

自動プログラミングにおける自然言語仕様から オペレーションを中心とした表現への変換

白井 勝利† 山田 宏之‡ 馬場 口 登† 手塚 慶一†
†大阪大学 工学部 ‡愛媛大学 工学部

自動プログラミングシステムにおいて、ユーザの記述のしやすさ、読みやすさの点から、自然言語仕様を受理できることが望まれている。そのためには、仕様に含まれる意味を形式的な表現に変換する必要がある。そこで我々は、仕様の各文の意味を表現するための枠組みとして、対象領域における基本的な操作を実現する上で必須の情報が直接記述できる、オペレーションフレーム(OF)を導入する。仕様記述における情報の欠落は、OFにおける必須情報の欠落に対応するので、OFの集合に対して対象領域のオペレーションに関する知識を適用することにより、単語レベル、及び文レベルの欠落情報が容易に補完できる。本稿では、対象領域を待ち行列シミュレーションに設定し、OFが自然言語仕様に含まれる情報の一表現形式として有用であることを明らかにする。

Transformation of Natural Language Specification into Operation-centered Frame in Automatic Programming

†Katsutoshi SHIRAI, ‡Hiroyuki YAMADA, †Noboru BABAGUCHI and †Yoshikazu TEZUKA

†Faculty of Engineering, Osaka University, 2-1 Yamadaoka, Suita-shi, 565, JAPAN

‡Faculty of Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama-shi, 790, JAPAN

It is desirable to realize an automatic programming system that can deal with natural language specifications in order to improve users' facility. Such a system must transform the information in the specification into the formal representation. For this purpose, we introduce an useful scheme for representation, called an operation-centered frame (OF). An OF represents the prerequisites to perform one basic operation in the domain. Since ellipsis in the specification corresponds to that in the OF, a series of OFs enables us to complement various ellipsis for words or sentences. In this paper, focusing on queueing simulations, we demonstrate that OFs provide the versatile functions in automatic programming.

1. まえがき

ソフトウェアの生産性を向上させる一手段として、自然言語仕様を受理できる自動プログラミングシステムの開発が挙げられる。自然言語による仕様記述には、人間が日常的に行っている思考過程を反映させやすく、また、記述の簡便さ、直感的な理解容易性等、他の手段による表現にはない特徴があり、記述効率が向上する¹⁾。しかし、これらの特徴は仕様に不完全性を許すことになる。ソフトウェア開発の下流工程において、この不完全性が検出されるとそれまでの作業を無効にすることがあり、開発効率が低下する。したがって、仕様中の不完全性はできるだけ早い段階で検出し、排除する必要がある。

従来から、自然言語仕様中の不完全性を排除する方法として、対象領域固有の知識を利用する方法があるが、対象領域固有の知識の表現形式、及び利用法が明確でない。そこで本研究では、対象領域の知識を“一つの機能を実現するために必要とされる情報”とし、それを表現するための枠組み（記述形式）を考察する。本記述形式により記述された情報には対象領域に現れる基本的操作を実現する上で必須の情報が直接的に記述されるため不完全性のうち情報の欠落に対処できる。

本稿では、具体的な対象領域を待ち行列シミュレーションとし、①自然言語仕様からの自動プログラミングシステムの概要、②対象領域の操作に対する必要情報の記述形式、③仕様中に現れる欠落情報の分類、及び各々に対する補完の方法、④ソフトウェアモデルに関する考察、⑤システムの動作例を述べる。

2. システムの概要²⁾

本研究で考察するシステムの構成を図1に示し、各処理モジュールの概略、及び自然言語仕様記述に課し

た制約を以下に述べる。

2. 1 自然言語処理部

自然言語処理部では、動詞を中心とした解析に適した格文法に基づいて要求仕様記述中の各文を処理し、以下の3つの操作を行う。

- ①文法に従って要求仕様を検査することによる自然言語仕様に含まれる構文的誤りの排除。
- ②一般的な世界の知識を適用することによる一般的な意味の誤りの排除。
- ③動詞をフレーム名とする表層的な格フレーム構造（以下、文フレームと呼ぶ）への変換。

システムの構成上は、自然言語処理部とモデル形式化部の変換過程は一つのモジュールとして統合することが可能である。しかし、自然言語処理の研究で得られる新しい成果を自然言語処理部に反映するために、構文上の曖昧性、多義性の解決など自然言語の知識に依存する部分を自然言語処理部とし、対象領域の知識に依存する部分と分離している。

2. 2 モデル形式化部

モデル形式化部では、モデル形式化知識ベースにある対象領域におけるオペレーションに関して体系化された知識を利用して以下の3つの操作を行う。

変換：ソフトウェアの仕様に対する表層的な記述である文フレームの集合から、個々の文フレームに対応した対象領域に依存する意味を表すフレーム（オペレーションフレーム、以下OFと略記）の集合に変換する。

補完：OFの集合内の欠落情報を補完し、完全なOFの集合にする。

構造化：OFの集合を形式モデルに変換する。

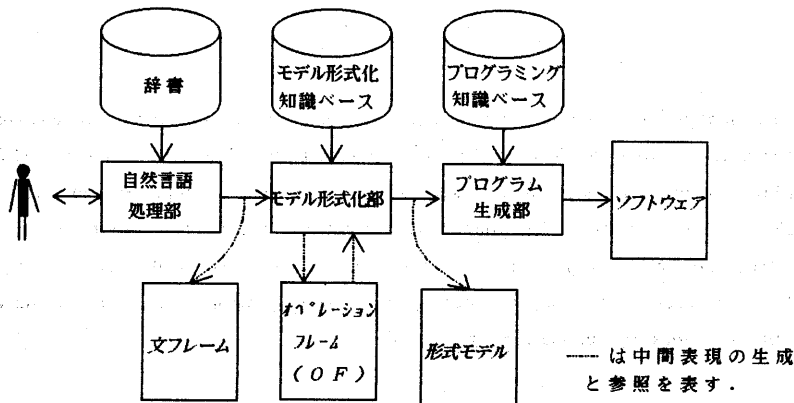


図1 システム構成

検証 : 作成された形式モデルをユーザに提示し、ユーザの意図したプログラムモデルと一致するかどうかを検証する。

2. 3 プログラム生成部

プログラミング知識ベースに含まれる目標言語の構文や意味に関する知識を用いて抽象的な形式モデルを詳細化し、具体的なソースコードに変換する。

2. 4 自然言語仕様記述に課した制限

本システムで受理する自然言語仕様に対して課した制約は次のとおりである。

言語: 日本語

構文: 単文, 条件文

3. オペレーションフレーム

自然言語仕様からソフトウェアを作成するには、自然言語仕様に含まれる意味をシステムが処理できるように表現することが重要である。そこでユーザが仕様記述で、「どのようなことを行いたいのか。」といった機能的な操作を中心に記述することに着目して、仕様記述の意味を対象領域におけるオペレーションを中心とした表現に変換する。この表現形式をOFと呼ぶ。

OFは図2で示すようにプリミティブタスク(以下PTと略記)、格役割、格要素で構成される格フレ

ムである³⁾。また、格要素となる単語が持つべき属性としてケースプロパティ(以下CPと略記)がある。

以下、各構成要素を説明し、文献4)の待ち行列シミュレーションの14個の例題に対する仕様記述に基づいた設定を示す。

なお、条件付き選択や、条件付き反復を表す条件文、また、その条件部で用いる数値比較の演算文も同じOFの枠組みで扱う。

3. 1 プリミティブタスク

同一の内容もしくは処理について自然言語で表現するとき多様な表現が可能となり、意味理解をきわめて困難にしている。この困難さを取り除くには自然言語での多様な表現をシステム内の一意の表現に変換しなければならない。そこで対象領域における処理を実現するために必要とされる基本的操作を設定し、対象領域の用語を用いて記述したものをPTと呼ぶ。

待ち行列シミュレーションのPTを表1に示す。

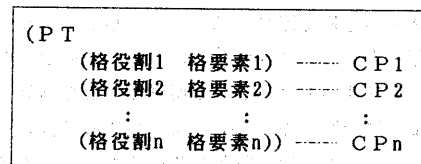


図2 OFの構造

表1 PT

PT	説明
施設のタスク	
start_at	開始する
end_at	終了する
初期リソースの定義	
init_exist	初期状態で存在する
客のタスク	
客の到着定義	
arrive	到着する
リソース無関係のタスク	
exit	行うタスクがなくなり退去する
waste_time	時間を消費する
リソース関係のタスク	
ask_resource	リソースの有無を調べる
get_resource	リソースを獲得する
rel_resource	獲得したリソースを解放する
make_resource	リソースを生成する
条件によるタスクの制御	
if	条件付き選択を行う
while	条件付き繰り返しを行う
=, <, <=, >, >=, =	数値比較の演算を行う

3. 2 格役割

OFは「どのような行為主体がどのような規範に基づいてPTで表されるオペレーションを行うのか。」を表現する。したがって、それぞれの単語がPTに対してどのような役割をもって結合しているのかを表すものが格役割である。

シミュレーションにおいて重要な役割をなす時間、リソース、数量等に重点をおいて設定した、それぞれのPTが取り得る格役割を表2に示す。

3. 3 ケースプロパティ

格要素となる単語とPTとの意味的な整合性を検査するために、格要素の満たすべき意味的な制約条件としてCPを設定する。

シミュレーションの仕様記述に現れる単語の意味的な制約条件を検討した結果に基づき設定したCPを表3に示す。

4. 欠落情報の補完

自然言語仕様記述では単語レベルや文レベルの情報の欠落が生じる。これはOFにおいては格要素の欠落や、OF自体の欠落となって現れる。この結果、ユーザの意図に合致したソフトウェアが作成できない。そこで本システムでは、OFの情報をもとにして省略要

表2 PTに対する格役割

格役割	説明
agent	行為者
time	時刻
duration	時間
place	場所
quantity	数量
resource_name	対象リソース
priority	優先権
condition	条件
then	成立時のタスク
else	不成立時のタスク

表3 CPの例

CP	説明
resource_user	リソース使用者
number	数
resource	リソース
time_a	時刻(一回)
time_i	時刻(反復)
of	OF

素(格要素、文)を検出し、対象領域に関する体系化された知識を適用してその要素を補完する。

本研究では、文献4)の例題に対する自然言語仕様記述を対象として、仕様中で起こり得る省略について考察し、その補完方法に応じて4種に分類した。なお、各分類の命名については文献5)を参考にした。

ここでは各省略の型について、省略の背景(ユーザの意図)、省略の例、補完に用いる知識、補完法を述べる。さらに、各型の適用戦略を述べる。

4. 1 単語の省略補完

4. 1. 1 連想型(I, II型)

ある文中の単語から、その文中の他の単語が容易に連想される場合には、ユーザはその連想される単語を省略することがある。

>補完法:

I型: 格要素としてOFに存在している単語から連想される種々の情報をその語の付属情報から取り出して、欠落情報を埋める。

例1: 「客は30分の指数分布で到着する。」

→「指数分布」に対し30分は平均である。

→「客は平均30分の指数分布で到着する。」と補完する。

II型: 欠落している格要素のCPをユーザに提示し、その属性を持つ単語を入力してもらうことで補完する。

4. 1. 2 文脈型(I, II, III型)

既に記述した情報であり、かつ文脈を参照することによりその情報が復元できるときには、ユーザは冗長な繰り返しを避けるために、その情報を省略することがある。

>補完法:

省略を含むOFを処理する以前のOFを参照し、欠落情報を補完する。

I型: agent省略の場合→前文のagentが候補となる

ただし、PTがifまたはwhileの場合のcondition部のagent省略の場合は、then部またはelse部のagentが候補となる。

II型: PT省略の場合→前文のPTが候補となる

例2: ①と②が連続して入力されるとする。

①「客1が来る。」

→(arrive (agent 客1) ...)

②「間隔は平均30分の指数分布である。」

→(PT? (time_i exp ...)...)

省略されたPT?を arriveと補完し、2つのOFを合成する。

Ⅲ型：agent以外の格要素の省略の場合→

欠落している格要素のCPを属性として持つような単語を文脈中に探索し候補として提示する。

これは連想型Ⅱ型において、CPにより指定された属性を持つ単語が文脈中に存在する場合である。

4. 1. 3 デフォルト型

対象世界で常識的な数値、あるいは特に注意を払わない数値に対してはユーザは省略することがある。

>補完法:

欠落情報をあらかじめシステムで有するデフォルト情報により補完する。

例3：リソースの有無の問い合わせ(ask_resource)の場合

PTがask_resourceであるOFにおいて

quantityの記述が省略されていれば

→ユーザは「とりあえずその資源があるかどうかを知りたい。」のだと判断する。

→quantity=1と補完する。

例4：リソース要求の優先度の場合

リソースの要求においてpriorityは省略される可能性があるが、その場合ユーザは「特に優先度を問題としていない。」と解釈してpriorityを最も優先度の低い0にする。

4. 2 文の省略補完

4. 2. 1 関係利用型

客が一連のタスクによりある機能を達成する場合、ユーザは中心的なタスクを記述し、それに付随するタスクを省略する場合がある。そのため仕様記述において、付随的なタスクを記述する文が抜けてしまうことになり、モデル形式化部では省略された文に相当するOFを補完することが必要となる。

>補完法:

PT間の順序関係を対象領域の知識として規定しておき、この順序関係を用いて省略されたOFを補完する。

例5：客のトップレベルの行動

[arrive→タスク群→exit]

一般にシミュレーションにおいて客は到着後に一連のタスクを行い、その後退去するが、退去

に関する記述を忘れることがある。そこで、客が退去するかどうかをユーザに問い合わせ、退去する場合にはexitをPTとするOFを補完する。

例6：タスク群の順序関係

リソースの獲得に関するPTの組合せには、

① [ask_resource→get_resource→waste_time→rel_resource]

② [get_resource→waste_time→rel_resource]

③ [get_resource]

のように3つのパターンがあるが①、②においてユーザはリソース獲得の中心となるget_resourceについては記述するが、付随的なタスクであるask_resourceについては省略する場合がある。そこでask_resourceをタスクとして持つのか否かを問い合わせ、①、②のどちらのパターンであるかを決定する。

4. 3 補完の戦略

上述の各補完方法の適用順序を効率の面から検討し、以下のような順序とする。

①連想型Ⅰ型を用いる。

②デフォルト型を利用して、数値的な欠落情報を補完する。

③文脈型Ⅱ型を利用して、2つのOFの合成を行うことで省略を補完する。

④文脈型Ⅰ型を利用して、省略された格要素のうちagentを補完する。

⑤文脈型Ⅲ型を利用して、省略された格要素のうちagent以外の省略を補完する。

⑥⑤のCPを持つ単語を文脈内に検索できなかった時は連想型Ⅱ型を利用してユーザに問い合わせて補完する。

⑦関係利用型で候補とされるOFを提示し、ユーザがOFを必要とした時には上述の①～⑥を繰り返してそのOFを埋める。

この方法により、欠落情報が補完され、モデル形式化部の次の操作であるモデルへの構造化に進む。

5. 形式モデル

形式モデルはユーザが記述した自然言語仕様に含まれる情報をソフトウェアのモデルとして構造化したものである。

5.1 オブジェクト指向のモデリング

オブジェクトとは、対象とその対象を操作する手続きを一まとめにした概念単位である。オブジェクト・モデルとは、計算すべき問題領域に存在する対象やその問題を説明するのに必要な概念を、それぞれオブジェクトおよび、それらのオブジェクトが解釈しうるメッセージとしてモデル化したものである。

このオブジェクト・モデルの特徴を以下に挙げる。

- ・対象領域をシミュレーションとした場合、さまざまな“もの”がシミュレーションの世界のオブジェクトとして素直に表されるのでモデル構築性に優れている⁶⁾。
- ・ある一つの個体に関する情報をまとめて記述するので可読性の点で優れている。
- ・目標言語がオブジェクト指向型言語の場合には、制御文、変数を付加することにより容易にソフトウェア化が可能であり、移行性が高い。

以上のことから、本研究において形式モデルにはオブジェクトモデルを採用する。

5.2 形式モデルの記述形式

形式モデルはフレームの集合で表され、個々のフレームは次のような記述形式をとる。

```
(object_name
  (condition_tag_1 task_1)
  :
  :
  (condition_tag_n task_n))
```

object_nameには客の名前または施設名を記述する。condition_tagにはif, then, else, endif, while, wend, null (空白)のうちいずれかを記述する。taskには客または施設の振舞いをOFの形式またはそのagentを省略した形式で記述する。

object_nameを見ることでシステムがどのような要素で構成されているかを理解でき、condition_tagにより、条件によるタスクの制御が読み取れる。

また、object_nameが施設名であるフレームは、施設の初期状態や施設に対する客の到着(施設が呼を生じたものと見なす)を表現する。

object_nameがその他であるフレームは客のタスクを表現する。

5.3 OFから形式モデルへの構造化

複数のOFの情報をオブジェクト毎(ここではagentになる名詞毎)にまとめる。その際の手続きを次に示す。

それぞれのOFについて、PTが、

- ①start_at, end_atであるとき、agentをobject_nameとするフレームのtaskに、そのOFのagentを除いて記述する。
- ②arriveであるとき、placeをobject_nameとするフレームのtaskに、そのOFのplaceを除いて記述する。
- ③ifまたはwhileであるとき、格役割thenのOF中のagentをobject_nameとするフレームに対して、condition_tagがifであるtaskにconditionのOFを、thenであるtaskにthenのOFを、elseであるtaskにelseのOFをそれぞれ記述する。
- ④その他の時、agentをobject_nameとするフレームのtaskに、そのOFのagentを除いて記述する。

上述の手続きのうち、agentをobject_nameとするフレームが形式モデルに存在しなければそのフレーム(ただし、空きフレーム)を作成してから、手続きを継続する。

例として、客1が平均12分の指数分布で店に到着し、その後のタスクの一部が「もし本が3冊あるならば、客は2冊買う、さもなければ、15~20分の一様分布で待つ。」の場合の形式モデルを図3に挙げる。

```
(店
:
(arrive (agent 客1)
(time_i (exponential (mean 12))))
:
)
(客1
:
(if
(ask_resource
(quantity 3)
(resource_name 本)))
(then
(get_resource
(quantity 2)
(resource_name 本)))
(else
(waste_time
(duration
(uniform (from 15) (to 20))))))
(endif)
:
)
)
```

図3 形式モデルの例

6. 具体例

本システムの動作を具体的な例で説明するため、ハンバーガショップに来店する3タイプの客と納入業者の振舞いをシミュレートするための自然言語仕様記述を形式モデルに変換するまでの過程を考える。ここでは、図4において、客3の振舞いに関する記述である①、②、③、④の文がどのように変換されていくかを中心に説明する。

6. 1 仕様から文フレームへの変換

まず格文法に基づく解析により⑤、⑥、⑦、⑧のように動詞をフレーム名とする文フレームに変換する。

6. 2 文フレームからOFへの変換

次に、モデル形式化知識ベースに含まれる、動詞をPTに変換する規則(例: 来る→arrive)、文フレームの格役割をOFの格役割に変換する規則(例: 主格→agent, 時刻→time)の適用、並びに時間単位の変換(ここでは分を基本単位と仮定して、時間・秒から分への変換)により⑨、⑩、⑪、⑫になる。

6. 3 OFの欠落情報補完

OFの格要素で??で表されるのが欠落している情報である。⑩において“??facility”はplaceの格要素としてfacilityを属性としてもつものが要求されることを表している。これに対し文脈型II型を用い、既に処理したOFの格要素を検索し、“店”がfacilityを属性として割り当てられているので“店”を補完する。⑩で時刻を表す表現のtime_dが欠けているが、連想型I型により一様分布を表す“uniform”を補完する。また⑩、⑪、⑫でagentの格要素が抜けているが、文脈型I型を用いて⑩で用いられている“客3”を順次補完する。⑪でリソースの個数を表す??numberが抜けているがデフォルト型により1を補完する。

6. 4 OF自体の補完

客3のタスクにおいて退出を表すタスクが見られない。そこで関係利用型によりユーザに問い合わせ、退出のタスクを表すOFを⑬のように補完する。

6. 5 OFの集合から形式モデルへの変換

客3のOFのうち、到着を表すPTつまりarriveをフレーム名とするOF⑨を⑭のように形式モデルの施設を表すフレームに記述し、その他のOFすなわち、⑩、⑪、⑫、⑬を客3がフレーム名であるフレームに

⑯、⑰、⑱のように記述する。

以上の操作により自然言語仕様は形式モデルに構造化される。

7. まとめ

本稿では、自動プログラミングシステムにおける仕様に含まれる情報の表現形式として、対象領域でのオペレーションを中心とした表現(OF)を提案した。つぎに、OFの集合に現れる単語レベル、文レベルの欠落情報を分類し、各分類に対する補完法を述べた。また、OFからソフトウェアモデルに構造化する際にオブジェクト中心に構造化することによりモデルの可読性・移行性が向上することを示唆した。

今後の課題として、今回提案した補完法の評価、さらに複雑な自然言語仕様の受理が挙げられる。

参考文献

- 1) 辻井, 上原: “ソフトウェア工学と自然言語処理”, 情報処理, Vol.28, No.7, pp.913-921(1987).
- 2) 白井, 山田, 手塚: “待ち行列シミュレーションモデル作成のための自然言語インタフェース”, 第36回情全大, 6N-3(1988).
- 3) 白井, 山田, 馬場口, 手塚: “自然言語仕様からモデル表現への変換—シミュレーションソフトウェアを対象とした場合—”, 第37回情全大, 2J-1(1988).
- 4) Goldberg, A. and Robinson, D.: “Smalltalk-80 The Language and Its Implementation”, Addison Wesley, pp.439-537 (1983).
- 5) 唐沢, 松永, 小川, 田村: “省略表現を含む不完全テキストを知識を用いて復元する英文補完システム”, 情学論, Vol.26, No.2, pp.374-381(1985).
- 6) 山本: “オブジェクト指向とシミュレーション”, 情報処理, Vol.29, No.7, pp.374-381(1988).

<p>自然言語仕様記述</p> <p>店は午前7時に開店する。 午後10時45分に閉店する。 客1は平均4分の指数分布で来店する。 10~25秒でメニューを見る。 ハンバーガを2個買う。 5~16分かけて食べる。 客2は平均15分の指数分布で店にくる。 もし、ハンバーガが3個あるならば、3個買う。 さもなければ、店を出る。</p>	<p>客3は午前12時から午後1時まで 平均2分の指数分布で店にくる。...① 10~25秒メニューを見る。...② ハンバーガを買う。...③ 5分~25分かけて食べる。...④ 納入業者は平均2時間の指数分布で店にくる。 ハンバーガを200~300個納入する。</p>
<p>文フレーム</p> <p>(開店する (主格 店) (時刻 午前7時))</p> <p>(閉店する (時刻 午後10時45分))</p> <p>(来店する (主格 客1) (時刻 (平均4分の指数分布))</p> <p>⋮</p> <p>(来る (主格 客3) (時刻 (午前12時から午後1時まで 平均2分の指数分布))</p> <p>...⑤</p>	<p>(見る (時間 10~25秒) (対象 メニュー)) ...⑥</p> <p>(買う (対象 ハンバーガ)) ...⑦</p> <p>(食べる (時間 5~25分)) ...⑧</p> <p>⋮</p> <p>(納入する (対象 ハンバーガ) (個数 200~300))</p>
<p>OF</p> <p>(start_at (agent 店) (time 7*hour))</p> <p>(end_at (agent ??facility) (time 22*hour+45))</p> <p>(arrive (agent 客1) (time (exponential (mean 4))) (place ??facility))</p> <p>⋮</p> <p>(arrive (agent 客3) (time (exponential (mean 2)(starting 12*hour) (ending 13*hour)) (place ??facility))</p> <p>...⑨</p> <p>(waste_time (agent ??resource_user, ??resource_producer) (time (??time_d (from 10*second) (to 25*second))))</p> <p>...⑩</p>	<p>(get_resource (agent ??resource_user) (quantity ??number) (resource_name ハンバーガ)) ...⑪</p> <p>(waste_time (agent ??resource_user, ??resource_producer) (duration (??time_d (from 5) (to 25)))) ...⑫</p> <p>⋮</p> <p>(make_resource (agent ??resource_producer) (quantity (discrete_uniform (from 200) (to 300)) (resource_name ハンバーガ))</p> <p>----- 関係利用型により補完したOF -----</p> <p>(exit (agent 客1))</p> <p>(exit (agent 客3)) ...⑬</p> <p>(exit (agent 納入業者))</p>
<p>形式モデル</p> <p>(店 (start_at (time 7*hour)) (end_at (time 22*hour+45)) (arrive (agent 客1) (time (exponential (mean 4)))) (arrive (agent 客2) (time (exponential (mean 15)))) (arrive (agent 客3) (time (exponential (mean 2)(starting 12*hour) (ending 13*hour)))</p> <p>...⑭</p> <p>(arrive (agent 納入業者) (time (exponential (mean 2*hour))))</p> <p>(客1 (waste_time (duration (uniform (from 10*second) (to 25*second)))) (get_resource (quantity 2) (resource_name ハンバーガ)) (waste_time (duration (uniform (from 5) (to 16)))) (exit))</p>	<p>(客2 (if (ask_resource (quantity 3) (resource_name ハンバーガ)) (then (get_resource (quantity 3) (resource_name ハンバーガ)))) (else (exit)) (endif))</p> <p>(客3 (waste_time (time (uniform (from 10*second) (to 25*second)))) ...⑮</p> <p>(get_resource (quantity 1) (resource_name ハンバーガ)) ...⑯</p> <p>(waste_time (duration (uniform (from 5) (to 25))))...⑰</p> <p>(exit) ...⑱</p> <p>(納入業者 (make_resource (quantity (discrete_uniform (from 200) (to 300)) (resource_name ハンバーガ)) (exit))</p>

図 4 具体例