変化を基本単位とした時系列情報の抽出と可視化

加藤 恒昭[1]　松下 光範[2]
東京大学[1]　日本電信電話株式会社[2]

概 要

統計量等の時系列情報について述べた文書群を要約可視化して動向情報を得ることを検討している．本稿では，そのために抽出すべき情報として，従来の情報抽出が基本としていた数値情報，つまり時系列情報の様々な時点での値ではなく，ある時間幅での値の変化が重要であることを述べる．変化を情報抽出の基本単位とし，数値情報はそのパラメータであるととらえる．MuST コーパスの分析を通じて変化の基本的な特徴とその言語表現の関係を明らかにするともに，それらの特徴の可視化と抽出の方法を提案する．

1　はじめに

可視化は視覚情報による要約と考えることができる．報知的要約であれば，それは要約対象となっている情報の内容をそれだけで概観できるものである必要があるし，指示的要約であれば，それらの情報の取捨選択を可能として情報アクセスを支援するものでなければならない．いずれにせよ，ここで重要なことは「要約対象となっている情報」についての情報を提供するということである．

我々は「MuST:動向情報の要約と可視化に関するワークショップ」[2] に参加しながら，統計量等の時系列情報について述べた文書群を要約可視化して動向情報を得ることを試みている．この動向情報は時系列情報に関するある種のグラフによって視覚的に表現されると考えており，そのようなグラフ描画に必要となる数値情報（時系列情報の様々な時点での値もしくは時点と値の対）を背景となる文書群である新聞記事テキストから抽出することに力点を置いてきた [5][6]．多くの研究グループが同じ観点で研究を進めている [7][3][8]．さて，上述した要約と可視化の役割に鑑みれば，この時系列情報に関するグラフは背景となる文書群に関する情報を提供するためのものであり，その文書群で述べられている情報を反映したものでなければならない．この点を強調しないと，このような研究は，白書等，他の情報源から容易により正確に入手できる情報を，テキスト情報という貧弱な情報源から，わざわざ多大な計算労力をかけて収集するという実用的価値のないものとなってしまう．

このように考えた場合,「グラフ描画を可能とする数値情報を抽出する」ことが本質であるかを改めて考える必要がある．我々は，これまでの研究 [6] で，このような数値情報の抽出が充分でないという問題を解決するために「定性的なグラフ概形情報を抽出するための定性表現の利用」を提案しているが，本稿では，これを一歩すすめて，あるいは，あえて主客転倒し，定性的な表現によってグラフの概形を示唆する「変化」の情報を中心においた情報抽出と可視化を提案する．

本稿の構成は以下の通り．まず，次節で，なぜ変化の情報に着目したかを説明し，続いて，変化の情報がどのように表現されるかを述べる．その後，そのような変化を可視化する仕組みについて述べる．最後に，それをテキストから抽出する枠組みについて述べた後，全体をまとめる．

2　情報の基本単位としての変化

下のふたつのテキストは，MuST コーパスに含まれる原油価格についての新聞記事から抜き出したものである．

> 原油価格（ドバイ原油）も，昨年１０月ごろ１バレル＝約２０ドルをつけたのをピークに下落が続き，今年１月下旬に同約１２ドル５０セントまで落ち込んだ．その後，イラク情勢の緊迫化で一時上昇したものの，現在はまた１２ドル前後で低迷，…（1998/2/14）

> 原油価格は指標となるドバイ原油（８月渡し）が１５日、１８ドル３０セント台まで急伸した．２月には１０ドルだったので８０％以上の上昇になる．(1999/7/17)

このようなテキストから抽出できる情報として最初に目につくのは,「昨年１０月ごろ１バレル＝約２０ドル」等，時点と値の対という一般のグラフにおいてプロット

される点となる情報である，しかし同時に,「下落が続き」「〜のをピークに」等，ある時間幅や特徴的な時点の変化に関しての定性的な情報が多く含まれていることもよくわかる.「今年１月下旬に同約１２ドル５０セントまで落ち込んだ」のように変化の終点や起点として時点と値の対が示されているし，その記事の執筆時点での統計量の値という「基本的な」情報も,「１５日、１８ドル３０セント台まで急伸した」のようにその時点である値となったという変化の一種として表現されている．これらのことから，むしろ，変化の情報が中心的であるといえる．情報の分節も,「つけた」「下落が続き」「落ち込んだ」「上昇した」「急伸した」等の変化に基づいていることがわかる．

加えて,「昨年１０月ごろから下落が続き，今年１月下旬に同約１２ドル５０セントまで落ち込んだ」というような変化に関する情報は，それ自体が３ヶ月間の原油価格の変化についての要約であり，個々の点よりも抽象度が高く，時系列情報の変化の概要，動向を把握するために重要である．同様に「低迷」や「急伸」等は，ある種の評価を含んでおり，そこには単なる数値以上の情報が含まれている．これらの観察が，変化を基本要素とする情報の抽出と可視化の動機となっている．

3　変化の表現

分節されたひとつの変化に関する情報は，多くの場合，テキストの節に対応し，その中心となる動詞によって，表現される変化やその項となる値との関係が規定される．例えば「$10 であった」という表現は，言及されている時点，以下やや正確でないかもしれないが出来事時点（event time）と呼ぶ，の統計量の値が$10 に等しいという情報であるが，値の変化には言及していない．一方,「$10 になった」はその時点の値が$10 に等しいことに加えて，その直前に何からの変化があったことを表すし，逆に「$10 にとどまった」はそこに変化がなかったことを表している．その時点の値との関係は同値関係だけでなく,「$10 を上回っている」の場合は出来事時点の値が$10 より高いことを示している[1]．更に，項となる値との関係は言及されている時点のものに限らない.「$10 を上回った」の場合，上昇方向の変化があり，$10 に等しい時点が出来事時点より過去にあったことを意味している[2]．そして，含意として出来事時点の値は$10 を上回っていることが示される.「$10 に迫っている」の場合は，予測される未来の時点との同値関係が表され，それに向けての方向が明示されない変化があることを示してる．

複合的な動詞を含んだ節，もしくは統語的には単純であるが複雑な意味を持つ動詞によって，変化の始まりや継続等，変化の変化が表現される.「上昇を始めた」は上昇方向の変化が出来事時点で始まったことを示すし,「上昇に転じた」は出来事時点以前に下降方向の変化があったことを示す.「$10 まで盛り返した」というような表現は，出来事時点の値が$10 であること，上昇方向の変化があること，加えて凹の変化があり，値が$10 に等しかった過去の時点があることが意味されている．動詞ではないが,「ピークとして」という表現も，同じように変化の変化に関する言及である．表 1 に代表的な動詞とそれが持つ変化に関する情報を示す.

「上昇」に対する「急騰」,「とどまっている」に対する「低迷」のように，ある種の動詞は，その変化や値についての含意，付随的な意味を持っている．前節のテキスト例からもわかるように，このような語の利用によって，変化やそれに関連する値にある種の評価が付与されている．これらは，表 1 のような形で分類することは難しいが，言語的な情報としてそのまま用いることで，値や変化の意味づけや評価を表現することが可能である．

表 1 にまとめた，動詞が持つ変化に関する情報は，そのまま，変化に注目した情報抽出の基本的なテンプレートを定義する．この枠組みを変化に関する情報のプリミティブ（IPC）と呼ぶ．IPC は，その出来事時点を示す情報に加え，値との関係の情報（type0），変化の情報（type1），変化の変化に関する情報（type2）とそれぞれのパラメータを持つ[3].

type0 の情報があることは値に関する言及があることを示し，パラメータとして，着目している統計量の値をとる．type0 の値は，"dur_eq"，"dur_above"，"dur_below"，"after"，"before"のいずれかであり，前３つはパタメータとなっている値との関係が出来事時点の値との関係であることを示し，それぞれ，その値との同値関係，それを上回っているという関係，それを下回っているという関係を示す．後 2 つは，パタメータとなっている値が出来事時点よりそれぞれ過去，未来のものであることを示す．これらの場合，値との関係は共に同値関係である．

type1 は変化があったかなかったかで，"somechange"もしくは"nochange"をとり，"somechange"の詳細値として上昇/下降に対応する"upward"/"downward"がある．type1 の値が"somechange"の場合は，パラメータとして，「いつ」に比べて「どのくらい」を表現する ref と diff を持つ.

type2 は変化の変化について以下の 4 つの値のいずれ

[1] このような大小関係の表現は項となる名詞句を「$10 以上」として動詞は同値関係を示す「である」を用いることで行うこともできる．このふたつは情報内容としては等価と考えるが，表現に着目した場合，情報の在処が異なるように思う.

[2]「10 日には$11 と，$10 を上回った」という表現を考えて頂きたい.

[3] type1，type2 等の名前は，変化を 1 次微分量，変化の変化を 2 次微分量と考えたことに由来する.

表 1: 動詞句と変化に関する情報

動詞句の表現（例）	値との関係	変化の情報	変化の変化の情報
$10 だった	出来事時点/同値		
$10 になった	出来事時点/同値	変化あり	
$10 にとどまった	出来事時点/同値	変化なし	
（$10 まで）上昇した	（出来事時点/同値）	上昇方向の変化あり	
（$10 まで）下落した	（出来事時点/同値）	下降方向の変化あり	
$10 を { 上回った|超えた }	過去/同値	上昇方向の変化あり	
$10 を { 下回った|割った }	過去/同値	下降方向の変化あり	
$10 に迫った	未来/同値	変化あり	
$10 を上回っている	出来事時点/より大きい		
$10 を下回っている	出来事時点/より小さい		
上昇を始めた		上昇方向の変化あり	開始
上昇を続けた		上昇方向の変化あり	継続
上昇に転じた		上昇方向の変化あり	反転
$10 まで盛り返した	出来事時点/同値	上昇方向の変化あり	凹もしくは凸

かをとる.

"continue" 同じ変化が連続していることを示す.

"start" 現在の変化が直前のそれとは異なることを示す.

"reverse" 現在の変化が直前の変化と逆方法のものであることを示す（start の詳細値）.

"convex" 以前に逆方向の変化があり（値としては凸もしくは凹で）過去に同じ値の時点があったことを示す.

これらは「3ヶ月続けて上昇している」「3ヶ月ぶりに上昇を始めた」の「3ヶ月」を表現するパラメータ dur を持つ．ただし，type2 の情報については更に整理が必要であることが明らかで，"reverse"と対になる"start"の詳細値，つまり，現在の変化の直前に変化がなかったことを示す表現はないのかが疑問であるし,「底を打った」のように直前までの変化が終了したことを表現する値がないのも問題である．また,「3ヶ月続いて上昇し，半年ぶりに$10 まで盛り返した」のような表現は今では2つの IPC となるが，これをひとつでまとめられるように"convex"を別の情報と考える必要があるかも検討する必要がある.

IPC の例を図 1 に示す.「１０日に$10 であった」に対応する IPC は，出来事時点での値との同値関係を述べているいうことで，type0="dur_eq"を持ち，出来事時点の情報と type0 のパラメータとして値を持つ.「上昇した」は type1="upward"という IPC に対応する.「$10 まで上昇した」を表現する IPC はこれらの両方の情報を持つことになる.「3ヶ月ぶりに$10 まで盛り返した」の場合は，type2="convex"が加わる.

「１０日に$10 であった」

```
<ipc date = "19981010"
     type0 = "dur_eq"
     val = "$10" /ipc>
```

「上昇した」

```
<ipc type1 = "upward" /ipc>
```

「$10 まで上昇した」

```
<ipc type0 = "dur_eq"
     val = "$10"
     type1 = "upward" /ipc>
```

「3ヶ月ぶりに$10 まで盛り返した」

```
<ipc type0 = "dur_eq"
     val = "$10"
     type1 = "upward"
     type2 = "convex"
     dur = "00000300" /ipc>
```

図 1: IPC の例

4 変化の視覚的表現

我々の目的は，統計情報等，特定の時系列情報に関心を持ち，それについて述べている一連の文書群について知りたい（アクセスしたい）という利用者の情報要求に対応する枠組みを構築することである．このために，必要な情報を時系列情報のマルチモーダルな要約として利用者に提示し，その概要を理解させるとともに，その要

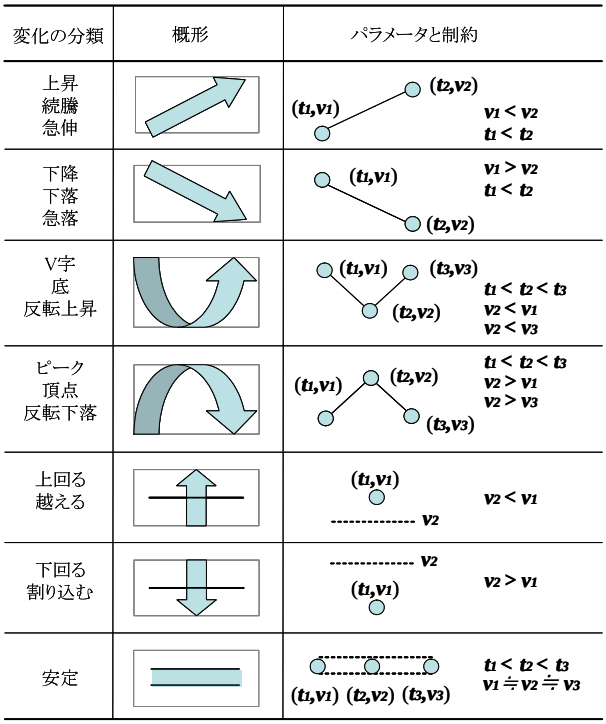


図 2: GPC の例と制約

約を対話的なインタフェースとしてその背後にある文書群への効率的なアクセスを可能にしなければならない．

時系列情報のマルチモーダルな要約は，例えば，それを図示したグラフである．テキストから抽出した数値情報は，そのようなグラフにおける点としてプロットされる．本稿では，変化を表現する情報をシンボリックな概形の配置によって可視化することを提案する．このグラフのシンボリックな概形を変化に関するグラフプリミティブ（GPC）と呼んでいる．GPC は IPC の分類に対応して，その種類が決定され，グラフ上に配置され，変化を表現する．配置される GPC のインスタンスの大きさと位置は変化の期間や幅に対応し，それぞれの GPC が持つパラメータによって決定される．下降を例にとれば，下降を始めた時期とその時点での値と言及されている時点とその値がパラメータとなる．図 2 に GPC の例とそのパラメータを示す．IPC の分類に収まらない「急騰」「低迷」等の補足的な意味を持つ情報の場合は，その動詞そのものを文字情報として GPC に注釈づけることで，グラフ中で表現することができる．このような GPC の利用とそこへの言語的な注釈によって，ここでいうグラフ表現は単なる数値情報の羅列である客観的なグラフではなく，その変化やトレンドを表現し，それに対する評価までを含んだマルチモーダルな要約となる．

テキストから得られる情報で GPC のすべてのパラメータが決定できるわけではない．例えば,「３ヶ月連続で下落した」という表現からは下落の期間は明らかとなるが，その幅，つまりどの値からどの値まで下降したかは明らかでない．これらの値が次節で述べる情報抽出の範囲で明らかにならない場合には，デフォルトの値を与えて配置を行う．このような配置の場合は，配置可能な領域や可能な変形が示され，利用者はその範囲で対話的に位置やサイズを変えることができ，それによってグラフを整えることができる．

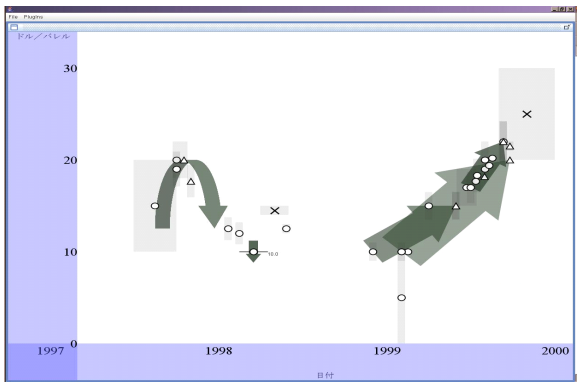
図 3 に GPC を用いたマルチモーダル要約の例を示す．それぞれドバイ原油価格と PHS 加入者数の変化を表現している．ここには GPC に加えて２種類の点と矩形が描かれている．点は「先週末，１バレル＝２０ドル２０セントと２０ドルの大台に乗り」のように変化の項として表現され時点と値の対である数値情報を示し，抽出の過程に応じて異なる記号でプロットされている．矩形は「２２ドル台」「１５ドル前後」「今年前半」「夏頃」等の慨然表現を含む情報の値の範囲を示している．配置された GPC を含むすべてのデータは，それが抽出されたテキストの部分およびそれを含む文書と結びつけられており，利用者は自分が注目した時点の情報に対話的にアクセスすることが可能である．その意味で，このグラフは文書群中に記述された時系列情報のマルチモーダル要約であると同時にそれら文書群へのアクセスインタフェースとなっている．

図 2 に示した現状の GPC は IPC の整理を行う以前に設計したもので,「上昇が始まった」「〜に迫っている」等，一部の IPC は対応する GPC を持たない．また,「急騰」「低迷」等の補足的情報の文字情報による注釈も未実装である．これらについては今後実装を進めていく予定である．

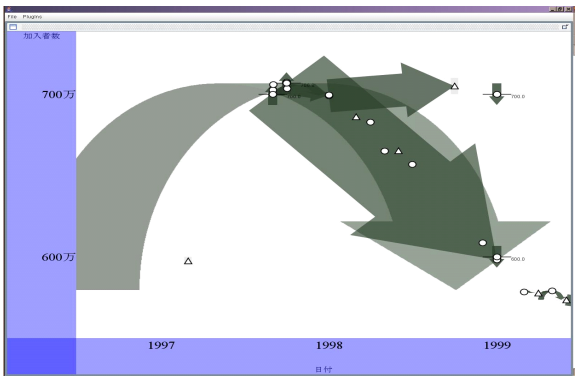
5　変化に基づく情報抽出

時系列情報の基本的単位を変化であると捉えた時，その要約・可視化の素材を自然言語テキストから抽出することをを考える．手法としては，FASTUS[1] など多くのシステムで用いられている典型的な手法である階層的なパタンマッチングによる情報抽出となろうが，一般的なそれに比べて次の２点で特徴的となる．

まず，一般的な情報抽出は，企業合併や役員交替等，対象とする分野の出来事に対応する上位のテンプレートを持ち，その出来事を構成する要素を抽出していくが，本手法では時系列情報の変化に関する一般的な表現に着目し，それぞれの変化に対応する情報をプリミティブと考えることになる．そのため，一般的な情報抽出が（少なくともその上位のレベルで）必然的に分野依存であるのに対し，この手法は多くの時系列情報に幅広く対応でき



(a) ドバイ原油価格



(b) PHS 加入者数

図 3: GPC を用いたマルチモーダル要約

る．もちろん時系列情報に関する表現にも,「$20 で取引されている」「20 万人が新たに加入した」のように対象とする分野に独特の表現があるが,(少なくとも新聞記事においては)「現在の価格は$20 である」「加入者数が 20 万人増加した」のように動詞としては一般的な表現が利用される場合がはるかに多く，それを利用することで，適用範囲の広いシステムを構築できる．この点で，時点と値の対である数値情報を抽出すること比べても困難は大きくならないと考える．加えて，後述する曖昧性解消の問題と関連して，数値情報の抽出でもここで着目する変化に関する表現への考慮は必要であろう．

次に，主に節で述べられている変化の情報に加え，その項として，あるいは同格表現等の言語的道具立てを用いて，随伴的に表現されている値に関する情報の抽出がやはり必要となるので，それに対応できる処理とする必要がある．例えば「ドバイ原油価格は５月には$15 と，昨年末に比べて 50%も急騰した」「ドバイ原油価格は昨年末の$10 から$15 にまで高騰した」からは，上昇という変化だけでなく，言及されている時点の価格が$15 であることと，昨年末の価格が$10 であること，もしくは昨年末に比べて 50%の上昇であることを抽出する必要がある．もちろんその一部は後処理としての演算や推論によるが，そのための材料の抽出は必要であるので，これらの表現に対する情報抽出の仕組みを用意しなければならない．更にこのように抽出される情報は断片的で重複も多いので，抽出された断片的な IPC を組み合わせることで，信頼性の高い総合的な情報の抽出を進めていく必要がある．

処理の流れは以下のようになる．これらの処理の一部はまだ実装が完了しておらず，Step 2, Step 4 に示すパタンマッチの有効性を MuST コーパスを用いて確認するにとどまっている．

Step 1. 与えられた統計量名，ドバイ原油価格等，を入力として，一般的な情報検索システムを用いて，それに関する文書を検索する．

Step 2. 数値表現，時間表現をパタンマッチによって抽出する．

Step 3. 抽出された数値表現，時間表現の存在と統計量名を構成する語の存在を利用して，情報抽出の対象となる文を選択する．

Step 4. 階層的なパタンマッチにより，IPC の情報を決定し，そのパラメータを埋めていく．

Step 5. IPC に含まれる情報を正規化し，それらを組み合わせることで，断片的な情報を完全なものに近づける．

数値表現，時間表現の抽出は固有表現抽出の中でも組織名等に比べ比較的容易であることが知られているので，Step 2 として行い，その後の処理へ情報を提供するのが適切と考えるが，ここでの困難は「$1 上昇して$11 となった」からわかるように，統計量の値そのものであるかその差分であるかが表現が出現する文脈をみないと明らかでない点である.「50%」のように一般には比率を表すものも政党支持率のような統計量では値そのものとなり，その区別も自明ではない．同様に時間表現も，グラフにおいて点となる時点の表現と時間幅の表現が重なる場合がある[4]．これらについては Step 4 のパタンマッチングの副作用として曖昧性を解消する．なお，時間表現については「去年」等の相対的表現を絶対時間に変換するいわゆる時間処理 [4] が必要であるが，これは Step5 の正規化の一部として行われる．

[4]前者は MuST の注釈付けにおいて `val` 要素と `rel` 要素の区別の問題，後者は `date` 要素と `dur` 要素の区別の問題に対応する．Step 2 の出力はこれらの曖昧さを残し，かつ `name`, `del`,`ins` 要素を含まない MuST コーパスとなる．

Step 2 と Step 4 のパタンマッチはすべて属性を持ったタグによって注釈づけられたテキストに新たなタグ付けを繰り返すという階層的パタンマッチングの枠組みで行われる．Step 4 で用いられるパタンマッチによる注釈付けの規則の例を以下に示す．それぞれにおいて，==>の左辺にマッチした部分が右辺に書き換えられる．新しく注釈づけられるタグの属性の値は既存の要素の属性から導かれる．<\xyz>は<xyz>...</xyz>の略記で，xyz 要素の開始タグと終了タグで囲まれた部分を示す．右辺における$0 は左辺にマッチした部分全体を示し，$n は左辺に現れる n 番目の要素もしくはグループを示す．val, rel, date は，それぞれ統計量の値，その差分や割合，時間を表現する要素である．右辺に現れる ipc 要素は IPC に対応する要素でその属性によって情報内容が表される．

Step 4 の規則は，慣用的な複合表現をまとめあげるレベル，主に同格表現によって並べられた複数の情報をまとめあげるレベル，節として表現された IPC を得るレベル，「3 ヶ月ぶりに〜となった」「3 ヶ月連続で〜だった」のような内側に IPC を含む情報を得るレベル，統計量名と時間情報を組み入れるレベルに分けられ，パタンマッチの階層をなす．ここに示した例は最初の 3 つのレベルからとっている．

```
<date>に比べて<\rel> ==>
   <rel ref=$1 diff=$2>$0</rel>

<\val>と，<\rel> ==>
   <rel ref=$2.ref diff=$2.diff>
     <ipc type0="dur_eq" val=$1>
       $0</ipc></rel>

<\rel>(も)?{ 急騰|急伸|...} した ==>
   <ipc type1="upward"
        ref=$1.ref diff=$1.diff>
     $0</ipc>
```

「ドバイ原油価格は 5 月には$15 と，昨年末に比べて 50%も急騰した」をこれらの規則を使って解析した例を図 4 に示す．このテキストは Step 2 によって図の最上段のように注釈づけられている．ここで<val|rel>は数値表現についての曖昧さを残したタグである．以下，簡単のために各要素の終了タグをすべて</>で表現する．これに先程の規則が順次適用される様子を図に示す．ここで，2 番目から 3 番目への書き換えで,「$15 と，昨年末に比べて 50%」がまとめあげられ，その後は単純な割合差分表現である「50%」や「$1」等と同じように扱われるが，この時点で,「$15 と」に関する情報抽出が随伴的に行われる．最終的に得られた結果には以下のふたつの IPC が含まれる．

```
<ipc type0="dur-eq"val="$15">...</>
<ipc type1="upward"
     ref="昨年末" diff="50%">...</>
```

前者はこれが出来事時点の値の同値関係に関する情報でその値が$15 であること，後者は上昇方向の変化に関する情報で，昨年末時点からの 50%の変化であることを示している．前者が随伴的な情報として獲得されている点に注意されたい．このふたつの IPC は統合することが可能で，得られる情報はオリジナルのテキストが表現しているそれに他ならない．

Step 4 の処理の対象となる文は Step 3 によって選択されたものであるが，ひとつの文に複数の統計量に関する情報が含まれている場合もあるので，得られた IPC が目的とする統計量に関する情報であるかを確認する必要がある．これもパタンマッチによるが，統計量の名前はその表現に大きな広がりがあり，それを数値表現のように上昇的に，つまり，名前表現がどのようなパタンを持つかを記述することで，見つけ出すことは難しい．このため，以下のような統計量名が来やすい文脈を指定する規則を用い，そこを name 要素とすることで，統計量の名前であろう部分を同定する．そして，この要素の内容が目的とする統計量の名前として矛盾しないかによって確認を行う．統計量名の抽出について，それを値や日付の要素と同様の上昇的なパタン記述によらず，下降的に，その文脈の記述によって行っていくという点は，統計量名の複雑さとその分野依存性を考えた時，重要な提案になっていると考えている．

```
(.+) は<\date>には<\ipc> ==>
                        <name>$1</name>
```

Step 5 では，矛盾のない IPC を統合していくことで断片的な情報を繋ぎ合わせて総合的な情報の抽出を進めていく．

6 おわりに

統計情報等，特定の時系列情報に着目し，それについて述べている一連の情報について知りたいという情報要求に着目し，それらに応える情報アクセスインタフェースとなりうるマルチモーダルな要約について検討した．時点と値の対である数値情報よりも，むしろ時系列情報の変化の情報が重要であることを指摘し，その点に注目して，それら変化に関する定性的情報を言語情報と合わせてグラフ化することで，単なるデータの羅列である「客観的な」グラフではなく，そのトレンドを表現しそれに対する評価までを含んだマルチモーダルな要約が行える

```
ドバイ原油価格は<date>５月</>には
<val|rel>$15</>と，<date>昨年末</>に比べて<val|rel>50%</>も急騰した
⟹
ドバイ原油価格は<date>５月</>には
<val|rel>$15</>と，
<rel ref="昨年末" diff="50%"><date> 昨年末</>に比べて<rel>50%</></>
も急騰した
⟹
ドバイ原油価格は<date>５月</>には
<rel ref="昨年末" diff="50%"><ipc type0="dur_eq" val="$15">
 <val>$15</>と，
 <rel ref="昨年末" diff="50%"><date>昨年末</> に比べて<rel>50%</></></></></>
も急騰した
⟹
ドバイ原油価格は<date>５月</>には
<ipc type1="upward" ref="昨年末" diff="50%">
 <rel ref="昨年末" diff="50%"><ipc type0="dur_eq" val="$15">
  <val>$15</>と，
  <rel ref="昨年末" diff="50%"><date>昨年末</>に比べて<rel>50%</></></></></>
 も急騰した</>
```

図 4: 階層的パタンマッチングによる変化の情報の抽出

ことを示した．加えて，そのような変化に関する情報を抽出するための，時系列情報の変化をプリミティブとする適用範囲の広い情報抽出方式を提案した．

今後は，まず，システムの実装を行い，様々な種類の統計量で提案手法が有効であるかを確認する．また，IPC の統合については，時間推論や慨然推論等の応用が期待できるので，その点も掘り下げていくことを考えている．更に，現在の検討は時間のみを自由変数とする統計量に限られているが，それ以外のパラメータを持つもの，政党支持率や業界各社のシェア等も扱えるように拡張していきたい．

参考文献

[1] Hobbs, J. R., Appelt, D., Bear J., et al.: FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. Roche, E., Schabes, Y. (ed)*Finite-State Language Processing* The MIT Press (1997).

[2] Kato, T., Matsushita, M., and Kando, N.: MuST: A Workshop on Multimodal Summarization for Trend Information. *Proc. NTCIR-5 Workshop Meeting* (2005) 556–563.

[3] 今岡裕貴，桝井文人，河合敦夫，井須尚紀．動向情報抽出における相対表現の利用効果に関する考察．知能と情報（日本知能情報ファジィ学会論文誌） Vol.18, No.5, pp. 735-744, 2006.

[4] Mani, I. and Wilson, G.: Robust Temporal Processing of News. *Procs. the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)* (2000) 69–76.

[5] 松下光範，加藤恒昭．動向情報に基づく情報可視化の基礎検討．第 19 回人工知能学会全国大会　1E3-03, 2005

[6] 松下光範，加藤恒昭．数値情報の補填とグラフ概形の示唆による複数文書からの統計グラフ生成．知能と情報（日本知能情報ファジィ学会論文誌） Vol.18, No.5, pp. 721-734, 2006.

[7] 難波英嗣，国政美伸，福島志穂，相沢輝昭，奥村学．文書横断分関係を考慮した動向情報の抽出と可視化．情報処理学会研究会 NL168-11, pp. 67–74 (2005).

[8] 曽我直也，斉藤博昭．動向情報提示システムの構築．言語処理学会第 12 回年次大会ワークショップ「言語処理と情報可視化の接点」　pp.5–8, 2006.