専門用語のユーザに合わせた言い換え支援システムの構築
-言い換えデータベースの提案-

藤沢仁子† 神門 典子†‡ 相原 健郎†‡ 安達 淳‡
†総合研究大学院大学 情報学専攻 ‡国立情報学研究所
E-mail: †satoko@grad.nii.ac.jp, ‡ {kando, kenro.aihara, adachi} @nii.ac.jp

専門家によって書かれた文化遺産分野の解説文は，非専門家には理解が難しい専門用語を多く含む．本研究では，小学生の読解支援を目的とした，文化遺産デジタルアーカイブのコンテンツに付随される解説文の言い換えの自動化の第一段階として，用語の言い換えに着目した言い換えデータベースを提案する．また，用語・言い換えパターンの二つのテーブルから機械的に生成した言い換えが，一文中で適切な表現となっているかを判定した結果と考察を報告する．

Support System for Paraphrasing Technical Terms according to User Knowledge Level
- Lexical Paraphrasing Database based on Structured NE -

Satoko FUJISAWA†, Noriko KANDO†‡, Kenro AIHARA†‡ and Jun Adachi‡
†The Graduate University for Advanced Studies ‡National Institute of Informatics

The terminology and the descriptions about cultural heritage are too technical and difficult for nonprofessional users of the domain, especially children and people from different cultures. This paper proposes a paraphrasing database focusing on paraphrases of technical terms to assist children in reading about cultural heritage. Moreover, the paper analyzes the sentences that are automatically created and determines whether they would be proper expression or not.

1．まえがき

近年，電子化された文書を利用者や利用形態に適した形に自動編集する技術の必要性が説かれており[1][2]，ある目的を満たす表現へ文章を変換する言い換え処理の研究が多数されてきている．

筆者らが所属する研究グループでは，教育機関において学習者が文化財コンテンツを利用して知識や多様な知見を発見することを支援するソフトウェアの開発を目指したCEAXプロジェクト[3]を推進している．ここで問題となるのは，既存の文化財の解説記述には専門用語が多数使われているため，記述内容を理解するには専門知識を必要とすることである．そこで，1つのオブジェクトの説明にも「受け手」に応じた多様性を考慮した記述の必要性に着目し，従来の専門家向けの記述に加え，小学生や非専門家にも理解し易い平易な表現など複数のメタデータや記述，関連するコンテンツを柔軟に管理できる仕組みを提案している[4]そして，提案システムのプロトタイプを用いた西東京市立田無小学校 6 年生 89 名を対象とした研究授業を実施したところ，受け手にあわせて記述を作成することで理解を支援できることが示唆された．

しかし，現在，子ども向けの記述がされているものは大変限られており，全てを人手で作成するのは困難である．そこで本研究では，文化財に関する解説文を子ども向けに言い換えることを支援するシステムの構築を目指した，言い換えデータベースを提案する．

2．専門用語の言い換えの分析

既存研究のなかで，読解支援を目的とした言い換えの研究の例を見てみると，小学生を対象とした榎本ら[5]，福祉的利用を目的とした乾[6]，日本語学習者を対象とした佐藤ら[7]の研究などが上げられる．

言い換え研究の中での位置づけとして，本研究は，文化遺産分野の記述を子どもの読解支援のために易しく解りやすい文章に，漢字から文章全体にわたり，簡潔な表現の中で言い換えることを目指している．特徴としては，漢字かな変換だけでなく語彙・構文・文章全体を考慮した子どものため（教育目的）の言い換え研究である点，量的複雑さは難易度のひとつの要素ととらえ，説明的になることで文章が冗長になることを避け簡潔な文章で言い換えることを目指す点などが挙げられる．

筆者らは今までに，40組の大人向け・子ども向けの文化遺産分野の解説文の比較により表現の使い分けの分析をした結果，13個の言い換えパターンを見出した［8］［表1］．

言語レベル	言い換えパターン
文章全体の意味	(P-1)補完文の追加 (P-2)詳細な記述の削除
談話的要素	(D-1)指示語・代名詞の内容の補完
形態・構文	(S-1)連体節の主節化
用語	専門用語を残した言い換え (L-1)カテゴリの補足 (L-2)用途・目的の補足 (L-3)形状の補足 (L-4)位置の補足 易しい用語への言い換え (L-5)一般名詞への言い換え (L-6)身近なものへの言い換え (L-7)一般名詞に位置/場所を補足 (L-8)小学国語辞典の語釈文と同等表現への言い換え
漢字	(K-1)読み仮名の付与

表１　検出された言い換えパターン

そのなかで、用語の言い換え（Lexical Paraphrase）で特に重要を思われるものを以下に列挙する．

専門用語を残した言い換え
(L-1) カテゴリの補足
「＜専門用語＞とよばれる＜カテゴリ＞」
大人：挂甲（けいこう）に身を固め
子供：挂甲（けいこう）とよばれる甲（よろい）

(L-2) 用途・目的の補足
「＜用途/目的＞ための＜専門用語＞」
大人　草摺を革紐で縅してつけています。
子供　太ももを保護するための草摺（くさずり）

(L-3) 形状の補足　「＜形状＞の＜専門用語＞」
大人　鞍や障泥・輪鐙をつけています。
子供　…どろよけや輪の形のあぶみを…

易しい語への言い換え
(L-5) 一般名詞への言い換え
大人　脚結で膝のあたりを結んでいます。
子供　膝のあたりをひもで結んでいます。

(L-6) 身近なものへの言い換え
「＜身近なもの＞のような＜一般名詞＞」
大人　庇は透かしが施されています。
子供　野球帽のようなひさしには…

(L-7) 一般名詞+ 位置・場所 への言い換え
「＜位置/場所＞の＜一般名詞＞」
大人　胸繋には馬齢が付けられている。
子供　首の下の革ひもには馬鈴がつけられています。

特に注目したいのは，上記(L-1) (L-2) (L-3) (L-4) の専門用語を残した説明である．その時代や分野を語るうえで重要であり，知識として学習させたい専門用語は，完全に平易な別の表現に言い換えてしまわずに残す．そして，辞書へのリンク機能などによりその言葉を説明するのではなく，画像に付随する説明書きとしての読みやすさを考慮し，その専門用語が含まれるカテゴリや用途・目的などを文章のなかに補足することで理解を支援する点である．

今回はこの点に着目し，例えば「美豆良」を「美豆良とよばれる髪形」と自動的に言い換えるための言い換えデータベースの構築を提案する．

３．システム概要

図１は，提案する言い換え支援システムの概要である．乾［6］の指摘にもあるようにテキスト評価と言い換え生成の２つのモジュールがシステムの中心となる．

おおまかな仕組みとしては，言い換えを必要とする文章を入力すると，ある評価基準に基づいてテキストを評価し，言い換える必要のある箇所を提示する．国語辞典や小学国語辞典などの辞典，教科書や指導要領，日本語能力試験の出題基準などを，評価基準として用いることができると考えられる．

そして，指摘された用語が既に言い換えデータベースに登録されている場合は，提示された生成された言い換え候補リストの中から適当なものを選択するか，リストの表現を参考に人手で書き起こす．言い換えた結果はデータベースにフィードバックされる．また，指摘された用語が未登録の場合はリンクされている美術用語辞典を参考に人手または半自動で用語テーブルへの入力を行う．用語辞典としては，美術の専門用語を扱った美術用語辞典や日本史用語辞典，百科辞典などが使用可能である．

４．言い換えデータベース

我々が対象とする文章の特徴としては，画像に付けられた説明書きであり，用語辞典のようにすべてを言葉によって説明するのでなく，画像から判断できる情報がある点，200 字程度の短い文章という制約の中で簡潔に表現されている点が挙げられる．各解説文は独立していて，それを読む順番はユーザに委ねられているので，同じ専門用語の説明が各解説文の中で繰り返し必要となる．また，教師が WWW など外部からデータを追加する際に説明文を書換える場合には，表現が統一されることが有用であると考える．

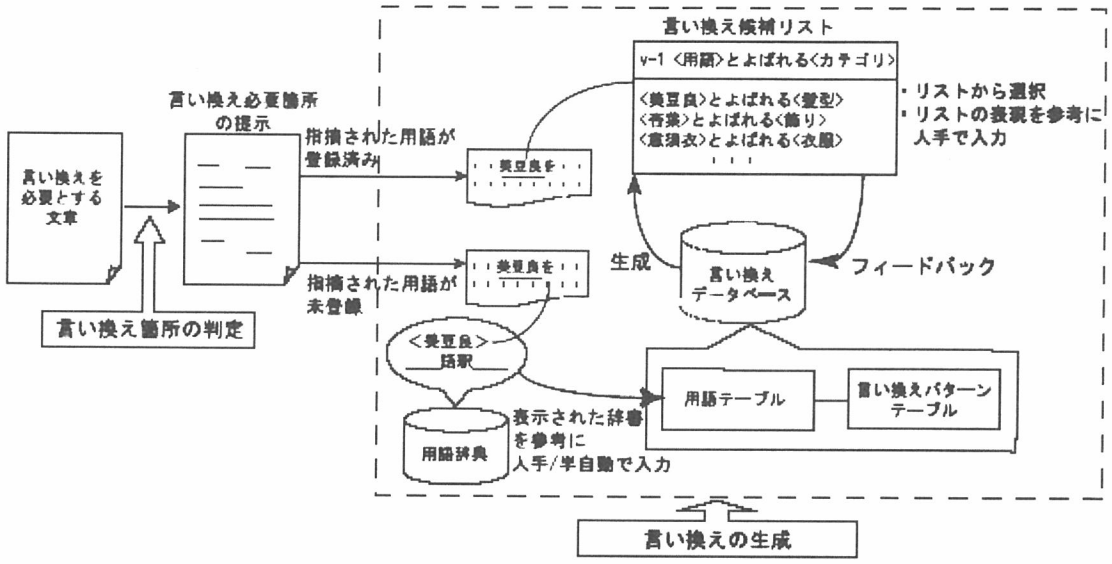

図１　システム概要図

そこで，言い換え文例をデータベース化し，ユーザに対して言い換え例を提示する為の言い換えデータベースの構築を提案する．類似のシステムとしては翻訳メモリが挙げられるが，翻訳メモリが対訳文を原文とともに保管し，入力文に対するマッチングと置き換えを行う［9］のに対し，提案システムでは用語と言い換えパターンの 2 つのテーブルから言い換え例を生成する．

用語テーブルでは，＜カテゴリ＞＜用途/目的＞＜位置/場所＞＜形状＞など各用語の属性を定義する．これらの属性は，先の人手による解説文の分析の中で見出されたものだが，AAT や CDWA にはこれらを記述するためのファセットや項目が定義されており，ドキュメンテーションにおいても重要であるといえる．しかし，カテゴリなどその用語が示すもの全てに当てはまる属性と，形状や位置のように一部の例でのみ有効な属性がある．それらをどのように管理していくかは今後の課題の一つである．一方，言い換えパターン テーブルは「＜専門用語＞とよばれる＜カテゴリ＞」「＜用途/目的＞ための＜専門用語＞」など，言い換えのテンプレートを集めたテーブルである．

言い換えを必要とする用語に対し，まず用語テーブルを検索し属性を取り出す．その属性を含む言い換えパターンを取り出し，用語テーブルのデータをテンプレートに当てはめて言い換え例を提示する．また，ユーザが実際に言い換えた結果を記録しておき，そのデータを基に頻度情報を作成し，同じ文例を提示する場合のランキングに使用する．

用　語	カテゴリ	用途/目的	位置	形状
歩揺	飾り	飾りの		スパンコールのような
短甲	甲	体を保護する	胴部分の	丈の短い
美豆良	髪型		頭部の	肩まで下がった
脚結	束帯	正装の	膝のあたりの	
・・・	・・・	・・・	・・・	

図３　用語テーブル

Ph_id	pattern
L-1	＜用語＞とよばれる＜カテゴリ＞
L-2	＜用途/目的＞ための＜用語＞
L-3	＜形状＞＜カテゴリ＞
L-4	＜位置＞＜用語＞

図４　言い換えパターンテーブル

・<スパンコールのような><歩揺>　(L-3)
・<体を保護する>ための<短甲>　(L-2)
・<美豆良>とよばれる<髪形>　(L-1)
・<膝のあたりの><脚結>　(L-4)

図５　生成された言い換えの例

5．正解判定

自動化の第一段階として，この 2 つのテーブルから機械的に生成された言い換えが，一文内で適切な表現になっているかを判定した．

テストデータとしては，「群馬の埋蔵文化財」［10］の“古墳時代”にある 96 件の解説文のから，先の分析で言い換え対象として検出さえた用語を含む 47 件の解説文を使用した．

正解判定の基準は，解説文全体ではなく一文ごとに，適正な文になっているかを，以下の５つのどれにあてはまるかで判定した．これらのカテゴリは最初から設定していたのではなく，判定作業を進めるうえで導き出された．

Y　：正解
Y2　：複合語の一部だけど正解
N　：N2，N3 以外の不正解
N2　：複合語の一部なので不正解
N3　：表現の重複により不正解

Y　：正解
言い換含む一文が，表現として適切である．

全国で発見される短甲とよばれる甲の大半はこの形式である。

N2　：複合語の一部なので不正解
複合語の一部が言い換え対象の用語となっているために，言い換えの結果，文が成立しない．

体を保護するための挂甲武人埴輪

足の甲まで鉄細長い小札の表現が

N3　：表現の重複により不正解
言い換えによって追加された説明が，元の文内の記述と内容的に重複してしまい，不適切な文となる．

使用された*甲は*、短甲とよばれる甲という形式

*筒状*の円筒形の円筒埴輪

N　：N2，N3 以外の不正解
N2，N3 に分類されないが，文として不適切なもの．

腰下には*立派な*小さな刀子が提げられ
⇒ 「立派な」と「小さな」が意味的に相反

（副葬品の種類は）農工具類と腰の刀子に限定
⇒ 画像内の位置と無関係の記述での言い換え

今回の判定は，文として適当か否かであり，不正解の場合が明確に判断できるので，人によって判定が異なる作業ではないと判断し，判定は一人で行った．

6．実験結果

以下の表は各言い換えパターンごとの判定結果である（表４～７）．なお，正解率の算出式は次の通り．

$$\frac{Y+Y2}{Y+Y2+N+N2+N3} = \text{正解率}$$

L_1	正解		不正解			合計
	Y	Y2	N	N2	N3	
animal equipment	14	4	0	0	7	25
Weapons and Ammunition	47	1	0	6	2	56
Costume	9	4	0	0	2	15
design element	2	0	0	0	0	2
equipment for costume	4	0	0	0	0	4
Tools and Equipment	9	0	0	0	0	9
others	17	5	0	1	0	23
合計	102	14	0	7	11	134
正解率						0.86

表４　実験結果 L-1 〈用語〉とよばれる〈カテゴリ〉

L_2	正解		不正解			合計
	Y	Y2	N	N2	N3	
animal equipment	12	0	0	5	9	26
Weapons and Ammunition	44	1	0	5	6	56
Costume	0	0	0	0	0	0
design element	0	0	0	0	0	0
equipment for costume	4	0	0	0	0	4
Tools and Equipment	0	0	0	0	0	0
others	3	0	0	0	0	3
合計	63	1	0	10	15	89
正解率						0.72

表５　実験結果 L-2　〈用途/目的〉ための〈用語〉

L_3	正解		不正解			合計
	Y	Y2	N	N2	N3	
animal equipment	14	0	0	5	2	21
Weapons and Ammunition	26	0	1	4	0	31
Costume	10	0	1	3	1	15
design element	2	0	0	0	0	2
equipment for costume	0	0	0	0	0	0
Tools and Equipment	9	0	0	0	1	10
others	0	0	0	0	0	0
合計	61	0	2	12	4	79
正解率						0.8

表６　実験結果 L-3　〈形状〉〈カテゴリ〉

L_4	正解		不正解			合計
	Y	Y2	N	N2	N3	
animal equipment	8	0	2	4	5	19
Weapons and Ammunition	22	0	8	3	8	41
Costume	3	0	2	4	2	11
design element	0	0	0	0	0	0
equipment for costume	3	0	0	0	1	4
Tools and Equipment	0	0	0	0	0	0
others	2	0	0	0	0	2
合計	38	0	12	11	16	77
正解率						0.49

表７　実験結果 L-4　〈位置〉〈用語〉

まず，不正解の殆どが N2，N3 のどちらかにあてはまることが判った．N2 とは複合語の一部が言い換え対象の用語となっているために，言い換えの結果，文が成立しなくなってしまうケースである．例えば「鈴杏葉」という言葉の「杏葉」の部分を「飾りのための杏葉」と言い換えると「鈴飾りのための杏葉」となってしまい不適切である．一方で，「杏葉とよばれる飾り」と言い換えた場合は，「鈴杏葉とよばれる飾り」となり，こちらは文として成立するので正解とした（Y2）．

今回の実験で言い換えられた複合語 55 個のうち正解に判定されたのは 15 個（正解率：0.27）であった．不正解になる主な原因は「小札とよばれる鉄板鋲留式眉庇付冑」「輪鞍から垂れている鐙」のように，言い換えによって追加された部分が複合語を切り離してしまうことによる．

一方，正解になったケースの殆どは（L-1）で，なかでも「鈴杏葉」「下げ美豆良とよばれる髪形」など語尾に言い換え対象語を含む場合に限られる．上記のような複合語の切り離しが起きないことも一つの原因であるが，このような構造の複合語が表すのが＜用語＞の細分類であるため，さらに上位の＜カテゴリ＞を補足した場合に意味的にも適切な表現となると考えられる．

反対に，「体を保護するための挂甲武人埴輪」のように語頭に＜用語＞がある場合は，すべてのパターンで不正解となった．これは，＜用語＞の部分が複合語のなかで修飾的な役割を果たしており，さらに＜用語＞の説明が言い換えで補足されると，その指し示す先が曖昧になるためと考えられる．例の場合だと「体を保護するための」「埴輪」であるかのようになり不適切である．

N3 は，「～は草摺で、腰から太股の部分を保護する付属具である」の「草摺」を「腰や太ももを保護するための草摺」と言い換えるようなケースである．言い換えた結果，「腰や太ももを保護するための草摺で、腰から太股の部分を保護する付属具である」というように，一文内で内容が重複してしまい不適切な文となる．

＜位置＞を説明に付与した（L-4）では，N2，N3 以外の不正解が多く見られた．これは，画像内での位置が明確な場合は説明があてはまるが，解説内の全ての用語が画像に写っているものを表してはいないためである．例えば「江戸・明治期の島田髷によく似た」のように用語が画像そのものを説明していない場合には「江戸・明治期の頭部の島田髷によく似た」となってしまい，不適切な表現となる．

AAT	件数
animal equipment	7
Weapons and Ammunition	8
Costume	4
design element	1
equipment for costume	1
Tools and Equipment	1
others	2

表８　AAT による用語の分類

用語	AAT
障泥	animal equipment
馬鎧	animal equipment
轡	animal equipment
錣	Weapons and Ammunition
挂甲	Weapons and Ammunition
意須比	Costume
島田髷	Costume
龍文	design element
脚結	equipment for costume
円筒埴輪	Tools and Equipment

表９　AAT による用語分類の例

また，用語が表現するものによって適する言い換えパターンに差があるかを調べるために，用語を AAT の Object Facet に従って分類した（表８）．表９は分類の例である．

分類したグループごとの判定結果も分析すると，用語のグループごとに適する言い換えパターンの傾向があることが判った．用語をグループ化してグループごとの言い換え結果を記録していくことで，言い換え候補のランキングなどに使用できると考える．

７．今後の課題

今回の実験では，一文として適切であるかのみを判定したが，言い換えた結果，実際に理解支援になっているのかを判定する必要がある．

また，用語テーブルの属性の管理も大きな問題である．属性は必ずしも一意に決まらない．まず，属性が用語が指すもの全体についてあてはまるのか，それとも一部についてあてはまるのかを区別して扱っていく必要がある．例えば，分析した解説文からは「美豆良」の形状として「肩まで下がった」という属性を検出したが，これは美豆良のなかでも「下げ美豆良」を表していて「上げ美豆良」に関しては当てはまらない．これらを区別し適格に扱えるようにする必要がある．

また，一つのプロパティが複数の属性を持つ場合がある．「縦割りの」と「スカート状の」はともに「草擦」の形状を表しているが，それぞれ適する文脈は異なるだろう．このような複数の属性を管理する仕組みも必要である．

８．まとめ

本研究では，文化遺産に関する解説文の読解支援を目的とした言い換えの自動化の第一段階として，用語の言い換えを取り上げた．この言い換えを実現するための言い換えデータベースを考案し，用語・言い換えパターンの二つのテーブルから機械的に生成した言い換えが，一文中で適切な表現となっているかを判定した．しかし，複合語の分割や内容の重複といった問題があり生成された文は必ずしも適切な表現とはならない．

今後は，まずこれら二つの問題の解消する手法について取り組む予定である．また，言い換えた結果が元の表現より分かりやすい表現になったのかをユーザ実験を通して検証していく予定である．

謝辞

本研究の実施にあたり，独立行政法人国立博物館東京国立博物館が所蔵する文化財の画像およびメタデータを使用させていただいた．

駒澤大学文学部講師の古庄浩明氏，サイエンスライターの財部恵子氏には解説の記述に関して多大な協力をいただいた．深く感謝する．

本研究の一部は，文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」の支援を受けて実施した．

参考文献

[1] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151-198, 2004.

[2] 山本和英. 換言処理の現状と課題. [言語処理ワークショップ, 2001], pp. 93-96, 2001.

[3] http://ceax.ex.nii.ac.jp/ceax/

[4] 山田太造, 相原健郎, 藤沢仁子, 神門典子, 上原祐介, 馬場孝之，長田茂美，安達 淳，"学校教育における文化財コンテンツ利活用のための教育支援システム"，日本教育工学会 研究報告集 JSET06-3，pp. 23-30 (2006 年 5 月)

[5] 榎本聡、室田真男、清水康敬: "「音訓の読み方」と「ふりがな表記」に対応した漢字かな自動変換サーバの開発"，教育システム情報学会誌, Vol.17, No.3, pp.275-284 (2000 年 10 月)

[6] 乾健太郎. コミュニケーション支援のための言い換え. 言語処理学会第 7 回年次大会併設ワークショップ, 2001.

[7]佐藤 理史，土屋 雅稔，村山 賢洋，麻岡 正洋，王 晴晴. 日本語文の規格化. 情報処理学会 研究報告 NL-153, pp.133-140 (2003 年 1 月)

[8] 藤沢仁子，相原,健郎，神門,典子. 文化遺産に関する説明文の対象ユーザに合わせた言い換えの提案. 情報処理学会研究報告，2006(82)，pp.7-12，2006

[9] 熊野正，後藤功雄，田中英輝，浦谷則好，江原暉将. 翻訳用例提示システムの設計・開発・運用. 電子情報通信学会論文誌. J84-D-II(6)，pp.1175-1184，2001.

[10] http://www.gunmaibun.org/osoretoinori/index.html