

自然语言中时间信息的模型化*

郭宏蕾 姚天顺

(东北大学计算机系 沈阳 110006)

摘要 在自然语言理解中,时间是一个重要的语境因素.本文提出一种独立于句子表层形式的多层次时间语义结构,浅层语义结构是时间描述的量化语义表示,深层语义结构描述事件的动态属性和存在特征.该时间语义模型能表示时刻和时段,将时间基点明确区分为物理时间基点和说话者时间基点,并提供通用的时间语义计算方法,将各语言的时间描述映射到时间轴上.在该模型基础上,从语义观点出发,建立时、体的可计算模型及各事件时间相关性计算模型.

关键词 自然语言理解,时间,语境,语义,多语翻译.

中图分类号 TP18

人的交流总是依赖于一定的时间语境,时间是人类共有的普遍概念,具有客观性和可比性.虽然,人们的时间观念具有统一性.但时间观念的描述却因语言的多样性而复杂化.时间轴及通用计时单位的引入,使所有时间描述都可映射为时间轴上的一点或一区间.这为研究统一的时间语义结构提供了基础.

在多语翻译中,为了准确理解句子的时间描述,分析时应方便完整地记录其时间语义,生成时应比较准确地再现源时间语义及表述风格.在情景处理中,我们要计算时间语境的语义,利用时间的可比性,确立各事件的时间相关性.在自然语言中时间处理主要涉及2个问题:(1)时间的内部表示,即时间结构的量化表示;(2)时间信息的流动和变换,即情景中时间语义的理解和推理.本文以真实世界(即已发生、正发生和将要发生的情景)中的时间信息为研究对象,排除了情态、意愿、否定及频度上下文这类时间信息.我们将时间知识和非时间知识相分离,建立了独立于句子表层形式的多层次时间语义结构,量化时间实体的浅层语义,描述事件的动态属性和存在特征等深层语义.同时,本文提供通用的时间语义计算方法,将各语言的时间描述映射到时间轴上,从而清晰地解释动词时态、时间副词和连词的组合,以便于计算复句的复杂时间语义及情景中各事件的时间相关性.

1 语言的时间语义结构

1.1 时间实体及量化表示

* 本文研究得到国家自然科学基金和国家863高科技项目基金资助.作者郭宏蕾,女,1970年生,博士,主要研究领域为计算语言学,人机系统.姚天顺,1934年生,教授,博士导师,主要研究领域为计算语言学,人机系统.

本文通讯联系人:郭宏蕾,北京100083,北京航空航天大学计算机系

本文1996-08-02收到修改稿

在自然语言中,时间常用时间片语或时间子句来描述,它们定位了时间轴上的一点或一个区间.具有时序性的简单时间片语称为时点实体,其具有前驱和后继,如“3 月份”的前驱(即上个月)为“2 月份”,“今天”的后继时点为“明天”;表示时间流逝量的时间片语称为时段实体,如“三个月”等.时点实体和时段实体统称为时间实体.为了量化一个时间实体,我们依据通用的科学计时单位,建立了时间实体的层次结构.每层以计时单位命名,时间实体的量值依据其依附的计时单位,分别归属于相应的时间层.因此,一个时间实体的量化表示可能需要多个计时单位(即由多个不同时间层上的量值合成),使用的计时单位的个数称为关注数.时间实体的量化值简称为时间量值,在句子中可采用不同的描述形式.根据其到时间轴上的映射方式,我们划分了时间量值指称的类型.在单值映射中,能直接在时间轴上定位的时间实体(如“1995 年 3 月”)为特指类型;象时间指代词(如“今天”)这类需计算才能定位的时间实体为泛指类型.泛指时间实体是主观的,与说话者有关.在计算泛指时点的坐标值时,我们以说话时刻为基点,引入了前驱函数和后继函数.象时频指代词(如“每天”)这类在时间轴上为多值周期性映射的时间实体为时频类型.时间量值可形式化为下述元组序列:

时间量值: := {⟨ $U_k i$, 指称类型, 数值特征, 数值域⟩}

n 为关注数; U_k 为 k 层上的计时单位, i 为 U_k 在本量化描述中的顺序号, $i=1, 2, \dots, n$; 特指型时间实体的数值特征为序数或数量, 用于指明数值域取值是时间序数还是时间数量. 泛指型时间实体的数值特征为前驱或后继, 其数值域为相对于计算基点的前驱或后继次数. 频指型时间实体的数值特征为周期, 其数值域为时间周期. 由于频指型时间描述是多值映射, 我们暂不予以考虑. 显然, “前年 12 月” = {⟨年, 泛指, 前驱, 2⟩⟨月, 特指, 序数, 12⟩}.

综上所述, 在多种语言翻译中, 借助于上述方法, 分析时我们可以准确记录时间描述信息及描述风格, 将采用不同描述风格的同一时间信息定位到时间坐标轴的同一时间区域^[1], 生成时可以给出原时间量值的准确译文, 并再现原文的描述风格.

1.2 表层语义结构

在真实世界中, 若简单句中陈述的活动或事实称为一个事件, 事件从发生到结束的时间则称为事件的存在时间, 句子中伴随事件而提及的时间描述则称为事件的参照时间(记为 RT), 它与一个显式或隐式的时间描述相关联. 句子通常只能提供事件的参照时间, 事件的存在时间一般需要通过参照时间、语法范畴(如动词的时态、体)和词汇语义的时间约束(如表示状态、过程的动词的时间顺序)间接推出.^[2] 在 1.2 节, 我们首先研究 RT , 在 1.3 节, 将考察事件的存在时间区及其与 RT 的关系.

现在, 我们考察一些含时间限定词(如“在…之前”、“在…之后”、“从…到…”等)的复杂时间片语. 这类时间描述通常具有时间坐标平移功能, 需要确定时间相对原点, 即时间描述基点, 简称为时间基点 TB . TB 由说话人选定, 可以为情景缺省的计时原点(记为 T_0 , 如公元元年)或说话时刻(记为 TS). TB 的引入可简化事件参照时间的描述, 在具体时间描述的量化过程中, 首先确定 TB , 然后, 在 TB 基础上, 进一步确定时间量及其与 TB 的位置关系. 这里的时间量是指句子中直接提供的时间描述信息. 在一个简单句中, 时间量可用序偶⟨时间量值 t , 增量值 Δt ⟩表示, 简记为 $(t, \Delta t)$. t 是相对于 TB 的时间偏移量, Δt 是相对于 t 的时间增量.

时间量与时间基点在时间轴上的位置关系由定位算子描述. 基本定位算子有 2 个: 前移

算子 B^* , 语义为“时间量定位在 TB 之前”; 后移算子 A^* , 语义为“时间量定位在 TB 之后”; 定位算子把由 $(t, \Delta t)$ 表示的时间量映射为时间轴上的一个具体区间 $[R', R]$ ——事件的参照时间(约定 $R' \leq R$), 即 $RT = [R', R]$. 基本算子能访问时间轴上的任一点或区间, 准确地记录各类时间描述. 现在, 我们给出时间描述的表层语义结构 STS(surface temporal semantics):

$STS ::= \langle \text{事件号}, \text{事件依存点}, \text{说话时刻}, \text{时间基点}, \text{基本定位算子}, \text{时间量}, \text{参照时间区} \rangle$

在一个情景中, 我们为各事件建立统一的序号(即事件号), 用以唯一标识其对应的 STS 结构. 事件依存点描述该事件在时间上直接依存的相关事件号, 主要用于记录复句中的依存事件(第 3 节详述). 时间基点的选取与时间量相关, 若时间量为绝对时间量(如“今年 3 月份(以前)”), 则 $TB = T0$; 若时间量为相对时间流逝量(如“3 个月(以后)”), 则 $TB = TS$.

下面, 介绍参照时间区 $[R', R]$ 的计算方法. 参照时间区 $[R', R]$ 的引入是为了准确理解情景中各事件的时间相关性. 设 O' 代表时间基点 TB 在时间轴上的坐标. TS 表示说话时刻, Tn 表示近 O' 点, Tf 表示远 O' 点, OP 代表定位算子, 则句子提供的时间描述(即 $\langle \text{时间基点}, \text{基本定位算子}, (t, \Delta t) \rangle$) 在时间轴上对应的参照时间区 $[R', R]$ 的确定遵循下述准则.

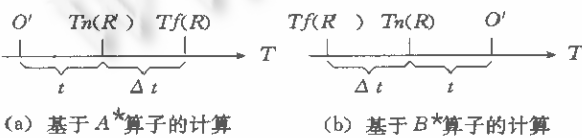


图1 参照时间区计算示意图

准则. 给定时间描述 $\langle O', OP, (t, \Delta t) \rangle$, 其参照时间区 $[R', R]$ 的计算为 (1) 后移运算: $OP = A^*, Tn = O' + t, Tf = Tn + \Delta t; [R', R] = [Tn, Tf]$ (见图 1(a)). (2) 前移运算: $OP = B^*, Tn = |O' - t|, Tf = Tn - \Delta t; [R', R] = [Tf, Tn]$ (见图 1(b)).

显然, $\langle O', OP, (t, \Delta t) \rangle$ 的语义为“从 R' 到 R 的区间, 并且 $[R', R]$ 在 TB 之后(或之前)”. 特别地, $\langle O', OP, (t, @) \rangle$ 表示“ Tn 点之后(或之前)”, $@$ 为 Δt 取缺省值, 即 ∞ . 此时 $Tf = -\infty$ 或 $+\infty$, 即只关心 Tn 点. 例如, “今年 3 月份以前” = $\langle T0, B^*, (t = \text{“今年 3 月份”, @}) \rangle$. $\langle O', OP, (t, 0) \rangle$ 表示“在 $Tn(Tf)$ 点上”. $\langle TS, OP, (0, \Delta t) \rangle$ 表示“在 Δt 段时间之后(或之前)”, 例如, $\langle TS, B^*, (0, 3 \text{ 个月}) \rangle$ 表示“3 个月以前”. 其他各种时间定位算子均可由基本算子 A^* 和 B^* 生成, 例如, WHILE 算子运算的语义为“从 $t1$ 到 $t2$ ”, 即 $\langle O', \text{WHILE}, (t1, t2) \rangle \Leftrightarrow \langle O', A^*, (t1, @) \rangle \wedge \langle O', B^*, (t2, @) \rangle \wedge t1 < t2$, 对应的参照时间区间 $RT = [t1, t2]$.

有时句中缺乏明显的时间描述, 而使用语法或特征词等语言学线索来补充欠缺的时间信息. 例如, 英语“*He finished the work.*”中, 用动词的“时”结构表述过去时间信息, 汉语“*他将做这个工作.*”中, 用时态特征词“*将*”表述将来时间信息. 这类缺省时间描述的参照时间区可如下计算: 设说话时刻 = TS , 由于“现在”不是一个绝对值, 必须定义为与说话时刻相关的概念, 因此, 用邻域 (TS, ϵ) 表示人们习惯的现在时间区域^[3], 从而, (1) 若时间语义为“在说话时刻之前”, 则量化为 $\langle TS, B^*, (\epsilon, @) \rangle$, 即 $RT = (-\infty, TS - \epsilon)$; (2) 若时间语义为“在说话时刻之后”, 则量化为 $\langle TS, A^*, (\epsilon, @) \rangle$, 即 $RT = (TS + \epsilon, +\infty)$; (3) 若时间语义为“在说话时刻之时”, 则量化为 $\langle T0, \text{WHILE}, (TS - \epsilon, TS + \epsilon) \rangle$, 即 $RT = [TS - \epsilon, TS + \epsilon]$. 此外, 对于缺乏表示时间信息的词汇和语法范畴的句子, 如汉语“*我学习英语*”, 其缺省时间语义为

$\langle T0, \text{WHILE}, (TS - \epsilon, TS + \epsilon) \rangle, RT = [TS - \epsilon, TS + \epsilon]$. 可见,在情景处理中,只有在时间轴上确定各事件的参照时间区间 $[R', R]$,才可能找出各事件的时间顺序.

1.3 深层语义结构

在语言表达式中,时间语境主要涉及:(1)说话时刻:是独立于语言表达式的固有客观信息;(2)事件的参照时间区 $[R', R]$;(3)事件的存在特征.这三要素常由谓词语义、谓词语法形态(时态、完成体、进行式)及时间副词描述,三者相互约束,构成时间的深层语义.

在各种语言中,说话者常以当前说话一刻为基点,将时间划分为过去和将来 2 部分.^[4]反映已发生、正发生或将要发生等概念.对于不同的说话者,现在、过去和将来是动态确定的.我们将这些特征称为事件的动态属性(记为 E_dym).事件的动态属性与说话时刻 TS 和参照时间 RT 有关,其确定方法如下:(1)若 RT 在 TS 之前,则 $E_dym = PAST$,即过去时间属性;(2)若 RT 在 TS 之后,则 $E_dym = FUTURE$,即将来时间属性;(3)若 RT 塌陷为 TS 点,则 $E_dym = PRESENT$,即现在时间属性;

一般地, RT 在句子中可有如下 3 种表现形式:(1)直接表示,如物理时间(“95 年 3 月”)和时间副词(now, yesterday)等.物理时间将 RT 直接定位于时间轴上,而时间副词则将 RT 同一个隐含的时间(即说话时刻)相关联.(2)间接表示,如复句中主句的 RT 可通过时间连词及从句的 RT 表示.^[5](3)隐含表示,如谓词的语法形态(aspect, tense).因此,事件的动态属性既可借助于语法手段或特征词获取,又可由时间描述的具体坐标值确定.在分析时,如果有明确的“时”结构,则可直接识别事件的动态属性.例如,在英语“I did the work.”中,由动词“do”为“过去时”可知 $E_dym = PAST$.当没有明确的“时”结构时,则需计算 RT 与 TS 的相对位置,以识别 E_dym .例如,汉语句子“今年二月份我去美国”没有明确的“时”标识,此时,我们必须计算 RT “今年二月份”的物理时间,由此判断 E_dym :(1)若 $TS = 1995$ 年 1 月 14 日,则 $RT = 1995$ 年 2 月,因 RT 在 TS 之后,故 $E_dym = FUTURE$.从说话者角度看,事情将要发生.在英语生成时,属性 $FUTURE$ 映射成动词的将来时结构.(2)若 $TS = 1995$ 年 3 月 14 日,由 RT 在 TS 之前,得 $E_dym = PAST$.从说话者角度看,事情已发生.这样,在英语生成时,属性 $PAST$ 映射成动词的过去时结构.

事件的发生、发展和影响是事件的存在方式,人们透过 RT 区间观察特定情景,可以感知事件的存在.我们将能观察到的状态或过程的持续时间称为事件时间(记为 ET).^[6]这与前面提到的事件存在时区不完全相同, ET 是存在时区的一个子时段,由于 ET 的边界是观测边界,因此, ET 的边界具有可延展性,由〈左延展性,右延展性〉描述,简记为〈 L_exp, R_exp 〉,且 $L_exp, R_exp \in [0, 1]$,取值为 0,表示边界已确定,不再具有延伸能力.取值为 1 表示无边界,可无限延展.取值在 $(0, 1)$ 区间,则表示有边界但未完全确定. ET 边界的延展性与谓词的时间约束有关.当〈 L_exp, R_exp 〉=〈0, 0〉时,表示事件的存在时间与事件时间是相同的.这样,事件的存在特征可由〈事件类型,影响相,持续相,延展性〉描述.事件分为状态和过程 2 类.^[7]若关注事件的影响性,则影响相取值为影响子,否则记为 *ignore*.若关注事件的持续性,则持续相取值为持续子,否则记为 *ignore*.给定事件的参照时间 $RT = [R', R]$,有如下一些定义:

定义 1. 存在子: $\exists t, t \in [R', R]$, 在 t 时刻,事件存在,记为 *EXIST*.

定义 2. 影响子: $\exists I, I = [R' - \delta_1, R' + \delta_2]$, 其中 $\delta_1, \delta_2 > 0$ 且 $R' + \delta_2 \leq R, \forall t, t \in I$, 在 t

时刻,可观察到事件 E , 记为 $EFFECT$.

定义 3. 持续子: $\exists I, I = [R' + \eta, R], \eta \geq 0, \forall t, t \in I$, 在 t 时刻, 可观察到事件 E , 记为 $KEEP$.

事件的影响性和持续性反映了事件的活性, 是事件的内在结构. 如果讲述者既不关注事件的影响性, 又不关注事件的持续性, 则陈述的只是事件的存在. 因此, 当 $\langle \text{影响相}, \text{持续相} \rangle = \langle \text{ignore}, \text{ignore} \rangle$ 时, 只说明在 RT 内可观察到事件 E (即定义 1), 对事件的活性则不作进一步描述, 英语中一般体的语义与此近似.

影响子 (即定义 2) 强调在 RT 之前, 事件 E 就可被观察到, 即在 RT 之前, 事件已开始对现实世界产生影响. 英语的完成体、汉语的时态副词“已经”的时间语义与之近似. 例如, 句子 $S1$ “John has read a book.”, $S2$ “John has walked for two hours.”, $S1$ 和 $S2$ 都关注事件的影响性, 由 $S1$ 的陈述, 可推出 $ET1 = [R' - \delta1, R' + \delta2]$, 其中, $\delta1 > 0, \delta2 = 0$, 说明事件 $E1$ 在 RT 前被观察到, 且在 RT 前已完成 (见图 2(a)). 由 $S2$ 的陈述, 可推出 John 在两小时前开始走, 现在或者继续或者停止, 即 $ET2 = [R' - \delta1, R]$, 其中 $\delta1 > 0, \delta2 = R - R'$ (见图 2(b)).

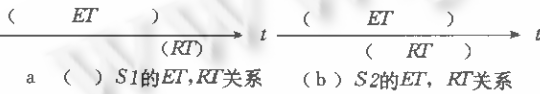


图2 事件的影响性

持续子 (即定义 3) 强调在 RT 的子时段内, 事件一直在持续. 英语的进行式、汉语的时态副词“正在”等的的时间语义与之近似. 例如

“He is reading a book.” 中, 事件的存在特征中含有持续子, 则在汉语生成时, 将持续子映射成时态副词“正在”.

当影响子和持续子共现时, 强调在 RT 以前, 事件就可被观察到, 而且在 RT 的子时段内, 事件一直在持续. 英语的复合时态完成进行式的时间语义与之近似.

由上可见, 英语中的时、体反映的是不同的 2 个时间概念, 时由说话时刻定义; 体则借助于参照时间定义, 且依赖于陈述者对事件的观察. 虽然, 汉语中没有明确的时、体语法形态, 但与时、体相对应的的时间语义描述却是汉语表达中必不可少的.

综上, 时间语境三要素间的深层关系组成深层时间语义 DTS (deep temporal semantics): $DTS ::= \langle \text{事件的动态属性}, \text{事件的存在特征} \rangle$

事件的动态属性说明参照时间与说话时间的关系. 事件的存在特征描述事件时间与参照时间、存在时区的关系. 通过 DTS 结构, 可计算出事件时间、参照时间、说话时间三者间的关系, 从而生成谓词的时态结构. 因此, 在英汉双向机器翻译中, 从语义的观点出发, 以 DTS 为中介结构, 可完成英汉 2 种时态体系结构的变换, 即

$$\text{时态体系结构(英)} \Leftrightarrow DTS \text{ 结构} \Leftrightarrow \text{时态体系结构(汉)}$$

2 事件的时间相关性

2.1 事件时间依存关系及复句的时间语义结构

在情景理解中, 任意两事件的时间关系可以用事件发生或结束的先后关系来刻画. 这种时间关系常由显式的因果信息或先后顺序给出, [8] 当两个事件的存在时区是确定的时, 事件间的时间关系可由下述定义得出.

定义 4. 若事件 $e1$ 的存在时间为 $[T1', T1]$, 事件 $e2$ 的存在时间为 $[T2', T2]$, 两事件的

开始顺序用 $Start(e_1, e_2)$ 表示, 则

$$Start(e_1, e_2) = \begin{cases} AFTER & \text{iff } T_1' > T_2', \text{ 即 } e_1 \text{ 后于 } e_2 \text{ 开始;} \\ BEFORE & \text{iff } T_1' < T_2', \text{ 即 } e_1 \text{ 先于 } e_2 \text{ 开始;} \\ WHEN & \text{iff } T_1' = T_2', \text{ 即 } e_1 \text{ 与 } e_2 \text{ 同时开始;} \\ UNKNOWN & \text{iff } T_1' \text{ 与 } T_2' \text{ 的关系未知;} \end{cases}$$

同样地, 通过对两事件存在时间的结束点 T_1, T_2 进行比较, 可以定义两事件的结束顺序为 $End(e_1, e_2)$ (定义从略).

显然, 在没有显式的时序说明或无法确定事件存在时区的情况下, 我们不能确定两事件发生或结束的先后, 只能借助于事件时间(即事件被观察的时间)判定两个事件谁先被观察到, 为此, 引入下述定义.

定义 5. 若事件 e_1 的事件时间为 $[E_1', E_1]$, 事件 e_2 的事件时间为 $[E_2', E_2]$. 两个事件被观察到的先后顺序用 $Obs(e_1, e_2)$ 表示, 则

$$Obs(e_1, e_2) = \begin{cases} AFTER & \text{iff } E_1' > E_2', \text{ 即 } e_1 \text{ 后于 } e_2 \text{ 被观察到;} \\ BEFORE & \text{iff } E_1' < E_2', \text{ 即 } e_1 \text{ 先于 } e_2 \text{ 被观察到;} \\ WHEN & \text{iff } E_1' = E_2', \text{ 即 } e_1 \text{ 与 } e_2 \text{ 同时被观察到;} \\ UNKNOWN & \text{iff } E_1' \text{ 与 } E_2' \text{ 的关系未知;} \end{cases}$$

由前面的叙述, 我们知道, 事件时间可以由参照时间确定. 若两个事件的参照时间区间是确定的, 则可以推出事件时间的区间. 这时, 通过比较两事件时间的开始点, 可以判断出哪一事件被先观察到. 需要指出的是, 事件“先于发生”和事件“先于被观察到”是 2 个不同的概念, 但我们经常将其混同. 事实上, 只有当事件的存在时间与事件时间是同一时间区间时, 二者才是一致的, 至于二者的更深关系, 则属于时间哲学范畴, 超出本文研究范围.

因此, 2 个事件 e_1, e_2 的时间依存关系可由元组 $\langle Start(e_1, e_2), End(e_1, e_2), Obs(e_1, e_2) \rangle$ 表示. 若元组中有一项的值不为 UNKNOWN, 则称两个事件具有时间依存关系, 在情景理解中, 我们用事件依存表 EDT(event dependency table) 记录各事件的时间相关性, EDT 的信息结构为 \langle 关注事件, 依存事件, 时间依存关系 \rangle , 其中关注事件指情景中当前谈论的事件, 依存事件指关注事件在时间上所依存的事件. 若将事件时间依存关系表中的事件看成结点, 用附加标签的有向线段 $\xrightarrow{\text{label}}$ 联结两个具有时间依存关系的事件点. 则构成情景中各事件间的时间关系网络. 其中 $label \in \{START, END, OBS\}$, 用于标注 $Start(e_1, e_2), End(e_1, e_2), Obs(e_1, e_2)$ 三类时间依存关系. “ \rightarrow ”由时间在前的事件结点指向时间在后的事件点. 若两个事件点具有同时性, 则用“ \leftrightarrow ”联接. 时间关系网络直观形象地描述了情景中各事件间的时间相关性. 例如, “ $e_1 \xrightarrow{START} e_2$ ”的语义为“ $Start(e_1, e_2) = BEFORE$ ”, “ $e_1 \xleftarrow{END} e_2$ ”的语义为“ $End(e_1, e_2) = AFTER$ ”; “ $e_1 \leftrightarrow e_2$ ”的语义为“ $Obs(e_1, e_2) = WHEN$ ”.

复句的时间关系是最简单的情景时间语境, 是分析复杂时间语境的基础. 复句的主句和从句构成两个事件, 具有各自的时间结构, 且二者存在着一定的时间依存性. 复句中的时间依存关系主要借助时间连词(如 after, before, when 等)进行描述, 这种时间关系首先体现在 $Obs(e_1, e_2)$ 上, 其中, 主句事件 e_1 为关注事件, 从句事件 e_2 为依存事件. $Obs(e_1, e_2)$ 的值与时间连词的语义相对应, 例如, 复习“ I finished the work after you came back.”中, 若“ you

came back. ”事件号为 e_2 , “I finished the work. ”事件号为 e_1 , 则 $Obs(e_1, e_2) = AFTER$.

在复句中, 由于时间连词的约束作用, 主句事件的说话时刻依存于从句事件的说话时刻.^[9]若主句事件 e_2 的说话时刻记为 $TS(e_2)$, 从句事件的说话时刻记为 $TS(e_1)$, $TS(e_2)$ 与 $TS(e_1)$ 的习惯时间差记为 $\delta(\delta > 0)$, 则主句事件的说话时刻 $TS(e_2)$ 由下面方法确定:

$$TS(e_2) = \begin{cases} TS(e_1) + \delta & \text{iff } Obs(e_2, e_1) = AFTER; \\ TS(e_1) - \delta & \text{iff } Obs(e_2, e_1) = BEFORE; \\ TS(e_1) & \text{iff } Obs(e_2, e_1) = WHEN; \end{cases}$$

复句中时间连词 after, before, when 与前节的定位算子 B^* 、 A^* 具有不同的语义. 复句中事件的参照时间的计算仍遵循上节准则, 不受事件间的时间依存关系的影响. 在时间语境的语义处理中, 主从事件说话时刻的时间差值 δ 、现在时间邻域 (TS, ϵ) 中的常量 ϵ 以及缺省计时原点 T_0 与情景中的具体时间描述无关. 我们将其称为时间语境的环境变量 ENV . 更复杂的情景, 涉及到更多的语义学知识, 本文不再叙述.

2.2 举例

我们以下述文本为例, 提取事件的时间语义层次结构, 并考察事件的时间相关性.

Yesterday, I went to school by bus at seven. The traffic had been very heavy since six o'clock. The exam had started before I got the school. My teacher was annoyed at my absence when I entered the classroom at nine. I sat for the examination after I told him the reason.

语义分析部件将该文本划分为如下 8 个事件: $\{e_1: go(I, school); e_2: be(traffic, heavy); e_3: start(exam); e_4: get(I, school); e_5: annoy(teacher); e_6: enter(I, classroom); e_7: sit(I, examination); e_8: tell(I, teacher)\}$. 时间分析器对各事件进行时间分析, 即:

(1) 根据文本提供的时间数据(即 $\langle O', Loc_op, (t, \Delta t) \rangle$), 计算各事件的时间区域. 若 $TS = 1996/7/18$, 则该文本的缺省计时原点 $T_0 = 1996/7/17$. 事件 e_i 的时间语义结构记为 $Tsem(e_i)$, 存在时间记为 $t_i, i = 1, \dots, 8$. “—”表示空项, 则:

$$Tsem(e_1) = \{\langle e_1, -, Ts(e_1), Ts(e_1), WHILE, ("seven\ yesterday", 0), [96/7/17/7:00, 96/7/17/7:00] \rangle, \langle PAST, PROCESS, ignore, ignore, (0, 0) \rangle\}$$

$$Tsem(e_2) = \{\langle e_2, -, Ts(e_2), T_0, A^*, ("six\ o'clock", @), [96/7/17/6:00, +\infty] \rangle, \langle PAST, STATE, EFFECT, KEEP, (0, 0.5) \rangle\}$$

$$Tsem(e_3) = \{\langle e_3, e_4, Ts(e_4) - \delta, T_0, B^*, (t_4, @), (-\infty, t_4) \rangle, \langle PAST, PROCESS, EFFECT, ignore, (0.5, 0.1) \rangle\}$$

$$Tsem(e_4) = \{\langle e_4, -, Ts(e_4), Ts(e_4), B^*, (\epsilon, @), (-\infty, Ts(e_4) - \epsilon) \rangle, \langle PAST, PROCESS, ignore, ignore, (0.1, 0.1) \rangle\}$$

$$Tsem(e_5) = \{\langle e_5, e_6, Ts(e_6), T_0, B^*, (t_6, @), (-\infty, t_6) \rangle, \langle PAST, STATE, ignore, KEEP, (0.5, 0.5) \rangle\}$$

$$Tsem(e_6) = \{\langle e_6, -, Ts(e_6), T_0, WHILE, ("nine\ o'clock", 0), [96/7/17/9:00, 96/7/17/9:00] \rangle, \langle PAST, PROCESS, ignore, ignore, (0, 0) \rangle\}$$

$$Tsem(e_7) = \{\langle e_7, e_8, Ts(e_8) + \delta, T_0, B^*, (t_8, @), (-\infty, t_8) \rangle, \langle PAST, PROCESS, ignore, ignore, (0.5, 0.5) \rangle\}$$

$$Tsem(e_8) = \{\langle e_8, -, Ts(e_8), Ts(e_8), B^*, (\epsilon, @), (-\infty, Ts(e_8) - \epsilon) \rangle, \langle PAST, PROCESS, ignore, ignore, (0.5, 0.5) \rangle\}$$

此时, 各事件在时间轴上的相对位置如图 3 所示.

(2) 依据事件的语义及时间约束, 计算各事件间的时间相关性, 即: (a) 在 $\{e_3, e_4\}, \{e_5, e_6\}, \{e_7, e_8\}$ 中, 每对事件均共现于同一复句, 具有显式直接依存关系; (b) $\{e_1, e_4, e_6, e_8\}$ 是同一主体发出的动作序列, $\{e_2, e_1\}$ 含有显式的时间参照区, 故它们具有很强的时序性. 图 4 描述了例 1 中各事件的时间相关性. 显然, 借助时间区域和依存关系的传递性, 我们可获取许

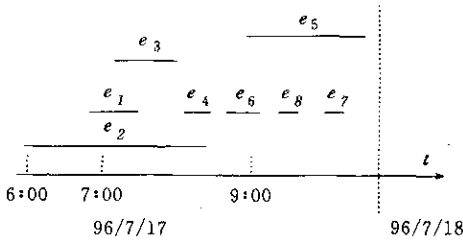


图3 例1的事件时间区域示意图

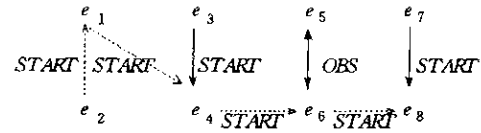


图4 例1的事件间的时间关系图

多事件间隐含的时间相关性,例如,由 e_3, e_4, e_6 间的时间依存关系,可推出 e_3 先于 e_6 发生,即考试在九点(即 t_6)以前开始。

3 时间语境的层次语义结构

综上,本节给出情景级时间语境的完整层次语义结构 TCS (temporal context semantics):

- TCS::={环境变量 ENV,事件时间依存关系表 EDT,所有事件的时间语义结构}
- ENV::={ δ, ϵ, T_0 }
- EDT::={ $\langle e_2, e_1, DEP(e_2, e_1) \rangle \mid \exists e_1, e_2 \in \text{事件集}, e_2 \text{ 为关注事件}, e_1 \text{ 为依存事件}, DEP(e_2, e_1) \text{ 为时间依存关系}$ }
- 时间语义结构::={表层语义 STS,深层语义 DTS}
- STS::={事件号,事件依存点,说话时刻,时间基点,基本定位算子,时间量,参照时间区}
- DTS::={事件动态属性,事件存在特征}
- 参照时间区::={ $[R', R] \mid [R', R] = f(\text{时间基点}, \text{时间量}), f \text{ 基本定位算子}$ }
- 时间量::={ $(t, \Delta t)$ }
- 时间量值::={ $\langle Uki, \text{指称类型}, \text{数值特征}, \text{数值域} \rangle^n, Uki, n, i \text{ 含义同第1节}$ }
- 事件存在特征::={事件类型,影响相,持续相,延展性}
- DEP(e_2, e_1)::={ $\langle Start(e_2, e_1), End(e_2, e_1), Obs(e_2, e_1) \rangle$ }
- 时间基点::={缺省计时原点 $T_0 \mid$ 说话时刻 TS }
- 基本定位算子::={前移算子 $B^* \mid$ 后移算子 A^* }
- 事件动态属性::={PRESENT | PAST | FUTURE}
- 事件类型::={STATE | PROCESS}
- 影响相::={EFFECT | ignore}
- 持续相::={KEEP | ignore}
- 延展性::={ $\langle L_exp, R_exp \rangle \mid L_exp, R_exp \in [0, 1]$ }
- 指称类型::={特指 | 泛指 | 频指}
- 数值特征::={数量 | 序数 | PRE | SUC | 周期}

4 结 语

在国家 863 高科技项目“汉英双向翻译系统 CETRAN2”中,我们增强了时间语义的处理功能,在句中每个谓词元素的语义分解建立之后,时间部件开始执行分析,利用本文的多层时间语义结构表达时间语境,为多语机器翻译和进一步的文本理解提供准确的时间语义。

参考文献

- 1 郭宏蕾,姚天顺. 时间语义的理解. 中国青年计算机研究新进展,第六届全国青年计算机会议,杭州:浙江大学出版社,1996.
- 2 Rebecca J Passonneau. A computational model of the semantics of tense and aspect. Computational Linguistics, 1988,14(2):44~60.
- 3 Bernard M. The presentation of linguistic information in an approach used for modelling temporal knowledge in

- discourses. In: Proc. of the 3rd ICCS, 1993.
- 4 奥托·叶斯柏森. 语法哲学. 北京: 语文出版社. 1988.
- 5 Chung H. Interpreting tense, aspect and time adverbial; a compositional, unified approach. In: Proc. of the 29th ACL, 1991.
- 6 Fei Song *et al.* Improving heuristic-based temporal analysis of narratives with aspect determination. In: Proc. of IJCAI, 1995. 1338~1345.
- 7 Rebecca J. Passonneau, situations and intervals. In: Proc. of the 25th of ACL, 1987. 134~137.
- 8 Irene Piments. Temporal structure of discourse. In: Proc. of the Coling-92, Nantes, 1992. 331~337.
- 9 Moens M, Steedman M. Temporal ontology and temporal reference. *Computational Linguistics*, 1988, 14(2):15~28.

MODELING TEMPORAL SEMANTICS INFORMATION FOR NATURAL LANGUAGE

GUO Honglei YAO Tianshun

(*Department of Computer Science Northeastern University Shenyang 110006*)

Abstract In natural language processing, temporal is an important situation. In order to understand temporal situation in the text, this paper presents a hierarchical temporal semantic structure independent of the surface description in sentences. The surface semantic structure formalizes the description of temporal situation; The deep semantic structure describes the dynamic attributes and the existing characteristic of the event in a given temporal situation. This temporal semantic representation model allows to express point in time as well as time intervals. It provides means to indicate time basic point with respect to “an absolute physical time”, as well as time basic point relatively to the speaker’s time. A universal calculation method of the temporal semantics provided in this paper localizes the temporal entities in all languages into the temporal axis. This enables to interpret combinations of verb tense, temporal adverbs and connective. In the meanwhile, within this model some constructs enable to express tense, aspectual properties of verbs as well as the semantic calculation of temporal relationship among events in the text.

Key words Natural language processing, temporal, situation, semantics, multilingual machine translation.

Class number TP18