

# 単文の意味を考慮した 常識的時間判断システムの構築

吉川義紀<sup>†1</sup> 吉村枝里子<sup>†2</sup> 土屋誠司<sup>†2</sup> 渡部広一<sup>†2</sup>

**概要:** 我々は日常生活を送る際に、時間軸を利用している。誰かとコミュニケーションをとる際に、時間やスケジュールを考えて話を行うことがある。本稿では、こういった人間が日常的に行っている時間に関する理解をシステム化することを目指した。このシステムを常識的時間判断システムと定義する。本システムでは、入力文中に複数の時間に関する語（時語）が出現した際の、各時語への重み付けの手法を提案した。また、入力文中の動詞の働きを考慮した時間判断機能をシステムに実装した。ここで動詞の働きとは、入力動詞により入力文全体が持つ時間や季節を変化させる働きのことと定義する。

## Construction of Common-sense Time Judgment System Based on the Meaning of Simple Sentence

YOSHIKI YOSHIKAWA<sup>†1</sup> ERIKO YOSHIMURA<sup>†2</sup>  
SEIJI TSUCHIYA<sup>†2</sup> HIROKAZU WATABE<sup>†2</sup>

**Abstract:** We are using the time axis as a reference of life. When we decide our schedule and we communicate with each other, we must consider time or schedule. Therefore, in this study, This study aim to the construction of common-sense time judgment system, which people do on a daily basis. In this system, for the sake of common sense judgment, This system have proposed a method of weighting when many time language is present in the input sentence. Moreover, it also was implemented function for time judgment based on the meaning of the verb.

### 1. はじめに

情報化社会の発展とともに、今日の情報技術は目覚ましい進歩を遂げた。デジタルデバイドの問題を考えると、高齢者を含むあらゆるユーザが特別な訓練をせずとも利用しやすいシステムの実現が望まれる。ここで、高度な情報処理システムの1つとしてコミュニケーションロボットを考えてみる。コミュニケーションロボットは人が日常で行っている会話をインターフェースとして利用している点において操作性が高いと言える。人とロボットのコミュニケーションを考える前に、人と人がコミュニケーションを行う場合を考えると、受け手は話し手の発話内容を受け取り、その相手が何を伝えたいかを理解することで会話が成立する。その理解のために、人間は「常識に基づいた連想」を無意識に行う。これは、人が日常的に行う会話中に存在する時刻や季節を用いた表現に対しても言える。例えば、「17時に集合して星を見に行こう」という発言からは、待ち合わせの日時が理解できる。また、人間は「クーラーといえは夏」のように、常識的な季節や時刻を連想することがで

きる。本稿では、季節・時刻の連想を「時間判断」と呼ぶ。人間が日常で何気なく行う時間判断を常識的時間判断と呼ぶ。時間判断を目的とした初期の研究として「時間の常識的判断メカニズムとその未知語処理(野村2003)」<sup>[1]</sup>がある。このシステムは入力された単語や入力文に季節や時間を表す語が存在する単文のみから時間帯や季節を連想して出力するものであった。一方で、実世界で用いられる時間判断は、単文中に存在する季節や時間を表す単語のみで時間判断を行うことはまれである。そのため、文章にほぼ対応できない初期の研究には限界があった。この問題解決のための研究として、「文の意味を考慮した常識的時間判断システムの構築(吉川2014)」<sup>[2]</sup>が行われた。本稿ではこれを先行研究とする。このシステムでは野村によって作成されたシステムをもとに、単語による時間判断だけでなく、より人間に近い単文からの時間判断実現を目指した。しかし、このシステムでは、入力された単文を1つの単語に変換することによって時間判断を行っているため、動詞の働きへの対応が不十分なことや、複数の時間や季節が入力文に出現した場合にどの季節や時間を優先するべきかシステムが判断できないなどの問題点から、正確に入力文全体の意味や内容を考慮して複雑な時間判断を行っているとは言えない。そこで本稿では入力文の意味を考慮して、入力文全体が持

<sup>†1</sup> 同志社大学大学院理工学研究科  
Graduate School of Science and Engineering, Doshisha University  
<sup>†2</sup> 同志社大学理工学部  
Faculty of Science and Engineering, Doshisha University

つ季節および時刻，時間帯が判断できる時間判断システムの構築を目的とする．ここで，入力文の意味を考慮した時間判断システムとは，入力文中の単語のみから時間判断を行うシステムではなく，入力文中に存在する動詞が文章に与える影響も考慮した時間判断を行うシステムである．

## 2. 関連技術

本稿で用いられている基礎技術について述べる．本稿では，概念ベース<sup>[3]</sup>，関連度計算方式<sup>[4]</sup>，関連度計算方式を応用した観点付き関連度計算方式<sup>[5]</sup>，逆引き国語辞書<sup>[6]</sup>と動詞項構造シソーラス<sup>[7]</sup>を用いた．

### 2.1 概念ベース

概念ベースとは電子国語辞書などを基にして機械的に構築された大規模な知識ベースである．概念ベース内で約9万の「概念」が定義されており，概念の意味的特徴を表す「属性」とその属性が概念にとってどれだけ重要かを表す数値である「重み」の対の集合より定義されている．

### 2.2 関連度計算方式

関連度計算方式とは概念ベースを利用し概念と概念の関連の強さを表現するものである．関連度の値は0.0~1.0の値を取り，その値が大きいほど関連が強い．

### 2.3 観点付き関連度計算方式

観点付き関連度計算方式は，関連度計算方式を応用したもので，観点を与えることで，その観点による概念同士の関連の強さを求める関連度計算方式である．

### 2.4 逆引き国語辞書

逆引き国語辞書は単語の意味を表現した文を入力とし，単語とその意味を定義した文(以下，定義文)を格納したデータ(以下，辞書データ)と比較することで，入力文の意味に適した単語を出力するシステムの実現のために構築された．ここで，辞書データには23万8千語の豊富なデータを保持する大辞林<sup>[8]</sup>を使用している．

### 2.5 動詞項構造シソーラス

複数の動詞を共通性質でまとめて，統語的に，あるいは意味的な振る舞いがどのタイプであるのか定義されている．それぞれの動詞に対して意味や統語的な共通性質でグループ分けを行うことで，操作的に動詞を選択することができ，また文中にある動詞表現が他のどの表現とどう異なるのかがわかり，文書理解への貢献のために，動詞項構造シソーラスは作成された．例えば「本を買った」と「本を借りた」は本が手元にあるという結果状態では違いは無い．しかし「買った」は所有権が移動しているが「借りた」は所有権は移動していない．このような関係を同時に記述するために，上位の共通属性として【他者からの所有物の移動】という概念が登録されている．さらに，ここから下位概念として【獲得】と【借用】が登録されている．これは，所有権に関する動詞の差異を表現するためである．そして【獲得】には「買う」を所属させ，【借用】には「借りる」を所属させる．またこの際に，動詞の同義語もインスタンスとし

て下位概念に所属させる．例えば「借りる」と同義にあたる「レンタルする」や「借り入れる」などがこれにあたる．

動詞項構造シソーラスは木構造シソーラスの形で構築されている．図1に動詞項構造シソーラスの構成を示す．

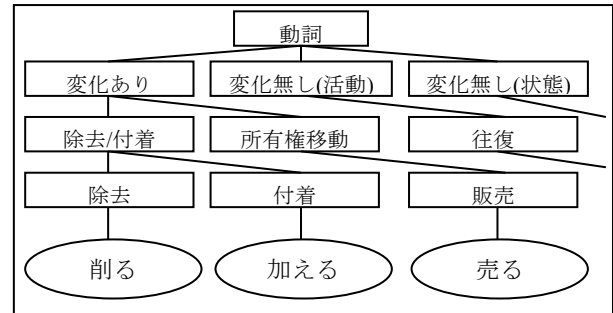


図1 動詞項構造シソーラスの構成

動詞項構造シソーラスには，4425個の動詞が格納されており，類義語の種類は709種類で，分類は5階層，意味役割の種類は71種類が定義されている．言語的利便性を考慮して構築されたシソーラス構造についてその制約などは以下の通りである．

- (a) シソーラスの各ノードは概念であり，単語ではない
- (b) 概念間には上位下位関係のみに限定し，多重継承は行わない
- (c) ある語義が複数の概念に属する場合はインスタンスのみ
- (d) 階層の数は特に制限しない

## 3. 先行研究

先行研究である時間判断システム(以下，旧時間判断システム)の定義や処理の流れ・使用技術について述べる．

### 3.1 旧時間判断システム

旧時間判断システムは，人とコンピュータとの会話を実現するために，人間が行う時間判断をシステム化したものである．例えば，人間は「梅雨が明けましたね」という発言に対して「もう夏ですね」という自然な返答を行うことができる．初期の時間判断システムでは，入力文中の「梅雨」という単語のみから時間判断を行っていた．しかし，こういった自然な会話を行うためには，「梅雨が明ける」から単純に季節「梅雨」を導くのではなく，「明ける」という動詞の働きにも着目し，時間や季節に関する判断を行うことが不可欠であるという考えから旧時間判断システムが構築された．ここで「梅雨」や「クリスマス」のように人が季節や時間帯を連想することができる単語を「時語」と定義する．

### 3.2 旧時間判断システムの流れ

旧時間判断システムの流れについて説明する．入力には「数字+時間単位(時，年など)」，「単文」，「単語」とし，複文・重文等は対象としない．ここで「単文」とは主述関係が1つの文章とする．まず入力に対して形態素解析を行い，単語か単文かを判定する．単語と単文の定義に基づいて，

入力が体言 1 つのみか用言 1 つのみの場合は単語と判定する。また、単語で無い場合は単文と判定し、入力文に動詞が複数存在する場合や主述関係が 2 つ以上見つかった場合は「主述関係は 1 つにしてください」と出力する(単文判定)。その後、入力が文章の場合はまず形態素解析で得られた単語全てに対し、「時変換語対応(3.4 節で後述)」において、「時変換語」が存在するかどうか判定する。その後、入力文に〇〇日後といった範囲時語が存在するかどうか判定する「範囲時語判定(3.5 節で後述)」を行う。そして、「時語生成(3.6 節で後述)」を行う。「時語生成」は単文から単語を作成する処理である。ここまでの処理で単語入力の場合は入力単語が、文章入力の場合は「時語生成」により作成された単語が獲得される。最後に獲得された単語に対して「時語理解(3.7 節で後述)」を行う。この処理は単語と時語知識ベース(3.3 節で後述)とを参照することで「時」の取得を行う処理である。ここで、「時」とは入力から連想する適切な日付、時刻、季節および時間帯を指す。また、入力に「範囲時語」や「時変換語」が存在する場合は取得された「時」を適切なものに変化させる処理も行う。また、時語知識ベースに存在しない語についても概念ベースによる未知語処理を行うことで対応することができる。旧時間判断システムの流れを図 2 に示す。

システムの流れを理解しやすくするために具体例を用いて説明を補足する。例えば「梅雨が明ける」という例について考えてみたい。「梅雨が明ける」を単文判定にかけると、主語「梅雨」と、述部「明ける」がそれぞれ 1 つずつ存在するため単文と判定される。次に、「時変換語対応」判定処理において「梅雨が明ける」の「明ける」が季節を変える言葉として登録されている。次に、「範囲時語判定」の処理では、今回の入力には範囲時語は含まれないため、影響を受けない。次に「時語生成」により、入力文より時語「梅雨」を取得する。最後に、「時語理解」により取得された時語「梅雨」に対して動詞「明ける」の意味を考慮して、季節を「梅雨」から「夏」へと進める。

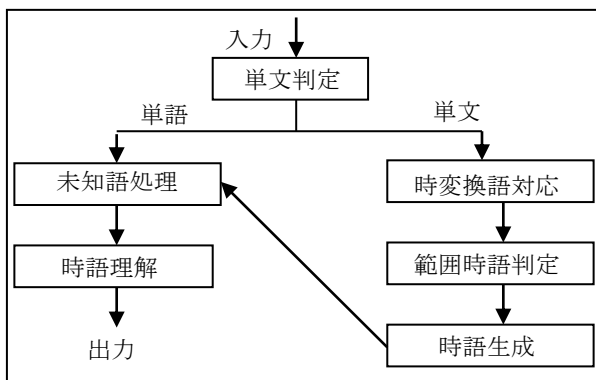


図 2 旧時間判断システムの概略図

### 3.3 時語知識ベース

旧時間判断システムにおける判断の基本である時語知識ベースには、大学生約 150 人に「時を表現する語」または

「時を連想する語」を 20 語以上、自由記述で回答してもらった。その回答のうち、5 名以上が回答した語を時語として登録している。知識数を増やせば精度が上がるのが考えられるが、人が時間や季節を連想できる語は固有名詞など無数に存在するため、知識を詰め込むには限界がある。そこで、この知識数でどれだけ多くの時間判断に対応できるかが重要なポイントとなる。時語知識ベースには、565 個の時語が登録されている。時語は、人が季節や時間帯を連想できる単語のことであり、明示の時語と暗示の時語に区別される。ここで、明示の時語は、「12 月 25 日」を指す「クリスマス」のように、明らかに時間や季節を示す言葉である。また、暗示の時語は、人間が「スキー」から「冬」という季節を連想するように、時を表す単語ではないが、特定の時に密接な関係を持つ単語である。表 1 に時語知識ベースの略例を示す。

表 1 時語知識ベースの略例

ID	名称	日時	代表語
276	クリスマス	12 月 25 日	
28	朝日		朝

### 3.4 時変換語対応

ここで時変換語の収集方法と処理について述べる。

#### 3.4.1 時変換語取得

例えば、「ストーブをしまう」、「ストーブを出す」のように、その文意によって時語が持つ「時」が変化してしまう。この場合、入力文の用言「しまう」と「出す」は「ストーブ」から得られる「冬」という季節に与える影響が異なる。人間は前者の発言を受けて「寒くなくなった」という印象を受け、「冬の終わり」や「春の訪れ」を感じる。そして、後者の発言からは「寒くなってきた」という印象を受けて「本格的な冬の始まり」を感じる。このような「しまう」や「出す」といった時語の季節感を変える働きをする語を「時変換語」とする。この「時変換語」を取得するためにアンケートを行った。まず、19 人の被験者に対して、時語知識ベースより選択した 30 個の時語とその時語が持つ「時」を与える。文意によって時語が持つ「時」が変化する単文を 1 つの時語に対して 1 つ作成するように指示をした。アンケートでは、例えば「時語:雪、季節:冬」を与えた場合、「単文:雪が溶ける、季節:冬→春」といった解答を取得した。採用語数は 52 語としてデータベースを作成した。時変換語データベースを表 2 に示す。

表 2 時変換語データベース

時変換語	時変換語フラグ
溶ける	1
望む	2

#### 3.4.2 時変換語処理

表 2 に示した「時変換語」データベースには、「時変換語」

と「時変換語フラグ」が格納してあり、語とその語が季節に与える影響をフラグとして登録している。入力文の動詞と時変換語データベースの「時変換語」が一致する場合、変換語が与える影響のフラグを「時語変換語フラグ」により取得する。フラグには季節を進める処理を示す「1」と季節を戻す処理を示す「2」が格納されている。フラグを用いた実際の処理は3.7節で述べる時語理解処理で行われる。

### 3.5 範囲時語判定

範囲時語判定について述べる。「今日から明日まで」といった範囲を示す時間(範囲時語)への対応を目指し、この処理を実装した。範囲時語については日本語時間表現の解釈法(表3)<sup>9)</sup>に基づいて定義する。表3より範囲を示す時語として「から、まで、後」などを範囲時語として定義し、その役割とともに範囲時語7語をデータベースに登録した。範囲時語データベースの一部を表4に示す。

表3 日本語時間表現の解釈法の分類とその例

分類	例
基準	今日から2日後
過去/未来	2日前/2日後

表4 範囲時語データベースの例

範囲時語	役割	範囲時語フラグ
から	開始地点	1
後	計算未来	3

範囲時語データベースには、フィールド「範囲時語」、「役割」、「範囲時語フラグ」が存在し、それぞれに範囲時語、範囲時語の役割、フラグが格納されている。範囲時語の「役割」とはフラグに対する説明書きであり、処理には使用していない。「範囲時語フラグ」とは、役割に対する実際の処理を制御するためのフラグであり、開始時間を決定するフラグを「1」、終了時間を決定するフラグを「2」、時間を進める計算を行うフラグを「3」、時間を戻す計算を行うフラグを「4」、〇日間や〇週間など開始時刻と終了時刻を計算するフラグを「5」、周期的な時間帯を表す「おき」を定義したフラグを「6」、「昨日」「今日」など時語と時語を結合する役割を示すフラグを「7」として、全7つのフラグを登録している。入力文と範囲時語データベースを参照し、一致する範囲時語が入力文中に存在すれば、その範囲時語に対応する範囲時語フラグを取得する。範囲時語判定においては、範囲時語の存在の有無および処理のフラグ取得しか行わず、実際のフラグを用いた処理は3.7節で述べる時語理解処理で行う。

### 3.6 時語生成

入力されたものが単語である場合、時語知識ベースで定義された時語と単純に比較することができる。一方で、入力が単文である場合、時語が単語の形で定義されている時語知識ベースと入力文を単純に照らし合わせることは困難

である。そこで、入力された単文を単語に変換する時語生成を実装し、2段階の処理で単文を単語へ変換している。1段階目は「体言のみによる手法」を用いており、ここでは、入力された単文に含まれる語句のうち体言と時語知識ベースを参照し「時」の取得を行う。入力文「落ち葉が舞う」に対して入力文中の体言である「落ち葉」と時語知識ベースを参照して「落ち葉」が「時語」であるかを検索する。

2段階目は2.4節で述べた「逆引き国語辞書」を採用する処理を行う。この処理を用いることで、入力された単文と辞書の定義文との意味の近さを考慮して、単文から単語を生成することができる。しかし、文中の体言や用言が「時語」である場合、文を単語に直す必要は無い。よって、旧時間判断システムでは用言・体言が「時語」かどうかを確認し、「時語」でない場合に逆引き国語辞書を用いる。入力文「葉が落ちる」を逆引き国語辞書にかけると辞書の定義文と入力の意味さを考慮して「落葉」という単語を取得する。この取得された単語と時語知識ベースを参照して「落葉」が「時語」であるかを検索する。

### 3.7 時語理解

時語理解では大きく分けて2つの処理を行っている。1つ目の処理は、入力された単語や時語生成により入力文より生成された単語と時語知識ベースを参照して「時語」に割り振られた「時」の取得処理である。例えば「雪」という時語と時語知識ベースを比較して、「冬」という「時」を取得することである。

2つ目の処理は、3.4節、3.5節で示した各種フラグに応じて適切な処理を行うことである。まず表2の2値を持つ「時変換語フラグ」に応じた処理について説明する。フラグ値が1の場合、はその文章から得る「時」を1つ進める処理を行い、フラグ値が2の場合、文章から得る「時」を1つ前に巻き戻す。例えば、「雪が溶ける」という入力に対して「溶ける」よりフラグ値1が取得される。さらに、時語理解処理の1つ目の処理により、雪から季節「冬」を取得する。そしてフラグ値1に応じて季節を「冬」から「春」へ進める処理を行う。次に、範囲時語処理における7つのフラグ値に応じた処理について説明する。ここでは、入力文と表4に示した範囲時語データベースを比較して「範囲時語フラグ」のフラグ値のうちいずれかが取得された場合に、「役割」に定義された処理を行う。例えば「今日から2日後」と入力された場合「から」と「後」という2語よりそれぞれフラグ値1と3が取得される。フラグ値1は「から」より前に存在する語を時間の開始点に設定し、範囲時語フラグ値3は「後」より前に存在する年月日や時間等を加算する処理を行う。これにより今日の日付を時間の開始点として、日付を2日間進める処理を行う。

## 4. 先行研究の問題点

先行研究には2つの問題点が存在する。1つ目は、「時変

換語」に関しての問題である。「時変換語」は3.4節で述べた「時変換語」取得アンケートにより収集された52語の動詞のみを定義した。しかし、先行研究では「時変換語」として定義した動詞52語では対応できない動詞が数多くある。1990年早稲田大学の森田らの調査<sup>[10]</sup>によれば、日本語における動詞には「単純動詞」として約2000語、「複合動詞」として約2400語の動詞が存在している。この調査から先行研究で定義した動詞50語では対応できない動詞が大量に存在することは明らかである。例えば「梅の木が枯れる」という例を考えてみたい。この例の正しい出力は「冬」だろう。しかし、先行研究で定義された「時変換語」には「枯れる」は定義されておらず、出力は「春」となる。これは、「時変換語」として定義した50語以外の動詞が入力文に含まれたために発生した問題である。

2つ目は入力文中に複数の時語が出現した場合に文意を考慮できないという問題である。例えば「19時に星を見に行く」という例を考えてみたい。この入力には「19時」という「具体的な時間」と「星」という「暗示的時語」の2つの時語が存在する。この「具体的な時間」や「暗示的時語」をここでは、「時語性質」と呼ぶこととする。既存システムでは、「時語性質」が異なる複数の時語が入力された場合に入力文中の最も後ろに存在する時語から「時」の取得を行うため、複数の時語を考慮できないという問題点があった。つまり、先ほどの例では最も後ろに存在する「星」という時語から「夜、18時～3時」と出力してしまうが、実際には、「暗示的時語」である「夜」よりも、「具体的な時間」である「19時」を優先して認識するほうが人間同士のコミュニケーションでは自然だと考えられる。これは、人間は「具体的な時間」をスケジュールや予定を設定する際に最も重要視するためである。次に「梅の木に雪が少し残る」という例を考えてみたい。この入力には「梅」という「季節が連想される時語」と「雪」という「季節が連想される時語」という2つの「時語性質」が同じ時語が出現する例である。この場合でも、旧時間判断システムでは入力文中の最も後ろに存在する時語から「時」の取得を行うので、出力は「雪」という時語から「冬」となる。また、「雪と梅の枯れ木」という例も考えてみたい。先ほどの例と同様に、時語は「梅」と「雪」という2つである。この場合でも、入力文中の最も後ろに存在する時語から「時」の取得を行うので、出力は「梅」という言葉から「春」となる。しかし、本来であれば、人は前者の季節を「春」、後者の季節を「冬」と連想すると考えられる。以上から、先行研究の出力は、「時語性質」が同様に異なるかに関わらず、人が行う時語理解の実現を目指したシステムの出力としては誤っている。

## 5. 提案する新時間判断システム

4章で述べた2つの問題点を解決するために、本稿で作

成した新時間判断システムについて、入力の定義やシステムの流れ・処理について述べる。なお本稿では、新時間判断システムに対して、旧時間判断システムで対応できなかった複数時語および動詞を考慮した入力文全体からの時間判断を実装することを目標として構築した。

### 5.1 入力の定義とシステムの流れ

新時間判断システムの入力は、1個以上の単語もしくは単文とする。ここで、「単語」とは「用言1つのみもしくは体言1つのみが存在するもの」と定義し、「単文」は「主述関係が1つしか存在しないもの」と定義する。本システムでは、「単語」の定義を「用言1つのみもしくは体言1つのみが存在するもの」としたため、入力に複数の時語が存在する場合も単文とみなして処理を行う。

新時間判断システムの流れを説明する。新時間判断システムは旧時間判断システムと共通の技術はそのまま使用し、意味理解処理(5.2節で後述)と複数時語処理(5.3節で後述)を追加した。また新時間判断システムの「時語理解(5.4節)」では追加した処理へ対応するため既存の「時語理解」を拡張した。図3に新時間判断システムの概略図を示す。

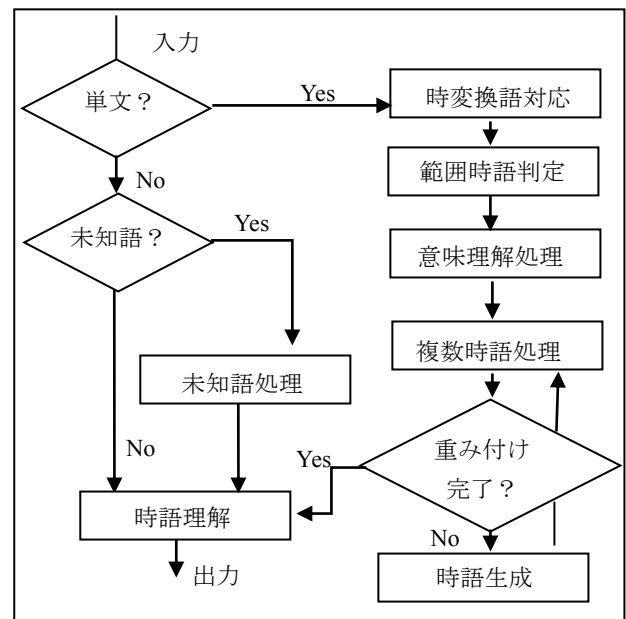


図3 新時間判断システムの概略図

まず入力に対して形態素解析を行い、単語か文章か判定する。単語と単文の定義に基づいて、入力に含まれる用言もしくは体言が1個の場合は入力を単語と判定し、入力が単語で無い場合や、入力に複数の単語が存在する場合は単文と判定する。ただし、先行研究と同様に、入力文に動詞が複数存在する場合や主述関係が2つ以上見つかった場合は「主述関係は1つにしてください」と出力し、回答を終了する。次に入力が単語であった場合は、その単語を用いて、時語知識ベースを参照する。また、入力が単文の場合はまず「時変換語対応」処理を行い、入力文の中に、「季節を進める」や「季節を巻き戻す」といった役割を持つ動詞が存在するかどうかを判定する。例えば「梅雨が明ける」

という入力の場合、「明ける」を「季節を進める」動詞であると判定する。その後、入力文に「範囲時語」が存在しないかを判定する「範囲時語判定」を行う。ここでは、「今日から〇〇日後」中における「から」のように時間的な幅を持つ語(範囲時語)が存在するかどうかを判定する。次に、入力文中の動詞がどういった働きを持っているかを判定する「意味理解処理」を行い、その後入力文中に存在する1個以上の体言に対して優先付けや重み付けを行う「複数時語処理」を行う。ここで、複数時語の抜き出しに、「時語生成」を用いる。以上の処理より得られた単語が時語知識ベースに登録されていない語である場合は未知語処理を行う。未知語処理は、入力文中に存在する未知の単語全てに実行される。最後に、「時語理解」処理を行い、「時変換対応」、「範囲時語判定」、「意味理解処理」で得られた入力文中に存在する語が持つ「役割」と入力文から得られた時語が持つ「時」を考慮して、最終的な「時」を決定し、それを出力する。

## 5.2 意味理解処理

意味理解処理は、4章で述べた問題点の1つ目を解決するために追加した機能である。既存システムが持つ1つ目の問題点とは、「時変換語」として登録された動詞が50語に限定されているために、それ以外の動詞に対応できないという問題点である。この問題点を解決するためには、未知動詞を「時変換語」に帰着させる未知語処理の手法も考えられるが、新時間判断システムでは2.4節で述べた動詞項構造ソーラスを利用した。未知動詞に対して未知語処理を採用しなかった理由は2つある。1つは未知語処理に用いられる重み比率付き関連度計算方式を利用した名詞間の意味の近さを判定した精度に比べて、動詞間の意味の近さを判定した場合の精度は低いことが2008年田中らの研究<sup>[11]</sup>によって明らかになっている。2つ目は4節で述べたように日本語動詞は無数に存在するものではない点である。ここでシステムに採用した動詞項構造ソーラスとは、動詞4425個に対して、読み、例文、動詞の役割などを体系的に割り当てた木構造の動詞データ群である。このうち役割には動詞「経つ」に対して役割「時の経過」といったものが割り振られている。動詞の役割について着目し、その「役割」に対応する形で「時を進める処理」や「時を巻き戻す処理」、「時を反転させる処理」を行うことで、「時変換語」に定義されない動詞に対応した。例えば、入力「梅が枯れる」において、動詞「枯れる」は動詞項構造ソーラスにおいてその役割が「死滅」と定義されており、この役割「死滅」に対応して「季節を戻す」処理を行う。これにより、入力文の「時」は「春」から「冬」に変化する。

### 5.2.1 動詞項構造ソーラスの利用

本稿では動詞項構造ソーラスのうち、動詞の役割を定義した「意味役割」に着目する。「意味役割」は動詞項構造ソーラスの3,4層目に存在する。この2層の役割から「時」

に影響を与えると考えられる役割を取り出し、70語でデータベースを構築した。データベースの例を表5に示す。

表5 役割知識ベース

役割	役割フラグ
存在	1
消滅	2

### 5.2.2 動詞の役割取得

動詞項構造ソーラスを用いた動詞の役割取得に関する流れについて説明する。まず、入力文中に存在する動詞と動詞項構造ソーラスに定義された動詞を比較し、動詞が一致した場合にその動詞に対する「意味役割」を獲得する。次に、それぞれの意味役割と役割知識ベースのフィールド「役割」に定義された語を比較し、役割が一致した場合に「役割フラグ」を取得する。この処理では、フラグの獲得のみを行い、フラグに応じた適切な「時」の出力は時語理解処理で行う。なお、フラグの値は3種類であり、それぞれ、フラグ値1は「季節を開始するフラグ」、フラグ値2は「季節を戻すフラグ」、フラグ値3は「季節を進める(終わらせる)フラグ」、である。

### 5.3 複数時語処理

複数時語処理とは、4章で述べた2つ目の問題点を解決するために追加された機能である。この問題点とは、入力中に時語が複数出現した場合、それらのうちのどの時語が、入力「時」を最も正確に表現しているかをシステムが判断できないという問題点である。例えば、「19時に星を見に行く」という入力には、「19時(具体的時間)」と「星→夜(暗示的時語)」の2つの時語が存在する。これらの「時語性質」は「具体的時間」と「暗示的時語」であり、入力は異なる「時語性質」を持つ時語を含む。一方、「梅の木に雪残る」という入力には、「梅→春(季節)」と「雪→冬(季節)」という2つの時語が存在するが、これらの「時語性質」は同一の「季節」である。「時語性質」が異なるか同様に、複数時語処理ではそれぞれ別の処理を行う。

#### 5.3.1 時語性質が異なる場合の複数時語処理

処理の説明の前に時語性質が異なる場合の処理実装のために行ったアンケートについて述べる。異なる「時語性質」の複数時語を含む例文5文から連想できる「時」を答えるように指示したアンケートを被験者10人に対して実施した。この例文5セットは、それぞれ例文中に「具体的時間」のみ、「具体的時間と暗示的時語」の2つの時語、「具体的時間と明示的時語」の2つ時語、「暗示的時語と明示的時語」の2つの時語、「具体的時間、暗示的時語、明示的時語」の3つの時語が存在していた。このアンケートの目的は、「時語性質」が異なる場合にどの「時語性質」を人は優先して認識しているのかを調査するためである。例えばこのアンケートでは被験者に「19時に星を見に行く」という例文を示して、この入力から連想される「時」は「19時」、「夜(18時~3時)」、「夜の19時」のうちどれが正しいと考えられ

るかといった結果を収集した。調査アンケートでは 10 人全員が「夜の 19 時」と出力するのが正しいと答えた。このようなアンケート結果から、「暗示的時語」から得られる「時」と「明示的時語」から得られる「時」と「具体的時間」から得られる「時」の優先順位を決定した。本稿では「具体的時間」 $\geq$ 「暗示的時語」=「明示的時語」の順で、これらから得られる「時」を優先することとした。

時語性質が異なる場合の複数時語処理では、入力内の複数の「時語性質」を持つ時語に対して優先順位を決定する。具体的な処理について説明を行う。入力文中に「具体的時間」と「暗示的時語」が含まれている場合は、具体的時間から得られる「時」を優先して出力する。また、時語の優先順位が同じ場合は優先付けを行わず出力する。例えば「クリスマスの夜」という例文を考えてみる。この入力には「夜(暗示的時語)」と「クリスマス(明示的時語)」が含まれる。この場合、優先順位は同じなので、「12月25日の夜」という出力をする。

### 5.3.2 時語性質が同様である場合の複数時語処理

処理を説明する前に、時語性質が同様である場合の複数時語処理を実装するために行ったアンケートについて述べる。同様の「時語性質」を持つ複数時語が存在する例文 5 文から連想できる「時」を答えるように指示したアンケートを被験者 10 人に対して実施した。このアンケートの目的は、「時語性質」が同様である場合に人はどのようにして「時」を認識しているのかを調査するためである。例えばこのアンケートでは被験者に、「梅の木に雪残る」と「雪残る梅の枯れ木」のような例を示した。先述のアンケートでは、前者の表す季節は「春」とであると 10 人中 7 人が、後者の表す季節は「冬」とであると 10 人中 9 人が回答した。しかし、既存システムでは前者を「冬」、後者を「春」と出力してしまう。この 2 つの例文において、片方の例文にのみ存在する単語は「木」と「枯れ木」である。ゆえに、「時」を表す時語ではない自立語(以下、「非時語」)である「木」や「枯れ木」が時間判断に影響を与えていると考えられる。つまり、この例文において「枯れ木」は「木」より冬を連想しやすいといえる。時語性質が同様の場合の複数時語処理は、入力内に同様の「時語性質」を持つ時語が存在する場合に、入力文の「時」を正しく表現していると考えられる時語を決定する処理である。旧時間判断システムでは入力中の複数の時語を同等に扱ってきたが、本来は複数の時語が文中に存在する場合、その入力を持つ「時」を最もよく表現しうる「時語」が存在するはずであり、同等に扱うのは問題があると考えた。そこで「時語」に対して「非時語」も考慮した重み付けを行うことで、「時語性質」が同一の場合でも優先付けできると考え、重み付けを用いることとした。この重み付けには、非時語と時語の意味的な近さを表す値である関連度を用いることにした。しかし、単純な意味の近さのみでは、時語が持つ「時」としての視点か

ら意味の近さを表現しているとはいえない。そこで、関連度計算方式を応用し、単純な意味の近さだけでなく、観点という形で与えた言葉を考慮した関連度を求めることができる 2.2.3 節で述べた観点付き関連度計算方式を採用した。

入力文中に複数存在する時語と非時語をそれぞれ概念として観点付き関連度計算方式を用いてそれぞれの意味の近さを判定する。ただし、本稿では観点として「季節」という語を与えることとする。この観点を与えたうえで、観点付き関連度計算方式を行い、そこで得られた関連度の値をそれぞれの時語に対する重みとする。ここで「梅の木に雪残る」と「雪残る梅の枯れ木」という 2 つの例を用いてこの処理の説明を行う。それぞれの例文には共通の時語である「雪」と「梅」が存在し、共通の非時語として「残る」が存在している。また、それぞれの例文内に異なる非時語として「木」と「枯れ木」が存在している。入力が「梅の木に雪残る」の場合は、観点「季節」を与え、時語「梅」と非時語「木」や「残る」とで観点付き関連度計算をする。一方で、入力が「雪残る梅の枯れ木」の場合は、観点「季節」を与え、時語「梅」と非時語「枯れ木」や「残る」とで観点付き関連度計算をする。それぞれの非時語との関連度を合計し、梅の重みとする。時語「雪」についても同様に計算し、重みを与える。この処理ではそれぞれの時語に重みをつける処理のみを行い、その重みに応じた「時語」から「時」を取得し、出力する処理は、時語理解処理で行う。

### 5.4 時語理解処理

時語理解処理では従来の時語知識ベース検索機能と「範囲時語」や「時変換語」のフラグ処理は既存システムと同様に行う。しかし、新たに追加された「意味理解処理」と「複数時語処理」に対応するためには、新システムにおける時語理解処理を、拡張する必要がある。まずは「意味理解処理」に対する「時語理解処理」について説明する。ここで、入力「梅の木が枯れる」という例を考える。「意味理解処理」より入力文中の動詞「枯れる」と小分類知識ベース参照することにより「死滅」に対する役割フラグ値 2 が獲得されている。このフラグ値 2 の役割は「季節を戻す」ことである。そのため入力文中の時語「梅」から取得された「春」から 1 つ季節を戻すことで「冬」と出力を行う。

次に、「複数時語処理」に対する「時語理解処理」について説明する。入力文中に時語が複数個存在する場合には「複数時語処理」により各時語について重みが割り振られている。また、「意味理解処理」により入力文中の動詞が「季節を終わらせる」役割であり、役割フラグ値 1 を取得している場合に、その動詞と主述関係にある「時語」の重みを 0 にする処理を行う。これは、「梅の枯れ木に雪が積もる」と「梅の枯れ木の雪が溶ける」における「時」の差を表現するために行う。後者の動詞「溶ける」の役割は「消滅」である。役割「消滅」から「季節を終わらせる」役割フラグ

が取得されるため、時語「雪」から「冬」を出力するのは間違いだと考えた。そこでこの重みを0にする手法を実装した。その後、この各時語に割り振られた重みに対してすべての組み合わせの差を取る。この差が閾値を超えるかどうかを判定する。超えた時語が存在する場合は、その時語と時語知識ベースを参照して「時」を取得する。閾値を超えた時語が存在しない場合は時語間に優位な差はないと考えて、入力中に存在するすべての時語と時語知識ベースを比較して、時語が持つそれぞれの「時」を出力する。

## 6. 評価と結果

旧時間判断システムと新時間判断システムの精度比較を行った。その評価方法と結果について述べる。

### 6.1 評価セットと評価方法

大学生34人に対して「時間、季節、時間帯」などを連想する単語と「時間、季節、時間帯」を連想できる単文を自由入力を書いてもらうアンケートを行い、評価セットを作成した。また、ニュース記事から収集した日常で用いる時間に関する表現を評価セットに追加した。これにより、単語50語と文章50文の評価セットを用意した。また、評価方法については次の通りである。まず、評価セットを新旧プログラムに通した出力結果を取得した。そして5人の被験者に、入力とその出力結果を示し、正しいと感じた場合は○、正しいとも間違っているともいえない場合は△、間違っていると感じた場合は×とする三段階評価を行った。

### 6.2 結果

被験者5人による評価結果を示す。評価○、△、×をそれぞれ2点、1点、0点として、評価は被験者5名の合計点が「8点以上の場合○、5点以上8点未満は△、4点以下は×」という評価を行った。本システムの精度は○と△を合わせた値とする。

表6 評価結果

システム ver	○	△	×
旧時間判断システム	38%	16%	46%
新時間判断システム	61%	10%	29%

表6より被験者が正しくないとして評価した×は17%低下する結果となった。

## 7. 考察

この章では評価結果からわかることを精度向上の要因とその課題について新旧時間判断システムの実例を示しながら述べる。

### 7.1 精度向上の要因

まず、例として「梅の木と雪」と「梅の枯れ木と雪」という2文についても考えてみる。この例文には時語「梅」と「雪」が存在し、非時語「木」と「枯れ木」が存在する。旧時間判断システムはどちらの例も「雪」から「冬」と出力していた。一方で新時間判断システムは時語「梅」や「雪」

だけではなく、非時語「木」や「枯れ木」がそれぞれの時語とどれだけ意味が近いかを考慮することで、前者の季節は「春」や後者の季節は「冬」と正しい出力を行うことができた。

### 7.2 課題

本システムには2つの課題がある。1つ目は固有名詞への対応である。テストセットに「バブル」、「紅白歌合戦」等の具体的な固有名詞が存在した。しかし、全ての固有名詞を時語知識ベースに登録することは不可能である。固有名詞に対応するには、時語になり得る固有名詞の収集やWebの利用が考えられる。2つ目はシステムが「時間判断を行うべき入力」であるかを判断できない点である。例「星を見る」と「プラネタリウムで星を見る」という2つの文について考える。これらは「星」という時語から「夜(18時~3時)」という結果が得られる。しかし、常識的にプラネタリウムが夜中の3時まで営業しているとは考えにくい。また、プラネタリウムで星を見ることは「夜」に限られたことではない。本当の意味で常識的時間判断を行う場合は、前者の例では「夜」と出力し、後者の例では「時」を出力することは不適切とシステムが判断する機能が必要である。

## 8. まとめ

常識判断システムの1つである時間判断システムに対して本稿では入力文中の複数「時語」や「動詞の役割」に着目し、入力文全体を考慮して時間判断を行うシステムの構築を行った。本稿においては、5章で述べた提案手法を実装することで、○と△を合わせた精度は比較システムと比べ17%向上し、71%となった。

**謝辞** 本研究の一部は、科学研究費補助金(若手研究(B)24700215)の補助を受けて行った。

### 参考文献

- [1] 野村理樹,渡部広一,河岡司,“時間の常識的判断メカニズムとその未知語処理”,FIT2003,pp.191-193,2003.
- [2] 吉川義紀,芋野美紗子,土屋誠司,渡部広一,“文の意味を考慮した常識的時間判断システムの構築”,FIT2014, pp.235-236, 2014.
- [3] 奥村紀之,土屋誠司,渡部広一,河岡司,“概念間の関連度計算のための大規模概念ベースの構築”,自然言語処理学会,Vol.14, No.5, pp.41-64,2007.
- [4] 井筒大志, 渡部広一, 河岡司,“概念ベースを用いた連想機能実現のための関連度計算方式”,FIT2002,pp.159-160,2002.
- [5] 松本和也,芋野美紗子,土屋誠司,渡部広一,“同義語・類義語を用いて観点を拡張した観点付き関連度計算方式”,FIT 2014, pp.207-208, 2014.
- [6] 三瀬慶久,吉村枝里子,土屋誠司,渡部広一,“文構造に着目した意味からの単語検索手法”,電子情報通信学会技術研究報告, pp.41-46,2010.
- [7] 松村明:大辞林第二版,株式会社三省堂,1995.
- [8] 溝渕昭二,住友徹,泓田正雄,青江順一,“日本語時間表現一解釈法”,情報処理学会論文誌,Vol.40, No.9, pp.3408-3419,1997.
- [9] 森田良行,“日本語と日本語教育”,凡人社,1990.
- [10] 田中翔,渡部広一,河岡司,“概念ベースにおける動詞の精練”,ICS2008,pp.127-132,2008.