談話構造に基づく語彙選択を行うXMLデータベースからのテキスト生成

関 洋平[†] 原田賢一[††]

本論文は，データベースからの報告書の生成というテーマに関して新たな実現方法を提案するものである．我々は，XML 変換に基づく日本語と英語およびフランス語とドイツ語の天気予報を出力する自然言語生成システムを，標準的な 3 段階パイプライン方式に基づいて実現した．人が作成するようなテキストを計算機が生成するためには，結束性の概念を考慮したうえで個々の文の構造を決定しなければならない．本論文において，DOM，SAX，XSLT などの XML の標準的な技術を使用することによって，談話単位ごとの局所的な制約に基づいて表層文を実現する方法について提案する．また，日本語と英語およびフランス語とドイツ語の 4 つの言語のテキスト生成を対象とすることによって，ドメインを対象とした伝達意図に基づく談話の構造の選択と，個別の出力言語に依存する談話中の個々の文の構造および表層語彙の選択との境界線を明確に示した．

Text Generation from XML-DB with Lexical Selection Based on Discourse Structure

Yohei Seki[†] and Ken'ichi Harada[††]

The purpose of this study is to propose a new method for the generation of reports from databases. We implemented an XML transformation-based natural language generation system for Japanese, English, French, and German weather forecast by using a three-stage pipeline architecture. In order to generate human-like texts, we must determine individual sentence structures by considering the concept of "cohesion". In this paper, we propose the method on how to realize surface sentences with local lexical constraints on discourse segment by using DOM, SAX, and, XSLT technique. We regard this process as text generation based on intentional structure to retrieve information from databases, and make a distinction between the selectional process of intentional strucuture based on domain knowledge and of linguistic structure by producing four language texts.

1. はじめに

自然言語処理の技術が発達するにつれて，文生成システム，すなわち計算機が素朴なデータ集合から人間に近いかたちの文章を生成する応用への期待が高まっている[1)]．自然なテキストを計算機上で生成するためには，人間のコミュニケーションにおいて重要な結束性（cohesion）を考慮して個別の文の構成を決定する必要がある．

我々は，この目的の実現のために，データの獲得意図の構造に対応した談話構造[2)]に基づいて，個々の文を生成するというアプローチを提案する．具体的には，自然言語生成における標準的な 3 段階パイプラインアーキテクチャ[3)]に基づいて，気象庁年報のデータを入力として予報文章を生成するシステム[4),5)]を実現した．出力は，日本語と英語およびフランス語とドイツ語の 4 つの言語を対象とした．作成したシステムを**図 1** に示す．なお，このシステムは天気予報以外に経済ドメインについても，月例経済報告の文章を生成できる．このシステムは，上側のプルダウンメニューを選択し，左側のボタンを押すとメニューの選択条件に対応するそれぞれの段階の出力結果が右下の窓に表示されることになる．

人間が天気予報を通して得たい情報は，その目的によって構造化されており，ある程度固定化されている．この情報を獲得したい意図の構造に対応した談話構造を構成することによって，人間の言語コミュニケー

† 青山学院大学理工学部情報テクノロジー学科/総合研究大学院大学情報学専攻
Department of Integrated Information Technology, Aoyama Gakuin University/Department of Informatics, The Graduate University for Advanced Studuies (Sokendai)
†† 慶應義塾大学理工学部
Faculty of Science and Technology, Keio University

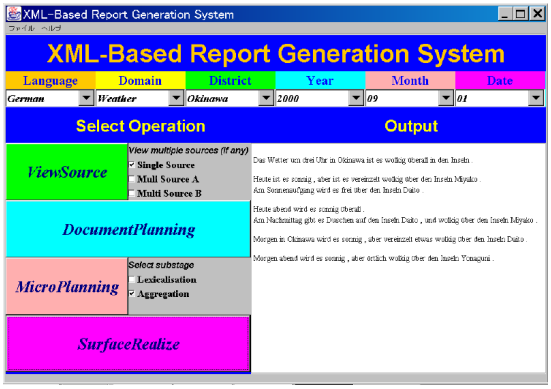


図 1 天気予報生成システム
Fig. 1 Weather forecast generation system.

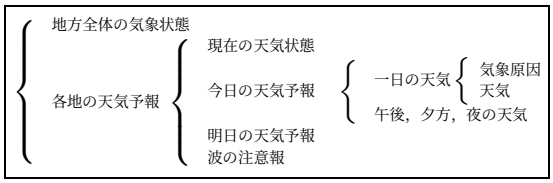


図 2 天気予報の談話構造（3 章で解説）
Fig. 2 Discourse structure on weather forecasts.

ションを意識した生成システムが実現できる．談話構造は，談話単位に基づいて支配と充足先行の 2 つの関係により定式化され，木構造として表現できる．天気予報の談話構造には，「気象→天気」「今日の天気→明日の天気」などの充足先行関係がある．支配関係も合わせたおおまかな構造を**図 2** に示す．

談話構造の実現にあたっては，XML（eXtensible Markup Language）を利用した．XML 形式で格納する利点として，談話単位に対応したタグでデータ集合を囲むことによって，データベースの入れ子構造を横断した検索および内容の決定が可能となる．

本研究の入力データは，気象業務支援センターの提供する気象庁年報 CD-ROM 2000 年版ならびに気象庁月報 CD-ROM 2001 年 9 月と 10 月版☆を，XML 形式に変換して，XML データベースである Yggdrasill☆☆に格納したものを利用する．

データは，特に，北海道地域と沖縄地域の 2 つの地域内の観測所のデータを使用した．北海道地方は，観測点の分類に応じて 2 つの入力データを構成している．それぞれに対する DTD☆☆☆形式とその差分を付録 A.1.1，A.1.2 に示す．沖縄地方のデータに対する

☆ http://www.jmbsc.or.jp/offline/cd0040.htm
☆☆ http://www.mediafusion.co.jp/seihin/ygg/index.html
☆☆☆ Document Type Definition，文書型定義

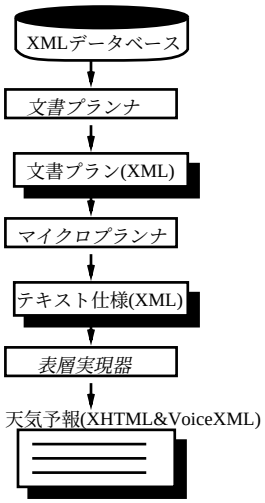


図 3 3 段階パイプライン方式による生成システムのフローチャート
Fig. 3 Flowchart for three-stage pipelined NL generation system.

DTD 形式の差分は，付録 A.1.3 に示す．

天気予報生成機構の実現に際しては，自然言語生成の標準的なアーキテクチャである 3 段階パイプライン方式[3] を採用した．3 つの段階はそれぞれ，文書プランニング，マイクロプランニング，表層実現と呼ばれる．3 段階パイプラインアーキテクチャによる文生成の処理過程を**図 3** に示す．

本研究では，入力データから談話構造に沿って，データを再構成する過程を文書プランニングとする．また，マイクロプランニングは，文書プランニングの出力を入力として，生成語彙の選択，文単位の集約，参照表現の生成を行う．この過程は，談話構造の順序を意識したタグの変換として，SAX（Simple API for XML）を利用して実現する．

しかし，談話構造に基づいてデータの順序を構成して生成語彙が選択できれば，そこから適切な文章が生成できるわけではなく，談話役割に応じて個々の文を適切に変換して生成する必要がある．本論文では，表層実現器において，談話役割の違いに応じて個々の言語表現の違いを生成することにより，適切な予報文章を生成する手法について提案を行う．特に，日本語と英語を生成し分ける際に重要な，大文字と小文字の区別や助動詞の活用など，それぞれの言語特性に応じた情報が談話構造から簡単に利用でき，望ましいテキストを実現できることを示す．

この実現にあたっては，XML の変換提示用言語である XSLT（eXtensible Stylesheet Language Transformations）を利用した．特に，表現の言い換えのた

A. 沖縄地方の天気
(1) 沖縄地方は，高気圧の範囲内にあっておおむね晴れています．
(2) 今日は，本島地方では気圧の谷の影響で雲が広がり，所によってはにわか雨か雷雨があるでしょう．
(3) 先島諸島や大東島地方では引き続き高気圧の範囲内で晴れる見込みです．
(4) 明日は，各地方とも高気圧の範囲内でおおむね晴れますが，所によってはにわか雨か雷雨があるでしょう．
(5) 沿岸の海域では各地方とも波がやや高い見込みです．
B. 北海道地方の天気
(1) 北海道付近は，寒冷前線が通過中で，今夜には上空 1500 メートルに 1 1 月上旬並の強い寒気が入る見込みです．
(2) 03 時の道内の天気は，太平洋側とオホーツク海側の一部で晴れていますが，日本海側や北部では雨が降っています．
(3) 今日は，日本海側やオホーツク海側は，曇り時々雨でしょう．
(4) 太平洋側西部は昼頃まで雨の降る所がある見込みです．
(5) 太平洋側東部は曇りで，一時雨が降る見込みです．
(6) また，夜には峠や山間部で雪に変わる所があるでしょう．
(7) 明日は，日本海側では朝の内まで雲が多いですが，のち晴れるでしょう．
(8) その他の地方は，晴れるでしょう．
(9) 網走西部では，河川の水位が高くなっていますので引き続き浸水に注意して下さい．
(10) 海の波の高さは，今日は日本海側やオホーツク海側では 3 メートルですが，太平洋側では 2 メートルのち 3 メートルでしょう．
(11) 明日は，各海域とも初め 3 メートルですが，のち 1 メートルから 2 メートルの見込みです．

図 4 天気予報の例文（気象庁発表,（株）CRC ソリューションズ提供）
Fig. 4 Weather forecasts example.

めに，談話役割の局所的な制約を，XSLT の変換規則テンプレート中のパラメータとして実現し，広域的なスタイル制約などと区別して整理することに成功した．これは，文献 6) によるクラスタリングに基づく言い換えのための提案より，実現の簡易さの点で優れている．以上から分かるように，入力データは，3 つのすべての段階において XML 形式で統一した．

本論文の構成についてまとめておく．2 章では，Web 上の天気予報文書から抽出した天気データと予報表現について整理する．3 章では，2 章の表現の分類に基づく天気予報文書の文章構成パターンについて説明する．4，5，6 章では，格納データから情報を提示する文章生成技術について説明する．7 章では，本研究におけるシステムの 3 段階それぞれの実現について定性的に，出力文について定量的に評価する．8 章で関連研究について述べ，9 章でまとめを行う．

2. 天気予報表現の抽出と分類

本研究では，天気予報の提示を，天気データ（「晴れ」「曇り」「雨」）からの文書生成システムとして実現する．そのために，まず，実際の天気予報の文書から天気データと予報表現の構造を分離して抽出する．抽出の対象データとして，WeatherEye☆のサイトにおける「お天気概況」の天気予報の 1 カ月分の文書集合を使用した．最初に，茶筌（WinCha，Version 2.1）を使用して形態素解析にかけ，次に Perl（ActivePerl, Version5.6.1）を使用して簡易パーサを作成して名詞句と前後の接続表現を頻度順に抽出する．続いて，名詞句を天気予報の専門用語として分類する．以上に基づき，予報生成システムの入力データと予報表現の基準について考察する．また，基本構成要素として,「天気」「時間」「場所」「気象」を分類した．

2.1 天気予報生成に必要なタグ付け

天気予報において使用される文章は，同じ「晴れ」でも文脈に応じてさまざまな表現を使用する．例として，WeatherEye サイトの「お天気概況」の文章を**図 4** に示す．この短い文章でも,「晴れ」の天気について「晴れています」「晴れる見込みです」「おおむね晴れますが」「晴れているところがおおいですが」「晴れますが」「晴れていますが」「晴れるでしょう」と，7 種類もの表現がある．

本研究の目的のために，文章を提示する場合について,「晴れ」と接続表現を分離したうえで格納することを考える．この応用に望ましいデータ表現の 1 つとして，XML 形式を用いることが考えられる．図 4 の例文をタグ付きデータとして格納する一例を**図 5** に示す．このタグ付けは，XSLT（eXtensible Stylesheet Language Transformations）を使用してデータ変換を行うことにより，図 4 の文章を提示できることから，情報提示用のタグ付けとして適切である．しかし，表現の属性値（「文末」の値など）に関して一般性に欠ける点がある．

☆ http://weather.crc.co.jp/index.html，気象庁発表,（株）CRC ソリューションズ提供

<文章 地方="沖縄">
<文 句点="。"><天気 属性="おおむね" 文末="ています">晴れ</天気><場所>沖縄地方</場所><気象 文末="の範囲内にあって">高気圧</気象></文>
<文 句点="。"><前 接続="">><文><天気 文末="が広がり">雲</天気><場所 助詞="では">本島地方</場所><気象 文末="の影響で">気圧の谷</気象><時間>今日</時間></文></前><後><文><天気 属性="所によっては" 文末="があるでしょう">にわか雨か雷雨</天気></文></後></文>
<文 句点="。"><天気 文末="る見込みです">晴れ</天気><場所 助詞="では">先島諸島や大東島地方</場所><気象 属性="引き続き" 文末="の範囲内で">高気圧</気象></文>
<文 句点="。"><前 接続="が"><文><気象 属性="各地方とも" 文末="の範囲内で">高気圧</気象><天気 属性="おおむね" 文末="ます">晴れ</天気><時間>明日</時間></文></前><後><文><天気 属性="所によっては" 文末="があるでしょう">にわか雨か雷雨</天気></文></後></文>
<文 句点="。"><天気 属性="各地方とも" 文末="見込みです">波がやや高い</天気><場所 助詞="では" 文末="各地方とも">沿岸の海域</場所></文>
</文章>

図 5 タグ付けの一例
Fig. 5 A tagging example.

2.2 「天気」データの抽出と分類

「天気」の表現は前後関係に応じて変化する．前後の活用を「基本表現」・「修飾表現」・「文末表現」として分類した結果を以下に示す．括弧（）の中が実際に出現した表現であり，中括弧 {} を使用して省略可能か選択可能な表現を示す．以下の分類もこの表記に従う．

(1) 基本表現

「基本表現」は天気を表現する基準であり，入力天気データに対応する．

A. 「晴れ」の表現（晴れ）
他の天気との組合せの表現（晴れまたは快晴，晴れまたは薄曇り，晴れている所も次第に曇り，晴後曇り）

B. 「曇り」の表現（曇り，雲，曇，雲って）
他の天気との組合せの表現（曇りで，一時雨，曇りで雨，曇って雨，曇りで所々で晴れ）

C. 「雨」の表現（一時雨や雷雨，一時雨，一時雨か雷雨，にわか雨か雷雨，一時にわか雨）

D. その他，「波」，「風」について天気の表現が見られるが，入力データの構成上，以下の表現も含めてここでは省略する．

(2) 修飾表現

「修飾表現」には，程度，時間，場所およびその組合せがある．

A. 程度（おおむね，次第に，大体，強い）

B. 時間（朝のうちまで，日中，朝晩を中心に，引き続き，現在）

C. 場所（所によっては，各地とも，所ところで，各地方とも）

D. 組合せ（日中は引き続きおおむね）

(3) 文末表現

「文末表現」の変化の基準は，現在の状況と予想の区別である．その他，一部の地域に対する場合，後ろに別の天気表現が続く場合などに応じて表現が変化する．

A. 晴れ
- 現在の状況（〜ています，〜ているところが多いです）
- 予報（〜ます，〜る見込みです，〜るでしょう）
- その他（〜て残暑が厳しいでしょう）

B. 曇り
- 現在の状況（〜となっています，〜っています）
- 予報（〜となるでしょう，〜でしょう，〜る所があるでしょう，〜の所が多いでしょう，〜の所が多くなるでしょう，〜となる見込みです）

C. 雲
- 現在の状況（〜が広がっています，〜が広がり，〜の出ている所があります）
- 予報（〜が広がってくるでしょう，〜が広がりやすいでしょう）

D. 雨
- 現在の状況（〜が降っている所があります）
- 予報（〜の降る所が { ある見込みです，あるでしょう }，〜が降りだすところがあるでしょう）

E. 一時雨（か，雷雨，にわか雨）
- 予報（〜{ になる，となる，の } 所があるでしょう，〜になる所もある見込みです，〜の所がある見込みです，〜となり，〜の残るところもある見込みです，〜があるでしょう）

以上の例から，実際に出現する表現の種類は多いことが分かる．本研究では，天気データから文脈に応じてさまざまな表現を生成できる天気予報生成システムを作成する．そのために，上記の「基本表現」を天気データとし，「修飾表現」は天気データに付属する属性値から生成することを考える．

また，「文末表現」については，属性値から生成するものと，よりおおきな文章レベルの構造に依存するものを区別する．後者については，談話単位の構成に応じて条件判定を行い予報表現を選択して生成することを考える．詳しくは 3 章で説明する．

2.3 「時間」・「場所」表現の分類

天気データの生成のための形式を決めるにあたっては，「天気」に直接関わる要素だけではなく，他の情報を考慮しなくてはならない．次に，「時間」「場所」の

- 明日，今日，午後，昼過ぎ，夕方，夜，09 時，明け方，朝晩，現在
- 今夜から明日，日中から夜，明日は午前中，今日 09 時の実況，今日，明日，今日・明日，今夜 21 時の予想，明日も，昼過ぎ，明日は，午前中

図 6 時間の出現語彙集合
Fig. 6 Occurring time-related phrases.

- 北部，南部
- 各地，道内の天気
- 日本海側やオホーツク海側，日本海側北部，北部や太平洋側の一部，太平洋側東部，日本海側，太平洋側，関東近海，東シナ海
- 山沿いの地方，山間部
- （気圧配置など）本州付近，奄美地方の北部や十島村，他，北部や十島村
- 北海道，宮城県，内海，本島地方，岐阜県の一部，九州南部，鹿児島市上空，先島諸島や大東島地方
- 沖縄地方，東海地方，東北地方，四国地方，中国地方，近畿地方，九州北部地方，関東甲信地方
- 九州北部地方の沿岸の海域，沿岸の海域，海上，東北地方は，太平洋側の一部

図 7 場所の出現語彙集合
Fig. 7 Occurring place-related phrases.

表現について整理する．なお，このほかに「気象」「注意」と呼ばれるグループも分類しているが，ここでは入力データの構成の都合上割愛する．

A. 「時間」の表現
「時間」の表現には，ある特定の時点を示すものと，時間の範囲を示す表現がある．**図 6** に具体例を示す．

B. 「場所」の表現
「場所」の表現には，ある特定の地方，広い地方，海岸，相対的方角などさまざまな表現が用いられている．**図 7** に例を示す．

3. 文章構成パターンの抽出

文章は文の連続から構成される．文章の構成を理解するためには，節および文の性質を分類した後，その構成について調べる必要がある．本研究では，文の性質を前章の「天気」「気象」の表現に基づいて分類する．続いて，分類に基づいた構成パターンを抽出しながら，区別できない要素に必要な基準を考えて，文の性質をより詳細に分類する．本章では，以上の手順を，図 4 の例に基づいて示すことで，構成パターンと必要なデータについて説明する．

3.1 文の分類とその構成

まず，図 4 の文章を 2 章の分類に基づき，文章の構成について考える．

A. 沖縄地方の天気
(1) 気象・現在の状況 → 天気・現在の状況
(2) 気象・予報 → 天気・予報
(3) 気象・予報 → 天気・予報
(4) 気象・予報 → 天気・予報
(5) 天気・予報

B. 北海道地方の天気
(1) 気象・現在の状況
(2) 天気・現在の状況
(3) 天気・予報
(4) 天気・予報
(5) 天気・予報
(6) 天気・予報
(7) 天気・予報
(8) 天気・予報
(9) 注意
(10) 天気・予報
(11) 天気・予報

以上から，(i)「現在の状況について記載した後，予報について記載する」，(ii)「気象原因を先に記載した後，天気について記載する」という 2 つの記載順序に関する構成規則が得られる．ただし，文末表現の分類が 2 種類のみであることから，それ以外の情報は得られない．そこで，次に「天気・予報」を区別する基準について考える．

まず，予報の対象としての「時間」要素に着目すると，「今日」と「明日」が区別できる．この情報を加えると，次のようになる．

A. 沖縄地方の天気
(1) 気象・現在の状況 → 天気・現在の状況
(2) 今日・気象・予報 → 今日・天気・予報
(3) 今日・気象・予報 → 今日・天気・予報
(4) 明日・気象・予報 → 明日・天気・予報
(5) 明日・天気・予報

B. 北海道地方の天気
(1) 気象・現在の状況
(2) 天気・現在の状況
(3) 今日・天気・予報
(4) 今日・天気・予報
(5) 今日・天気・予報
(6) 今日・天気・予報
(7) 明日・天気・予報
(8) 明日・天気・予報
(9) 注意
(10) 今日・天気・予報
(11) 明日・天気・予報

(i)	「現在の状況について記載した後，予報について記載する」
(ii)	「気象原因を先に記載した後，天気について記載する」
(iii)	「今日についての天気予報を記載した後，明日についての天気予報を記載する」
(iv)	「波についての天気予報は (iii) の条件に関係なく最後に記載する」
(v)	「天気の変化がある地方を，同じ天気が継続する地方より優先して予報を記載する」
(vi)	「地方全体についての予報を記載した後，山沿いの地方についての予報を記載する」
(vii)	「広い時間帯の天気予報を記載した後，細かい時間帯の天気予報について記載する」
(viii)	「同じ意味の表現を交代で用いる」

図 8 図 4 から抽出した構成パターン
Fig. 8 Construction patterns.

以上から，(iii)「今日についての天気予報を記載した後，明日についての天気予報を記載する」，(iv)「波についての天気予報は (iii) の条件に関係なく最後に記載する」 という 2 つの記載順序に関する構成規則が得られる．

以上の 4 つの規則をしても区別できない要素に注目してみよう．この例では，(A．2〜3)，(B．3〜6)，(B．7〜8) の 3 つの個所が該当する．この 3 カ所を構成する基準について考える．(A．2〜3) についてはどちらが先でも大差ないが，(v)「天気の変化がある地方を，同じ天気が継続する地方より優先して予報を記載する」という規則が考えられる．(B．3〜6) については (vi)「地方全体についての予報を記載した後，山沿いの地方についての予報を記載する」 (vii)「広い時間帯の天気予報を記載した後，細かい時間帯の天気予報について記載する」という 2 つの規則が考えられる．

最後に，予報の文末表現としてよく現れる「でしょう」と「見込みです」の区別であるが，これについては (viii)「同じ意味の表現を交代で用いる」ことで，表現の単調さを避けるという規則があてはまる．

3.2 構成パターンとその実現

図 4 の例文から抽出した構成パターンについて **図 8** にまとめる．この構成パターンのいくつかは，McKeown のテキストスキーマ，および Mann の修辞構造理論（RST）[7] の考えと共通する．テキストスキーマはシステムが所有するテキスト構造の知識へのパターンマッチングに基づきテキストを計画立案する技術であり，XML-DB からの文章の生成の実現に適した技術の 1 つと考えられる．最近の位置付けは文献 8) が参考になる．

図 8 に出現する構成パターンにはいくつかのレベルがある．すなわち，(i) は，「天気」の文末表現，(ii) は「気象」「天気」の構成要素順序，(iii) は「時間」の要素，(iv) は，「天気」の要素，(v) は，「時間」を通した「天気」の要素の比較，(vi) は，「場所」の属性，(vii) は，「時間」の属性の詳細化，(viii) は，文末表現によって決定する．このように，文章を適切に構成するためにはさまざまな水準の情報が必要となる．

以下の章では，以上の事実を考慮した予報文章生成システムについて説明する．システムは，文書プランニング（document planning），マイクロプランニング（microplanning），表層実現（surface realization）の 3 段階パイプラインモデル（three-stage pipeline model）に基づいて実現する．文書プランニング，マイクロプランニング，表層実現の詳細については，それぞれ 4，5，6 章で説明する．これは，文献 3) にも見られるように，自然言語の文章生成についての標準的なモデルである．

4. 文書プランニング

文書プランニングについて説明する前に，日本語と英語の生成処理の違いについて簡単に説明しておく．日本語と英語の文書プランニング処理はまったく異なるところはなく，出力結果である文書プランも共通のものを使用する．マイクロプランニング処理に関しては，英語の前置詞の取扱いおよび表層語彙の選択を除いては，共通の処理が行われる．ただし，語彙が異なることから，出力結果であるテキスト仕様は異なる．本質的な生成の仕分けは表層実現の段階が担当する．

文書プランニングは，文生成過程において最も重要な内容の決定（Content Determination）と文書の構造化（Document Structuring）の 2 つのタスクを行う．本研究では，1 章で説明したように，この過程を気象庁年報を入力データとして格納した XML-DB からの検索結果を構造化することで実現する．実現にあたっては，Java を使用した．XML-DB は Yggdrasill を使用しており，Yggdrasill の API は，COM-Interface として構成されていることから，検索結果を Java から使用することができる．また，検索にあたっては XBath☆と呼ばれる XPath に基づいた独自の検索言語を採用しており，時間や場所に応じて天気データの

☆ XBath の仕様については [9] を参照

集合を構成できる．

4.1 観測時間に応じたデータの集約

本研究の対象は，北海道地方および沖縄地方のある1日の天気予報概要文書を生成することであり，対象となる観測所は，北海道が22カ所，沖縄が7カ所となる．このうち，天気の観測を行っている地点は時間帯とともに変化する．本研究では，3章の談話構成パターン（図2の談話構造参照）に基づき，1回の文書プランニングに対して，3時，その日の昼間，6時，9時，12時，その日の夜，15時，18時，21時，翌日の昼間，翌日の夜の天気のデータ集合についてこの順序で天気の集合を地域別に検索する．検索したデータを，DOM（Document Object Model）に基づいてデータを階層的に構造化することにより，出力結果をXML形式で得た．これは文書プランと呼ばれる．次に，観測所の地域について説明する．

4.2 観測所の地域による構造化

本研究では，北海道地方と沖縄地方の観測所を，天気予報文書中に頻出する地域名に基づき，構造化を行い，それぞれの地域に対して天気集合を検索する．それぞれの構造は，**図9**，**図10**のように表される．北海道地域については2つの構造を作成した．次のマイクロプランニングの段階では，2つの構造に基づくテキストの仕様を作成して，それぞれの天気について記述量の少ない仕様を選択する．

出力となる文書プランの結果を付録A.2に示す．検索結果である各地の天気の集合は，各時間および今日と明日の昼と夜を単位として天気事象（WeatherEvent）を構成する．1つのWeatherEventの中では，それぞれの地域ごとに天気の分布確率を階層的に構成している．付録から分かるように，分布確率の値が100%である場合には，その構成要素である部分地域（および観測所）は展開しない．また，同一地域内の複数の観測所（例：与那国島地方と石垣島地方）の天気が同じ場合は，観測地点名の置き換え処理（例：八重山地方）を行う．

内容の決定にあたっては，内容の決定に対する2段階モデル[10]と類似の方法を本研究で実現する．内容の決定の2段階モデルとは，時系列データに対する要約手法の1つとして，入力データの質的な概要を作成したうえで参照して内容の決定を行うという提案である．本研究では，北海道地方および沖縄地方の全観測点に対する天気の分布確率を検索して概要とする．全地方に対するデータ集合と地域別のデータ集合をそれぞれ1つの問合せ（XBath）を使用して実現できることが，XML-DBに格納したデータに基づいて文生成

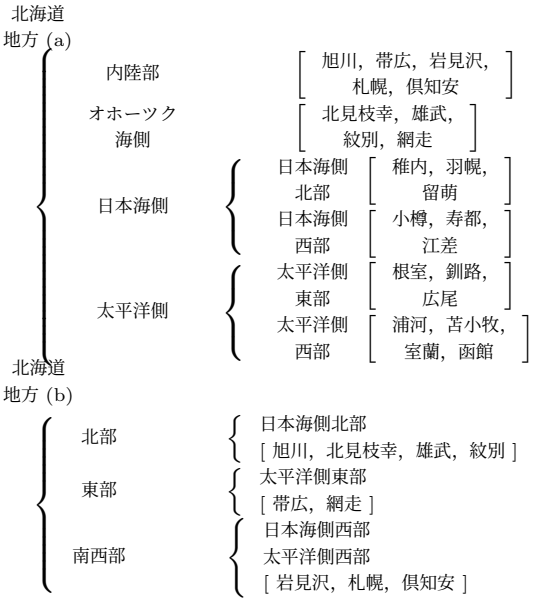


図9 北海道地方の観測所の地域による構造化
Fig. 9 District structure for Hokkaido observatory points.

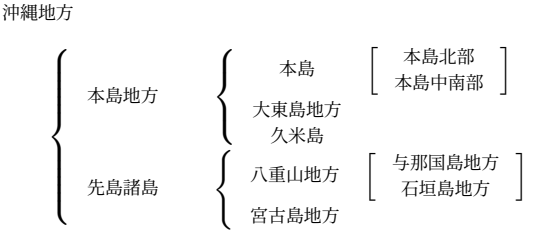


図10 沖縄地方の観測所の地域による構造化
Fig. 10 District structure for Okinawa observatory points.

機構を実現する最大の利点の1つである．次に，概要データと地域別のデータを比較し，概要データにおいて天気の分布確率が50%を超えるデータは，地域別の天気データから削除する．この処理は，DOMを使用して入力データ構造を走査することによって実現した．以上から，その地方において分布確率が過半数を超える天気について，地域別の説明を省く処理が行われる．

5. マイクロプランニング

マイクロプランニングは，ドメインに依存した処理である文書プランニングや言語知識に依存した処理である表層実現のいずれにも純粋に属さない処理を行う．具体的には，表層語彙の選択（Lexicalisation），1文単位のデータの集約（Aggregation），参照表現の生成（Referring Expressions Generation）の3つの処理が行われる．本研究では，このうち最初の2つの処理に

ついて，文書プランにおけるタグ構造の順序に基づく要素の収集およびタグ変換処理として実現した．XML に対するこの処理の実現方法として，SAX（Simple API for XML）を利用した．

5.1 語彙の選択の基準

語彙の選択に際しては，気象庁が天気予報などで用いる予報用語☆を参照したうえで，2 章の「時間」の表現と地域内の天気の分布確率に基づく修飾表現を実現する．時間については，天気事象の Time 属性の値を，時間要素として**図 11** のように置き換えた．

地域内の天気の修飾表現については，分布確率に基づいて**図 12** のような表現に置き換えた．

また，表層実現過程を意識して，天気と上記の修飾表現と観測所の集合を属性とする句仕様（Phrase）タグを導入した．このタグは，ほぼ 1 つの節に対応する．地域タグは Place タグに置き換えて，地域名を Name 属性の値とした．また，天気事象タグの属性として時制（Tense）属性を与えて，午前 3 時の天気事象についてのみ Present（現在形）を，その他の要素には Future（未来形）を値として与えた．

5.2 天気を単位とした 1 文単位の集約

生データから自然言語を実現するうえで，最も重要な処理は 1 文に対応するデータの範囲を決定することである．この目的のために，集約処理では，データの属性に基づきデータ集合の縮約を試みる．本研究の集約処理は 3 つの段階から構成される．

(1) 「午前 3 時」「今日の昼」「今日の夜」「明日の昼」の天気事象内の地域全体の天気を参照した部分地域の句仕様の収集
(2) 「今日の昼」「今日の夜」の天気事象中の天気を参照した「6 時，9 時，12 時」および「15 時，18 時，21 時」の句仕様の省略および変形
(3) 複数の句仕様の修辞接続関係に基づく 1 文単位への合成

最初の処理は，地域全体の天気で修飾表現が「ところにより」「局地的に」に対応する天気について，各部分地域中で対応する天気属性を持つ地域名や観測所名を収集して新たな句仕様を形成する．どの地域名または観測所の名前を含むかの判断は修飾属性の値に基づいて行う．また，値に応じて「〜の大半」「〜の多く」などの表現を地域名に付加している．北海道地方については，2 つの入力から句仕様を形成したうえで，短い記述を選択する処理が行われる．なお，収集した地域名は並列構造として実現しているが，「と」と「および」を組み合わせることで単調な言い回しを避けている．また，英文に関しては，“, ” で接続した後，最後の個所のみ “, and” で接続する．さらに，「局地的に」を属性値としていた句仕様は新たに Aux（助動詞）属性を与えることで表層実現過程で文末表現を変化させる．

(1) 3 時 ⇒ 午前 3 時の
(2) 6 時 ⇒ 朝のうちは
(3) 9 時 ⇒ 昼前は
(4) 12 時 ⇒ 昼過ぎは
(5) 15 時 ⇒ 夕方からは
(6) 18 時 ⇒ 宵の内は
(7) 21 時 ⇒ 夜遅くなると
(8) Today-Day ⇒ 今日は
(9) Today-Night ⇒ 夜には
(10) Tomorrow-Day ⇒ 明日は

図 11 時間要素の表層語彙表現
Fig. 11 Time-related phrases.

(1) 100% ⇒ 各地とも
(2) 80%〜100% ⇒ 全般に
(3) 50%〜80% ⇒ おおむね
(4) 20%〜50% ⇒ ところにより
(5) 20%より下 ⇒ 局地的に

図 12 修飾要素の表層語彙表現
Fig. 12 Modifying phrases.

「午前 3 時の」天気事象がこの処理に含まれるのは，気象庁が午前 5 時に発表する天気予報にならい実現すると，この天気事象要素は現在の天気として扱われ，一塊の談話単位として表層的に実現されることによる．

次の処理では，「今日の昼」の天気事象中の句仕様の天気属性の値から，6 時，9 時，12 時中の対応する天気属性を持つ句仕様を削除し，同様に，「今日の夜」の天気事象中の句仕様の天気属性の値から，15 時，18 時，21 時中の対応する天気属性を持つ句仕様を削除することで，それぞれ時間帯の特徴的な天気要素だけを残す．次に，時間帯を横断して同じ地域で同じ天気が続く場合には，「夕方から宵の内は」のように時間範囲を表す句仕様を形成する．結果は，6 時，9 時，12 時と 15 時，18 時，21 時の句仕様をそれぞれを 1 つの天気事象要素としてまとめる．

最後に，天気間の類似性を利用して修辞関係[7] に基づいた 1 文の構成を試みる．1 文の構成方法は，連続する句仕様のうち，先頭から偶数番目の句仕様に修辞（Rhetoric）属性を追加することで行う．修辞関係は，「晴れ」「快晴」の天気と，それ以外の天気をグループ化し，同じ天気グループに属する場合は「順接」，そうでない場合は「逆接」の関係を値とする．

☆ http://www.kishou.go.jp/know/yohouyougo/index.html

本章を終えるにあたり，英文を生成する際の異なる処理について説明する．英文の生成対象として,「各地に」「全般に」などは “in all districts” のような前置詞句を使用したのに対して,「おおむね」「ところにより」などは，“generally，partly” などの副詞を使用した．それぞれの表現は表層実現過程において生成される語順が異なることから，要素タグを区別した．

出力結果であるテキスト仕様を付録 A.3 に示す．

6. 表層実現

表層実現は XSLT（eXtensible Stylesheet Language Transformations）を利用して実現した．XSLT は，タグのパターンを発火点としたテンプレートルールの記述が基本となっており，本研究では，時制情報（現在形，未来形）を談話を単位とするテンプレートのパラメータから文を単位とするテンプレートのパラメータへと受け渡すことにより,「晴」「曇」「雨」などの天気表現の活用を実現した．この実現には $xsl:param$ コマンドを使用した．また，文末属性 Aux を利用して「ところがあります」に基づいた天気表現の活用を実現した．条件分岐には，$xsl:choose$ コマンドを使用した．

また，英文の生成については，大文字と小文字の区別などの正書法の処理が重要となるが，このために天気事象を談話単位とする中での句仕様の順序の情報を利用する．すなわち，各談話単位の先頭にはその時間を表す言葉を大文字で導入し，それ以外の文章は，修辞関係で接続されない奇数番目の文章と 6，9，12 時および 15，18，21 時を集約した談話単位の先頭の文章を大文字で開始する．また，並列構造を実現する際に主語を共通化するために，天気の形式主語や共通の時制句を省略することも同様に談話構造内の位置情報を利用して実現した．さらに，ドイツ語の生成においては，主節においては定動詞が 2 番目に来る語順の制約規則があり，主節か従節かといった文の構造に依存して語順が決まる．この制約に関しても，句仕様の談話単位内の位置の順序を計算することで語彙の選択を実現した．この情報は，XSLT における position() 関数と last() 関数ならびに mod などの演算子を利用することで実現した．

このほか，同じ語彙を何度も生成する場合には，出現回数を数えることにより言い換え処理を実現している．実現例としては，沖縄地方の “各地とも” の表現に対応する英文表現 “throughout the archiperago” は，カウンタを使用して 2 回目以降の出現は “throughout” にした．また，出現頻度の偶数/奇数番目の出現順序の判定に基づく “でしょう/見込みです” の言い換え（図 8 の構成パターン (viii)）も同様に実現した．

なお，XSLT 自身は Xalan[☆]を使用して実現している．また，言い換えのための出現回数の保存のために，Xalan の Java 拡張関数を使用した．出力結果である予報文章を付録 A.4 に示す．

7. 評価

本研究についての評価は，定性的な評価を 3 段階のそれぞれのモジュールの実現技術について考察した後で，平成 13 年 9 月 20 日から同 10 月 30 日までの 40 日間のデータに基づいて出力について定量的な評価を行う．

7.1 文書プランナについての評価

(1) 文書構成規則の実現容易性

生成文書の記載順序については，XML-DB に格納されたデータを抽出して DOM を使用して構成するというアプローチを採用することにより，入力文書の構造に依存せずに談話構造に基づいた内容の再構成が容易に実現できる．本研究のような，天気予報の場所と時間の順序に基づいた再構成以外に，他ドメインへの応用として，経済時系列データベース NEEDS（Nikkei Economic Electric Databank System）[☆☆]からの内閣府発行月例経済報告の生成[☆☆☆]に関して応用を試みた（文献 11) の付録.A 参照）．月例経済報告生成の特徴としては，データに応じて発表月の 2 カ月から 3 カ月前までのデータを使用して計算を行い，文書を構成する点があげられる．この場合も，本研究同様に，抽出した結果を談話順序に沿って格納するために DOM を使用することが，構造の理解のしやすさ，保守性などの点で有効である．

(2) 内容決定における重複性排除

内容の決定に関しては，4.2 節において紹介した 2 段階内容決定[10] の実現への有効性があげられる．まず，XPath のような XML-DB 特有の階層構造に応じた検索言語を使用することで，階層全体にわたるデータと階層の一部に対応するデータ集合を抽出する．次に，DOM 木を走査することにより，双方の重複を排除することができる．

☆ http://xml.apache.org/xalan-j/index.html
☆☆ http://www.nqi.co.jp/needs/n_top.html
☆☆☆ http://www5.cao.go.jp/keizai3/getsurei.html

表 1 実際の天気予報との一致率
Table 1 Agreement rates between weather forecasts on the World Wide Web and output texts.

一致の状態	北海道地方	沖縄地方
天気と地域名の対応が予報と作成文書との間で 3 カ所以上一致	$\frac{9}{41}$	$\frac{9}{41}$
天気と地域名の対応が予報と作成文書との間で包含的に一致	$\frac{11}{41}$	$\frac{13}{41}$
天気と地域名の対応が予報と作成文書との間で部分的に一致	$\frac{16}{41}$	$\frac{13}{41}$
天気と地域名の対応が予報と作成文書との間で相互補完に一致	$\frac{2}{41}$	$\frac{2}{41}$
天気と地域名の対応が予報と作成文書との間で不一致	$\frac{3}{41}$	$\frac{4}{41}$

7.2 マイクロプランナについての評価

(1) タスク別モジュール化に基づく保守性の向上
マイクロプランナは，SAX を使用して，語彙の選択および集約の 3 つのタスク（5.2 節参照）ごとにパイプラインにつなぐ ことで実現した．このため，サブタスクをモジュール化することにより保守性が向上した．また，3 段階パイプラインアーキテクチャの XML による実現において最もむずかしい点であると思われる，XML の構造を横断して内容と構造を大きく組み替える処理に対する，SAX を使用した実現の有効性を示した．

(2) 複数言語生成への応用
マイクロプランナは，複数言語（英語，フランス語，ドイツ語，日本語）の生成に対してそれぞれ共通のモジュール化手法を適用して実現した．各言語に応じた語彙の選択と 1 文単位の集約処理（英語であれば複数の名詞要素を “, ” に続けて最後に “, and” で構成，日本語の名詞要素は “と” や “および” を組み合わせて実現）を実現することで，SAX を使用したタスク別モジュール化が個々の言語に依存せず有効となることを示した．

7.3 表層実現器についての評価

(1) 音声媒体への対応
表層実現器を XSLT を使用して実現することにより最大の利点は複数媒体への対応である．近年の World Wide Web 上では携帯電話，PDA，テレビなどの複数媒体からのアクセスを行うユーザ数が増加し，また，視覚障害者向けの音声ブラウザの開発も進んできた．本システムの表層実現器は，XSLT の変換ファイルを複数用意し，XHTML フォーマットと VoiceXML フォーマット☆で文書を出力し，IBM Websphere Voiceserver SDK☆☆を使用することで，英語とドイツ語について音声による出力を実現した．これにより，XSLT により表層実現モジュールを実現することで複数メディアに対応できるという本手法の利点の 1 つが示された．

(2) 談話単位内の句の位置情報に基づく生成語彙の選択
生成のための語彙選択に関しては，XML 形式の入力を使用することで，段落内の位置情報を使用した大文字/小文字の区別などの語彙の選択を実現した．また，繰返し的な表現を避けるための言い換え処理を Xalan の Java 拡張関数を使用して実現した．

7.4 出力文についての評価

出力文について定量的に評価する．まず，平成 13 年 9 月 20 日から 10 月 30 日までの 40 日間の間のデータを使用して，World Wide Web 上で収集した天気予報との一致率を表す**表 1** に示す．

一致率が低くなる原因はいくつかあげられる．1 つは，台風などのときには同じ「雨」の天気についても，各地域ごとに注意とあわせて細かく天気を表現する傾向がある．また，そのような状況では，晴れの地域は「その他の地方」という言葉でまとめられる傾向がある．このように，表現されている天候の地域が生成した天気予報と実際の天気予報との間で補完関係になる事例は多く，これは本研究の数値情報に加えて焦点などのパラメータが実際の生成には必要となることを意味している．また，図 8 の規則 (v) は,「今日と明日」および「昼と夜」の間でも強く影響している．そのほかには，悪天候について詳しく紹介する傾向があり，各天候の情報を等しく扱っていないことがあげられる．「台風」などの気象原因パラメータに応じて談話構造を切りかえて生成することによる一致率の向上は入力データの不足から今回は行わなかったが，将来取り組むべき課題の 1 つとして検討している．

8. 関 連 研 究

まず，文献 12)〜14) を引用し，機械翻訳研究における位置付けについて説明する．

☆ http://www.voicexml.org
☆☆ http://www-3.ibm.com/software/speech/enterprise/ep_11.html

本研究のシステムは，数値データを入力として，日本語，英語のほかに，フランス語とドイツ語の天気予報を生成する．本研究の翻訳は，中間言語方式に基づいていると考えることができ，そのフォーマットは言語独立な XML 形式でタグ付けされた文書プランとみなすことができる．文献 13) にあるように，完全な意味解析は難しく，AI 完全な問題である．そこで，本研究では，言語独立な数値データに基づいて構成した XML 形式の文書プランを中間言語とすることを提案する．

さらに，生成手法に関しては細かい構文処理を行う代わりに部分文を単位として生成するアプローチを採用している．この方針は文の一部を翻訳単位としてデータベースに格納して対応付ける一連の手法（文献 14) など）と一部共通するが，本研究では，より基本的なデータに基づく中間表現から文を構成することにより，文の一部が持つ意味をデータとして明示化する．

また，生成処理に関しては，ドメイン依存の部分言語（sub-language）アプローチを採用している．この方針は，風速などの数値データから英語とフランス語の天気予報を生成する FoG☆[1),3)] と共通のものであり，また，文献 13) の後書きなどにもあるように，ドメインごとに複数の翻訳エンジンを採用することは，実際に翻訳を行う立場から捉えた場合にも有効であることが分かる．

FoG と比較した本研究の最大の利点は実現の見通しのよさにある．また，本研究のシステムは経済ドメインへの応用にも成功した（文献 11) の付録.A 参照）．他のドメインへの応用において生じた問題点としては，入力データからの推論が何段階もの複雑なステップを踏むため表層語彙にむずびつけることがむずかしい文章（例：景気が良い・悪いを数値データから総合的に判断する）に関しては実現が難しい点があげられる．何らかの統計的な解析手段と組み合わせたうえでこの問題が解決できるかどうかについては，気象解析技術などもあわせて将来検討する問題の 1 つと考えられる．

以上から，本研究の有効性について総括する．本研究の生成手法は，文書の談話構造がある程度定型化でき，入力データが必ずしもその談話構造に対応しない場合に，データを再構成して文章を生成する応用に特に有効である．逆に，談話構造が固定化できない対話生成処理や，言語独立な中間表現が設定できない応用は本研究の対象外となる．ただし，談話構造を横断した柔軟な生成処理が行えないという意味では決してなく，談話構造を横断した言い換え処理を実現した．

XML に関連した文章生成研究については，文献 15) がある．本研究と比較すると，文献 15) は，3 段階パイプラインのすべての段階で共通の技術の応用を試みたうえで，いくつかの手法を比較しているのに対して，3 段階のそれぞれに有効と見なせる技術を割り当てたうえで応用を実現しているという点で本研究に利点がある．すなわち，文書構造に基づいた内容の決定には，DOM 木を走査することにより，概要に基づいた内容の決定が可能となる．文書プランからテキスト仕様を構成するような大きな要素の組換えには SAX が適している．XSLT は談話構造に基づく言い換えを容易に実現でき，音声媒体の出力への対応も実現できる．

さらに，Java が限られたユーザしか使用できないことから DOM を使用しないという文献 15) の主張は，本来の生成アーキテクチャへの実現の有効性という論点から外れているように思われ，また現在，DOM は，libgdome-ruby を使用することにより Ruby のようなスクリプト言語からも使用可能である．以上は，文献 15) の著者と直接議論したうえでの主張である．このほかにも対話処理への応用があるが，本研究とは応用目的が若干異なり，紙数の関係からも，これ以上の比較は行わない．

最後に，参考にした研究についてまとめる．文献 3) は，WeatherReporter と呼ばれる天気予報生成 のプロトタイプシステムの作成過程について説明しており，本研究のシステムの実現にあたって参考にした．XML 形式からのデータ抽出については，文献 16) を参考にした．

9. おわりに

本研究はそもそも，生データからの報告文書の生成[1)] という一般的なテーマを実現するにあたっては，意図の構造と談話の構造が対応しているという指摘[2)] に基づき，生データからの検索自体を，自然言語の談話の構造に基づいた検索に置き換えることに本質があるという着眼を出発点としている．

データの構造化にあたっては，XML を利用するのが自然であり，自然言語生成の技術として標準的な 3 段階パイプラインアーキテクチャを採用したうえで，DOM，SAX，XSLT などの標準的な XML 技術を駆使して実現することにより，計算機上での実現可能性を示すとともに，大文字と小文字の区別や助動詞の活用など言語の特徴ならびに談話の構造に依存した生成処理を容易に実現できることを証明した．また，複数

☆ the **Fo**recast **G**enerator component of Environment Canada's Forecast Production Assistant workstation

の言語を対象とすることによって，ドメインを対象とした伝達意図に基づく談話の構造の選択と，個別の出力言語に依存する談話中の個々の文の構造および表層語彙の選択との境界線を明確に示した．

謝辞 本研究の入力データは，（財団法人）気象業務支援センター提供の地上気象観測原簿データ（気象庁年報）2000 年度版および 2001 年度 9 月と 10 月の月報の CD-ROM を使用した．また，本研究で使用した天気予報の例文は気象庁発表によるものであり，（株）CRC ソリューションズ社の Web サイト☆上からデータを取得した．使用のご許可をいただいた CRC ソリューションズの冨洋一様に感謝します．

XML データベースとしては，（株）メディアフュージョン社の Yggdrasill 評価版 Version 1.0 を使用した．使用に関してご尽力された方々に深く感謝します．

参 考 文 献

1) Kittredge, R.I. and Polguere, A.: The Generation of Reports from Databases, *Handbook of Natural Language Processing*, Dale, R., Moisl, H. and Somers, H. (Eds.), chapter 11, pp.261–304, Marcel Dekker (2000).
2) Grosz, B.J. and Sidner, C.L.: Attention, Intentions, and the Structure of Discourse, *Computational Linguistics*, Vol.12, No.3, pp.175–204 (1986).
3) Reiter, E. and Dale, R.: *Building Natural Language Generation Systems*, Cambridge University Press (2000).
4) Seki, Y. and Harada, K.: XML Transformation-based three-stage pipelined Natural Language Generation System, *the 6th Natural Language Processing Pacific Rim Symposium (NLPRS'01) Exhibition and Demonstration*, Tokyo, Japan (2001).
5) 関 洋平，原田賢一：天気予報を対象とした XML-DB からの動的な文章作成，2002 年情報学シンポジウム，日本学術会議講堂 (2002).
6) Edmonds, P.: *Semantic Representations of Near-Synonyms for Automatic Lexical Choice*, Ph.D. Thesis, Department of Computer Science, University of Toronto (1999).
7) Mann, W.C.: Rhetorical Structure Theory: Description and Construction of Text Structures, *Natural Language Generation: Recent Advances in Artificial Intelligence, Psychology, and Linguistics*, Kempen, G. (Ed.), chapter 7, pp.85–95, Kluwer Academic Publishers, Dordrecht (1987).
8) McDonald, D.D.: Natural Language Generation, *Handbook of Natural Language Processing*, Dale, R., Moisl, H. and Somers, H. (Eds.), chapter 7, pp.147–179, Marcel Dekker (2000).
9) メディアフュージョン XML ラボ：XML データベースによる Web アプリケーション開発，ソフトバンクパブリッシング (2001).
10) Sripada, S.G., Reiter, E., Hunter, J. and Yu, J.: A Two-stage Model for Content Determination, *Proc. 8th European Workshop on Natural Language Generation associated to ACL 39th Ann. Meeting and 10th Conf. of the European Chapter*, Toulouse, France, pp.3–10 (2001).
11) 関 洋平，原田賢一，野村直之：Ruby による複数資源要約システムの実現，情報処理学会第 66 回情報学基礎研究会第 32 回デジタルドキュメント研究会，国立情報学研究所 (2002).
12) Somers, H.: Machine Translation, *Handbook of Natural Language Processing*, Dale, R., Moisl, H. and Somers, H. (Eds.), chapter 13, pp.329–346, Marcel Dekker (2000).
13) Bond, F.: Toward a Science of Machine Translation, *Proc. MT Roadmap Workshop at TMI-2002*, Keihanna, Japan (2002).
14) Forcada, M.: Using multilingual content on the web to build fast finite-state direct translation systems, *Proc. MT Roadmap Workshop at TMI-2002*, Keihanna, Japan (2002).
15) Wilcock, G.: Pipelines, Templates and Transformations: XML for Natural Language Generation, *Proc. 1st NLP and XML Workshop*, Tokyo, Japan (2001).
16) 戌亥 稔，田中 聡，田中行広：実践 XML データベース構築，オーム社 (2001).

付 録

A.1 入力 DTD 形式

A.1.1 北海道地方の入力データに対する DTD 形式

```
<?xml version="1. 0" encoding="Shift_JIS" ?>
<!ELEMENT Hokkaido ( InlandAreas,theSeaofOkhotskSide,
theSeaofJapanSide,thePacificOceanSide ) >
<!ATTLIST Hokkaido Date NMTOKEN #REQUIRED >
<!ATTLIST Hokkaido Year NMTOKEN #REQUIRED >
<!ATTLIST Hokkaido Month NMTOKEN #REQUIRED >

<!ELEMENT InlandAreas ( Observatory+ ) >
<!ELEMENT theSeaofOkhotskSide ( Observatory+ ) >

<!ELEMENT thePacificOceanSide ( thePacificOceanEastSide,
thePacificOceanWestSide ) >
<!ELEMENT thePacificOceanEastSide ( Observatory+ ) >
<!ELEMENT thePacificOceanWestSide ( Observatory+ ) >

<!ELEMENT theSeaofJapanSide ( theSeaofJapanNorthSide,
```

☆ http://weather.crc.co.jp

```
theSeaofJapanWestSide ) >
<!ELEMENT theSeaofJapanNorthSide ( Observatory+ ) >
<!ELEMENT theSeaofJapanWestSide ( Observatory+ ) >

<!ELEMENT Observatory ( Time+, Day ) >
<!ATTLIST Observatory Name ID #REQUIRED >
<!ATTLIST Observatory Number NMTOKEN #REQUIRED >
<!ELEMENT Day ( Pressure,AvgTemp,MaxTemp,MinTemp,Humid,
AvgWindSpeed,MaxWindSpeed,WindDirection,Weather+ ) >
<!ELEMENT Time ( Pressure,Temp,Humid,WindDirection,WindSpeed,
Weather,RainFall ) >
<!ATTLIST Time Attribute ( 3 | 6 | 9 | 12 | 15 | 18 | 21 )
#REQUIRED >
<!ELEMENT Weather ( #PCDATA | Element )* >
<!ATTLIST Weather Attribute ( Night | Day ) #IMPLIED >
<!ELEMENT Element ( #PCDATA ) >
<!ATTLIST Element Number ( 1 | 2 | 3 ) #REQUIRED >
<!ATTLIST Element Modify ( While | After | AfterWhile |
AfterWithIntermittent | With | WithIntermittent )
#IMPLIED >

<!ELEMENT Pressure ( #PCDATA ) >
<!ELEMENT AvgTemp ( #PCDATA ) >
...
```

A.1.2 北海道地方の入力データに対する DTD 形式 (2)

```
<?xml version="1. 0" encoding="Shift_JIS" ?>
<!ELEMENT Hokkaido ( North,East,SouthWest ) >
<!ELEMENT North ( theSeaofJapanNorthSide,Observatory+ ) >
<!ELEMENT East ( thePacificOceanEastSide,Observatory+ ) >
<!ELEMENT SouthWest ( theSeaofJapanWestSide,
thePacificOceanWestSide,Observatory+ ) >
...
```

A.1.3 沖縄地方の入力データに対する DTD 形式

```
<?xml version="1. 0" encoding="Shift_JIS" ?>
<!ELEMENT Okinawa ( HontoDistrict,SakishimaIslands ) >
<!ELEMENT HontoDistrict ( HontoNorth,HontoSouth,
DaitoIslandDistrict,Observatory ) >
<!ELEMENT HontoNorth ( Observatory ) >
<!ELEMENT HontoSouth ( Observatory ) >
<!ELEMENT DaitoIslandDistrict ( Observatory ) >
<!ELEMENT SakishimaIslands ( YaeyamaDistrict,
MiyakoIslandDistrict ) >
<!ELEMENT MiyakoIslandDistrict ( Observatory ) >
<!ELEMENT YaeyamaDistrict ( YonaguniIslandDistrict,
IshigakiIslandDistrict ) >
<!ELEMENT YonaguniIslandDistrict ( Observatory ) >
<!ELEMENT IshigakiIslandDistrict ( Observatory ) >
...
```

A.2 沖縄地方 2001 年 9 月 1 日の文書プラン

```
<?xml version="1.0" encoding="Shift_JIS"?>

<Set Object="Weather" Year="2001" Month="9" Day="1">
  <WeatherEvent Time="3">
    <Okinawa ObservatoryNumber="4">
      <Weather>
        <Fair Distribution="50%" />
        <Cloudy Distribution="50%" />
      </Weather>
      <HontoDistrict ObservatoryNumber="2">
        <Weather />
      </HontoDistrict>
      <SakishimaIslands ObservatoryNumber="2">
        <Weather />
      </SakishimaIslands>
    </Okinawa>
  </WeatherEvent>
  <WeatherEvent Time="Today-Day">
    <Okinawa ObservatoryNumber="7">
      <Weather>
        <Rain Distribution="42%" />
        <Fair Distribution="28%" />
        <Cloudy Distribution="28%" />
      </Weather>
      <HontoDistrict ObservatoryNumber="4">
        <Weather>
          <Rain Distribution="75%">
            <Name>久米島</Name>
            <Name>本島</Name>
          </Rain>
          <Fair Distribution="25%">
            <Name>大東島地方</Name>
          </Fair>
        </Weather>
      </HontoDistrict>
      <SakishimaIslands ObservatoryNumber="3">
        <Weather>
          <Cloudy Distribution="66%">
            <Name>八重山地方</Name>
          </Cloudy>
          <Fair Distribution="33%">
            <Name>宮古島地方</Name>
          </Fair>
        </Weather>
      </SakishimaIslands>
    </Okinawa>
  </WeatherEvent>
  <WeatherEvent Time="6">
    <Okinawa ObservatoryNumber="4">
      <Weather>
        <Rain Distribution="50%" />
        <Fair Distribution="25%" />
        <Cloudy Distribution="25%" />
      </Weather>
      <HontoDistrict ObservatoryNumber="2">
        <Weather>
          <Cloudy Distribution="50%">
            <Name>本島中南部</Name>
          </Cloudy>
        </Weather>
      </HontoDistrict>
      <SakishimaIslands ObservatoryNumber="2">
        <Weather>
          <Fair Distribution="50%">
            <Name>石垣島地方</Name>
          </Fair>
        </Weather>
      </SakishimaIslands>
    </Okinawa>
  </WeatherEvent>
  <WeatherEvent Time="9">
    <Okinawa ObservatoryNumber="7">
```

```
    <Weather>
      <Rain Distribution="42%" />
      <Cloudy Distribution="42%" />
      <Fair Distribution="14%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="4">
      <Weather>
        <Rain Distribution="75%">
          <Name>久米島</Name>
          <Name>本島</Name>
        </Rain>
        <Cloudy Distribution="25%">
          <Name>大東島地方</Name>
        </Cloudy>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="3">
      <Weather>
        <Cloudy Distribution="66%">
          <Name>与那国島地方</Name>
          <Name>宮古島地方</Name>
        </Cloudy>
        <Fair Distribution="33%">
          <Name>石垣島地方</Name>
        </Fair>
      </Weather>
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="12">
  <Okinawa ObservatoryNumber="4">
    <Weather>
      <Fair Distribution="75%" />
      <Rain Distribution="25%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="2">
      <Weather>
        <Rain Distribution="50%">
          <Name>本島中南部</Name>
        </Rain>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="2">
      <Weather />
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="Today-Night">
  <Okinawa ObservatoryNumber="5">
    <Weather>
      <Cloudy Distribution="80%" />
      <Fair Distribution="20%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="2">
      <Weather>
        <Fair Distribution="50%">
          <Name>大東島地方</Name>
        </Fair>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="3">
      <Weather />
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="15">
  <Okinawa ObservatoryNumber="7">
    <Weather>
      <Cloudy Distribution="57%" />
      <Fair Distribution="28%" />
      <Thunder Distribution="14%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="4">
      <Weather>
        <Fair Distribution="25%">
          <Name>大東島地方</Name>
        </Fair>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="3">
      <Weather>
        <Thunder Distribution="33%">
          <Name>与那国島地方</Name>
        </Thunder>
        <Fair Distribution="33%">
          <Name>宮古島地方</Name>
        </Fair>
      </Weather>
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="18">
  <Okinawa ObservatoryNumber="4">
    <Weather>
      <Cloudy Distribution="50%" />
      <Rain Distribution="25%" />
      <Fair Distribution="25%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="2">
      <Weather>
        <Rain Distribution="50%">
          <Name>本島中南部</Name>
        </Rain>
        <Fair Distribution="50%">
          <Name>大東島地方</Name>
        </Fair>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="2">
      <Weather />
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="21">
  <Okinawa ObservatoryNumber="5">
    <Weather>
      <Cloudy Distribution="60%" />
      <Thunder Distribution="20%" />
      <Fair Distribution="20%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="2">
      <Weather>
        <Thunder Distribution="50%">
          <Name>本島中南部</Name>
```

```
        </Thunder>
        <Fair Distribution="50%">
          <Name>大東島地方</Name>
        </Fair>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="3">
      <Weather />
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
<WeatherEvent Time="Tomorrow-Day">
  <Okinawa ObservatoryNumber="7">
    <Weather>
      <Cloudy Distribution="57%" />
      <HeavyRain Distribution="14%" />
      <Rain Distribution="14%" />
      <Fair Distribution="14%" />
    </Weather>
    <HontoDistrict ObservatoryNumber="4">
      <Weather>
        <Rain Distribution="25%">
          <Name>久米島</Name>
        </Rain>
        <Fair Distribution="25%">
          <Name>大東島地方</Name>
        </Fair>
      </Weather>
    </HontoDistrict>
    <SakishimaIslands ObservatoryNumber="3">
      <Weather>
        <HeavyRain Distribution="33%">
          <Name>与那国島地方</Name>
        </HeavyRain>
      </Weather>
    </SakishimaIslands>
  </Okinawa>
</WeatherEvent>
</Set>
```

A.3 沖縄地方 2001 年 9 月 1 日のテキスト 仕様（日本語用）

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<Set Object="Weather" Year="2001" Month="9" Day="1">
  <WeatherEvent Tense="Present">
    <Time>午前 3 時の</Time>
    <Place Name="沖縄地方の天気">
      <Weather>
        <Phrase modify="おおむね" weather="晴" />
        <Phrase modify="おおむね" Rhetoric="並列" weather="
曇" />
      </Weather>
    </Place>
  </WeatherEvent>
  <WeatherEvent Tense="Future">
    <Time>今日は</Time>
    <Place Name="沖縄地方の天気">
      <Weather>
        <Phrase aux="ところが多い" weather="曇" locations="
先島諸島の多く" />
        <Phrase aux="ところが多い" weather="晴" Rhetoric="
逆接" locations="大東島地方と宮古島地方" />
        <Phrase aux="ところが多い" weather="雨" locations="
本島地方の大半" />
      </Weather>
    </Place>
  </WeatherEvent>
  <WeatherEvent Tense="Future">
    <Time>夜には</Time>
    <Place Name="沖縄地方の天気">
      <Weather>
        <Phrase modify="全般に" weather="曇" />
        <Phrase aux="ところが多い" weather="晴" Rhetoric="
逆接" locations="本島地方の大半" />
      </Weather>
    </Place>
  </WeatherEvent>
  <WeatherEvent Tense="Future">
    <Weather>
      <Phrase modify="夕方からは" weather="雷" locations="
与那国島地方" />
      <Phrase modify="夜遅くなると" weather="雷" Rhetoric="
順接" locations="本島中南部" />
      <Phrase modify="宵の内は" weather="雨" locations="
本島中南部" />
    </Weather>
  </WeatherEvent>
  <WeatherEvent Tense="Future">
    <Time>明日は</Time>
    <Place Name="沖縄地方の天気">
      <Weather>
        <Phrase modify="おおむね" weather="曇" />
        <Phrase aux="ところが多い" weather="晴" Rhetoric="
逆接" locations="大東島地方" />
        <Phrase aux="ところがある" weather="大雨" locations="
与那国島地方" />
        <Phrase aux="ところが多い" weather="雨" Rhetoric="
順接" locations="久米島" />
      </Weather>
    </Place>
  </WeatherEvent>
</Set>
```

A.4 沖縄地方 2001 年 9 月 1 日の天気予報の出力

A.4.1 日本語の出力

午前 3 時の沖縄地方の天気は全般に晴れていますか，または曇っています。

今日は先島諸島の多くは曇りとなるでしょうが，大東島地方と宮古島地方は晴れる見込みです。本島地方の大半は雨が降るでしょう。

夜には全般に曇りとなる見込みですが，本島地方の大半は晴れるでしょう。与那国島地方は夕方からは雷がある見込みです。また，本島中南部は夜遅くなると雷があるでしょう。本島中南部は宵の内は雨が降る見込みです。

明日はおおむね曇りとなるでしょうが，大東島地方は晴れる見込みです。与那国島地方は大雨となるところがあるでしょう。また，久米島は雨が降る見込みです。

A.4.2 英語の出力

The weather at three o'clock in Okinawa is generally fair and cloudy.

Today's weather will be partly cloudy over the Yaeyama District, but partly fair over the Daito Islands, and Miyako Islands. There will be partly rain on the Kume Island, and Honto.

Tonight's weather will be cloudy, but partly fair over the

Daito Islands. There will be showers on the the central South part of Honto, and thunder over the Yonaguni Islands. There will be thunder over the the central South part of Honto at night.

The outlook for tomorrow in Okinawa is generally cloudy, but partly fair over the Daito Islands. There will be partly rain on the Kume Island, and locally heavy rain over the Yonaguni Islands.

A.4.3 フランス語の出力

Le temps à trois heures dans zone de l'Okinawa est tres bien et est nuageux généralement.

D'aujourd'hui temps dans zone de l'Okinawa sera nuageux en partie dans zone de Yaeyama, mais sera parfait en partie dans Îles de Daito, et Îles de Miyako. Il sera pluvieux en partie dans Île de Kume, et Honto.

Ce soir temps dans zone de l'Okinawa sera nuageux, mais sera parfait en partie dans Îles de Daito. Il y aura de tonnerre dans Îles de Yonaguni, et il y aura de tonnerre dans la pièce de mi-sud de Honto. Il sera pluvieux dans la pièce de mi-sud de Honto en soirée.

Les perspectives pour le demain dans zone de l'Okinawa sera nuageux, mais sera parfait en partie dans Îles de Daito. Il sera pluvieux en partie dans Île de Kume, et sera forte pluie localement dans Îles de Yonaguni.

A.4.4 ドイツ語の出力

Das Wetter um drei Uhr in Okinawa ist es im Allgemeinen sonnig und wolkig.

Heute regnet es örtlich auf der Insel Kume, und dem Honto, und ist es örtlich wolkig über dem Bezirk Yaeyama. Es ist örtlich sonnig über den Inseln Daito, und den Inseln Miyako.

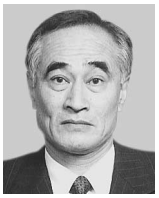
Heute abend es wird wolkig, aber örtlich sonnig über den Inseln Daito. Am Abend gibt es Duschen auf dem zentrale Südteil von Honto, und wird Donner über den Inseln Yonaguni. Es wird Donner über dem zentrale Südteil von Honto.

Morgen in Okinawa wird es im Allgemeinen wolkig, und wird örtlich Regen auf der Insel Kume. Es wird örtlich sonnig über den Inseln Daito, aber wird vereinzelt schwerer Regen über den Inseln Yonaguni.

(平成 13 年 11 月 20 日受付)
(平成 14 年 5 月 15 日採録)

関　洋平（正会員）

1971 年生．1996 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．2002 年より青山学院大学理工学部助手．同年より総合研究大学院大学に在学中．自然言語処理の研究に従事．ACL，言語処理学会各会員．



原田　賢一（正会員）

1940 年生．1966 年慶應義塾大学大学院工学研究科管理工学専攻修士課程修了．1967 年同大学工学部助手．1970 年〜1989 年同大学情報科学研究所助手，専任講師，助教授，教授．1989 年 4 月より同大学理工学部計測工学科教授．1998 年 4 月より同学部情報工学科教授．この間，1973 年〜1975 年米国メリーランド大学訪問研究員．工学博士．ソフトウェア工学，プログラミング言語およびその処理系の研究に従事．ACM，IEEE，ソフトウェア科学会会員．