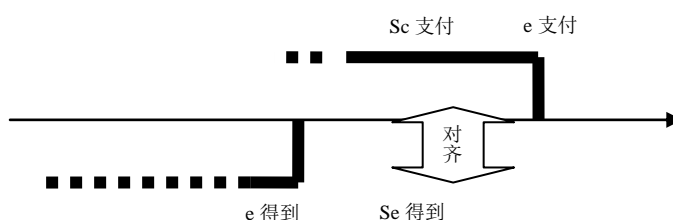


图 (2): “他得到 10000 块钱, 都用来支付税款了” 的对齐结果

需要特出的是, 当 T_i 突显某一边界时, 有一种极为特殊的情况, 即当 T_{i+1} 突显某一时段时, 会要求以 T_i 突显的那个边界作为时域 (这种特殊的时域为一个时点), 例如:

(7) 他到达终点 (时) (T_1), 我还在努力呢 (T_2)。

T_2 “在努力” 突显延续 Sc 状态, 这时, 前面 T_1 提供的时域就是它所突显的终结阶段 “e 到达”, 指在他得到的那一刻我在努力。这时的 “对齐” 操作较为特殊, 有等于关系, 即 “e 到达 = Sc 努力”, 意为 “在到达事件终结的同时, 努力事件正在持续”。

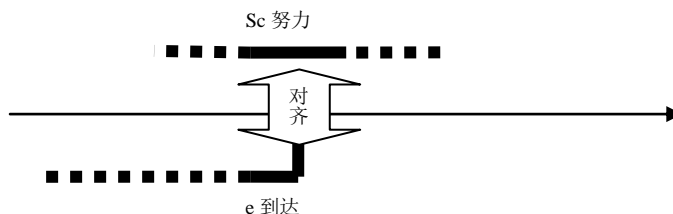


图 (3): “他得到 10000 块钱, 我还在努力呢” 的对齐结果

请注意, 图中实线指语句中能够确定下来的阶段, 即能准确提取的时间信息, 而虚线指句中尚未提到, 或无法确认的阶段, 它们只是可能存在, 句义并没有直接告诉我们它们已经发生或在何时发生。

背景与前景关系这一分析同样适用于主从结构:

(8) 前天 (T_{21}) 我听说 (T_{22}) 昨天 (T_{11}) 他会得到 10000 块钱 (T_{12})。

句中四个时间单位分为两个层次, 其中宾语从句事件 “他会得到 10000 块钱” 是主句事件 “我听说” 的内容, 二者也有直接的时间关系, 需用背景与前景关系来加以刻画。当句子只有结构焦点而没有对比焦点时, 主句代表前景事件, 是信息焦点所在; 而从句代表背景事

件，为主句提供背景知识，作为前者的先设信息。另外，(8) 中从句事件有一个“会”，它强制突显“得到”结果实现以前的所有可能阶段，故从句提供的时域为“Ss/s/Sc 得到”；而主句事件“听说”突显持续阶段，所以有“Ss/s/Sc 得到 \geq Sc 听说”，意为“在得到结果之前，有听说事件发生”。

主从结构也有一些特殊的情况，例如：

1) 当从句是祈使（包括命令、许诺、期盼等）句时，从句事件突显开始之前的 Ss 阶段，故“妈妈叫我赶紧穿好衣服”，从句提供的时域为“Ss 穿”，最终有“Ss 穿 \geq Sc 叫”，即“妈妈叫我”是在“我穿好衣服”还未发生的时候。

2) 当从句是静态事件时，从句事件突显延续 Sc 阶段，如“小王说这笔钱属于他”，从句提供的时域为“Sc 属于”，最终有“Sc 属于 \geq Sc 说”，即“小王说”是在“这笔钱属于他”的时间内发生的。

3) 从言语活动角度讲，每一句话都可以分析为主从结构，即使是一个简单句，如单说“这笔钱属于小王”，实际上也相当于“我正在说这笔属于小王”这样的主从结构。其中“我”指说话者，从句事件与我说话的时间对齐，这正是“说话时间是事件的基点”的本质。这也说明了言语活动理论的一个基本原理：形式上的隐而不显并不意味着语义上的缺失，恰恰相反，任何一个 T_i 都需要一个与之相关的主句 T_{i+1} 才能获得完整的语义解释，而这一解释最终必然归结于言语活动。

2.3 修正操作

时间单位初步对齐后的结果很可能是有缺陷的：图中虚线部分的有关信息并没有得到确认。这一部分信息不是话语的语义内容直接告诉我们的，所以不能准确地提取；但是在信息交流中，它们又是必要的。实际上，听说双方可以通过调用一些日常生活知识，以及具有普遍性的认知和语用规则，来尽力推导出这些信息。这就是“修正操作”，它主要包括“边界修正”、“解歧”、“真值蕴涵”、“动态性推定”等。

1) 边界修正，是所有修正操作中最重要的一個，其实质是心理上的“完形”(gestalt) 机制：陈振宇、陈振宁^[23]说，如果背景是有边界的时段，那么就会觉得呈现在这一窗口中的前景事件，也是有边界的。如前面图 2 中所呈现的情况，准确地讲，“支付”的持续阶段与“得到”的终结后状态对齐；不过，“Se 得到”是一个以“得到”的终结 e 为起点，并向后延伸的区间（下图中虚线箭头所示），它有前界点；完形机制使我们觉得在其中呈现的“支付”事件也应该有前界，于是把“支付”的开始 s 也呈现出来了，意为“在得到事件终结之后，支付事件开始并完成”：

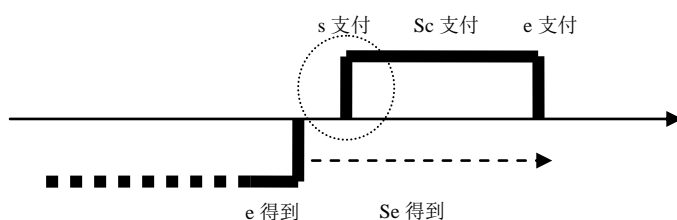


图 (4): “他得到 10000 块钱，都用来支付税款了”修正以后的结果

主从结构也有同样的情况。如例 (8) 中“Ss/s/Sc 得到”（即“得到”结果实现以前的所有可能阶段）是一个以“得到”的终结 e 为起点，并向前延伸的区间，它有后界点；完形机制使我们觉得在“Ss/s/Sc 得到”中呈现“听说”事件也应该有后界，于是把“听说”的终结 e 也呈现出来了，意为“在今天之前，就已经听说过了（听说事件已经结束了）”。

再如例 (6) 中 T_1 “昨天”和 T_2 “小王把股票转让给小李”之间，根据对齐规则有“Sc

昨天 \geq Sc 转让”，意为“在昨天的持续时间中，有转让事件的持续”。但是“Sc 昨天”是一个封闭区间，既有前界点，又有后界点；完形机制使我们觉得在其中呈现的“转让”事件也应该有前界 s 和后界 e，且它们都在“昨天”的边界之内，故是指“在昨天的持续时间中，转让事件开始、持续及终结”。

边界修正是一个十分重要的研究，有不少特殊性质，如一般来讲，背景性阶段不适用于边界原则，例如“他知道我们今天会去”，“知道”的终结 e 是被忽略的背景阶段，所以不能对它进行边界修正。边界原则又是语用原则，可以违反，例如“昨天他当了班长”，虽然“昨天”是个封闭区间，但“他当班长”这事，仅是对开始 s 进行了边界修正，对终结 e 却不能。我们发现，在汉语中存在一条“开始优先原则”：当开始和终结都可能适用于边界修正时，优先对开始进行修正；至于终结，需要考虑其它的限制条件，其中极为重要的是“时间长度限制规则”：当事件已有边界点在窗口呈现的时候，如果前景时间单位的时间长度，小于或等于背景时间单位的时间长度，则可以对两端边界点都进行边界修正；如果前景大于背景时间单位的时间长度，则最多可以对一端的边界点进行边界修正。例如“昨天他当了班长”，由于“当班长”是一个比较长的过程，其时间长度远大于“昨天”，所以不能对“当班长”的开始和终结都进行边界修正；根据“开始优先原则”，可以对开始进行边界修正，而“当班长”的终结则不知道在什么时候。最终推论得“昨天小王开始当班长”。

2) 解歧，有些事件可能的突显阶段有多个，如“转让”既可能突显过程，又可能突显结果，这必然带来歧解。目前最有效的解歧策略是根据语义内涵来进行选择。如(6)中 T_3 “得到”只突显“e 得到[possess (小王, 10000 块钱)]”；而 T_2 “转让”提供的时域却有两个取值：或者是突显转让过程“Sc 转让[cause (小王, change (possess (小王, 股票) \wedge \neg possess (小李, y) \wedge \neg possess (小王, n) \wedge possess (小李, n)))]”，或者是突显转让结果后的状态“Se 转让[\neg possess (小王, 股票) \wedge possess (小李, 股票) \wedge possess (小王, n) \wedge \neg possess (小李, n)]”。究竟哪一个对呢？根据指称机制判断出“n=10000 块钱”，在“e 得到”阶段，有“possess (小王, 10000 块钱)”，而“转让”的参考指针也必须满足这一点，因此我们只能选第二个，最后结论为“Se 转让 \geq e 得到”。

换一个上下文完全可能得到不同的结论，例如：

(9) 小王把股票转让给小李，但小李没要。

小李要股票 \Rightarrow Ss 要[\neg possess (小李, 股票)] s/Sc 要[cause (小李, change (\neg possess (小李, y)))] e/Se 要[possess (小李, 股票)]

由于否定的是事件的发生，也就是开始 s 及开始以后的阶段都不存在，故只有以下关系存在：

(9') 小李没要股票 \Rightarrow Ss 要[\neg possess (小李, 股票)]

“转让”的时域也必须满足“ \neg possess (小李, 股票)”的要求，故结论为“Sc 转让 \geq Ss 要”，即在转让的过程中，小李没要。

除此之外，我们还发现了其他一些修正操作，如基于时序的真值蕴涵规制、基于背景—前景关系的动态性推定等。

3 功能性成分的类型

3.1 改写突显阶段的功能性成分

在汉语中，这一类最为丰富，如光杆论元结构“他把股票转让给张三”或突显延续 Sc 阶段，或突显终结 e 阶段，加上功能性成分“正在”后“他正在把股票转让给张三”则强制性地只突显延续 Sc，不再突显 e；加上功能性成分“开始”后“他开始把股票转让给张三”则强制性地突显开始 s，不再突显 Sc 和 e，这就是改变了原有的突显性类型。下表为突显性改写成份中最常见的四大类：

VP 内成分	时、体类成分	时间单位外部的成分	语气类成分
<p>a、T 动量/时量：不论事件原来突显哪一阶段，加上动量或时量性成分后，都强制突显这一动量或时量达到的那个时点。</p> <p>b、T 到 T₁：强制以 T₁ 的达到作为突显的阶段，</p> <p>c、T 完、T 到、T 住、T 着（补语）、T 过（补语）、T 毕：强制突显 T 的 e 阶段，</p> <p>d、开始 T、着手 T：强制突显 T 的 s 阶段。</p> <p>……</p>	<p>a、T 着、在 T、正（在）T、进行 T、T 的过程等：强制突显事件的 Sc 阶段。</p> <p>b、T 了等：强制突显事件的 s 阶段或 e 阶段，</p> <p>c、将/会/要/即将 T、来/去 T[表祈使]等：强制突显事件的 Ss 阶段或 Sc 阶段，</p> <p>d、T 过、曾经 T 等：强制突显事件的 Se 阶段，</p> <p>e、已经 T、刚 T 等：强制突显事件的 Sc 或 Se 阶段，</p>	<p>a、（在/当）T 时/的时候、T 时期/期间/时分/时节/时刻、于 T：保持 T 的突显性结构。</p> <p>b、（在）T（以/之）前/后、T（以/之）前/后（的）这段时间、T 接着/然后、如果/若/假如 T 就/便/则、T 的话就/便/则：强制突显原来的 T 突显阶段之前或之后的状态；但如果原来的 T 突显时段，且该时段后没有其它状态或只有背景阶段时，则保持原来的突显状态。</p> <p>c、（在）T（以/之）外：强制突显 T 的 Ss 或 Se 阶段。</p> <p>d、（在）T（以/之）内/里/中：强制突显 T 的 Sc 阶段。</p> <p>e、过（了）T、T 过（了）：强制突显 T 的 Se 阶段。</p> <p>f、从/自/打/自从/自打 T（以/之）后、从/自/打/自打 T 起/以来/开始：强制突显原来的 T 突显阶段之后的状态；但如果原来的 T 突显时段，且该时段后没有其它状态或只有背景阶段时，则保持原来的突显状态。</p> <p>g、到/直到/等到（了）/至 T…才：保持 T 的突显阶段。</p> <p>到/直到/等到（了）/至 T…仍（然）/还（是）/依然/始终/一直：强制突显原来的 T 的突显阶段之前的状态。</p> <p>……</p>	<p>祈使（包括命令、许诺、期盼等）句中的 T：强制突显 T 的 Ss 阶段。</p>

表（1）：突显性改写成分

3. 2 否定

对时间单位的否定必须落实到具体的阶段上，有以下规则：

- 对时间单位整体的否定，即是对该单位开始阶段的否定；反之亦然。
- 对延续阶段的否定就是对开始阶段的否定。如“他没在看书”也就是“他没看书”。
- 对终结阶段的否定并不意味着否定过程，从语用性的角度来说，倒是趋向于肯定开始和持续阶段为真，只不过未达到结果。

与英语不同，现代汉语中的否定词需要进行分化：

- 不 T、别 T：否定 T 的 s 阶段。
- 没 T、未 T：否定 T 的突显阶段。

例如“小李不要股票”，否定的是“要”的事件的开始。而“小李没要到股票”，由于“到”强制突显终结 e 阶段，所以“没”就仅仅是对该终结的否定，而趋向于认为要的过程 Sc 为真（除非在后续句中加上一个进一步的修正，如“（小李没要到股票，）实际上他根本没去要”之类）。

3. 3 改写边界操作的功能性成分

主要是控制能否进行边界修正，分为三类：

- T 着、在 T、正（在）T 等：不允许进行边界修正。
- T 了、已经 T、将/会/要 T、T 过、曾经 T 等：强制要求至少对一端边界进行修正。

边界强制性操作成分是以往本体研究与信息提取研究的重点，限于篇幅，本文不再赘述。这里仅举一例说明：

（10）a. 我进来之前，他刚出去。

b. 我进来的时候, 他刚出去。

“刚 T” 强制突显事件“出去”的 Se 阶段, “T 之前” 强制突显“进来”事件的 Ss/s/Sc 阶段 (当“进来”被视为终结事件时), 故在下面例 a 中, 有“Ss/s/Sc 进来 (我) \geq Se 出去 (他)”, 又根据边界修正, “我进来之前”是一个以“e 进来”为边界, 向左延伸的时域, 故“出去”事件会因为完形把“e 出去”这一边界也得到突显, 即“Ss/s/ScSs 进来 (我) \geq e 出去 (他)”, 在我进来发生前, 他已出去。有时也会有例 b 的情况, “T 的时候”强制突显事件“进来”的突显阶段, 即 e 阶段, 这是一个时点, 故有“e 进来 (我) = Se 出去”, 即在我进来发生时, 他出去这一事件已终结。从命题意义看, 两个句子差不多, 但推理的过程与所依据的操作都是不同的。

参考文献

- [1] 郭宏蕾, 姚天顺. 时间语义层次结构及理解[J]. 中文信息学报, 1997, 11 (1): 11-19.
- [2] 龚千炎. 汉语的时相时制时态[M]. 北京: 商务印书馆, 1995.
- [3] 戴耀晶. 现代汉语时体系统研究[M]. 杭州: 浙江教育出版社, 1997.
- [4] 马红妹, 王挺, 陈火旺. 汉语篇章时间短语的分析与时制验算[J]. 计算机研究与发展, 2002, 39 (10): 1211-1220.
- [5] 吴平博, 陈群秀, 马亮. 基于时空分析的线索性事件的抽取与集成系统研究[J]. 中文信息学报》2006, 20 (1): 21-28.
- [6] 林静, 曹德芳, 苑春法. 中文时间信息的 TIMEX2 自动标注[J]. 清华大学学报 (自然科学版), 2008, 48 (1): 117-120.
- [7] 贺瑞芳, 秦兵, 刘挺, 潘越群, 李生. 基于依存分析和错误驱动的中文时间表达式识别[J]. 中文信息学报, 2007, 21 (5): 36-40.
- [8] 马红妹, 齐璇, 王挺, 陈火旺. 汉英机器翻译中汉语篇章时间信息系统模型[J]. 计算机工程与科学, 2002, 24 (4): 85-88.
- [9] 陈振宇, 陈振宁. 怎样计算现代汉语句子的时间信息[J]. 中文信息学报, 2005, 19 (3): 94-104.
- [10] 陈振宇. 时间系统的认知模型与运算[M]. 上海: 学林出版社, 2007.
- [11] 麻志毅, 林鸿飞, 姚天顺, 马佳琳. 基于情境的文本中的时间信息分析[J]. 东北大学学报 (自然科学版), 1999, 20 (3): 239-242.
- [12] Allen, J. F. Towards a general theory of action and time[J]. *Artificial Intelligence*, 1984, 23: 123-154.
- [13] 林闯, 刘婷, 曲扬. 一种不确定时段的扩展时段时序逻辑: 时间 Petri 网模型表示和线性推理[J]. 计算机学报, 2001, 24 (12): 1299-1309.
- [14] 刘婷, 林闯, 刘卫东. 扩展时段时序逻辑的推理机制[J]. 计算机学报, 2002, 25 (6): 637-644.
- [15] 徐永东, 徐志明, 王晓龙. 基于信息融合的多文档自动文摘技术[J]. 计算机学报, 2007, 30 (11): 2048-2054.
- [16] 徐永东, 徐志明, 王晓龙, 刘远超. 中文文本时间信息抽取及语义计算[J]. 哈尔滨工业大学学报, 2007, 39 (3): 438-442.
- [17] 徐永东, 王亚东, 刘杨, 王伟, 权光日. 多文档文摘中基于时间信息的句子排序[J]. 中文信息学报, 2009, 23 (4-1): 27-33.
- [18] Reichenbach, H. *Elements of Symbolic Logic*[M]. New York: The Free Press & London: Collier-Macmillan, 1947.
- [19] 王昀, 苑春法. 基本转换的时间-事件关系映射[J]. 中文信息学报, 2004, 18 (4): 23-30.
- [20] 林静, 苑春法. 汉语时间关系抽取与计算[J]. 中文信息学报, 2009, 23 (5): 62-67.
- [21] Vendler, Z. *Linguistics and Philosophy*[M]. Ithaca: Cornell University Press, 1967.
- [22] 陈振宇. 三层七元的时间系统[A]. 语言研究集刊 (第六辑) [C]. 上海: 上海辞书出版社, 2009, 159-179.

[23] 陈振宇, 陈振宁. 完形体“了₁” [A]. 现代汉语虚词研究与对外汉语教学 (第二辑) [C]. 上海: 复旦大学出版社, 2008, 187-203.

作者联系方式: 陈振宇 上海市邯郸路 220 号复旦大学中国语言文学系 200433
13585693648 chenzhenyu@fudan.edu.cn