TR-IT-0329

# TDMT 解析変換知識作成の手引 (日中版)

山本 和英　　垣谷 よし子†

2000 年 1 月

Abstract

TDMT 用解析変換データ (日中) 作成を目的とする、解析変換知識の構成とその仕様について解説を行ない、また作業手順などについて説明する。

エイ・ティ・アール音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories


---

† 株式会社コングレ

# 目次

# 第 1 章

# はじめに

本レポートは、本研究所で研究が進められている TDMT (Transfer-Driven Machine Translation, 変換主導型機械翻訳) の、日本語から中国語へ翻訳するシステムの解析知識、及び変換知識の記述方法をまとめたものである。対象として想定する読者は、日本語と中国語の知識を持ち、Mule 上に開発された TDMT のインターフェース [河井 94] の操作に慣れた作業者 (ルールライター) である。

日中版 TDMT はいくつかの言語ペアのうち最も最近に研究を開始した言語ペアであり、研究開始は 1998 年 1 月である。日英、日韓などと比較して、日中では日本語解析の際に使用するカテゴリ、マーカなどを独自のものに変更し、新たに設計した。また、従来は弊害が大きかったいわゆる複数ヘッドの導入を一切行なわずに記述を試みている。変換パタンによる変項の除去 (原言語にある変項が目的言語にない変換パタン) も危険であるとの判断からこのような操作は一切行なわず、代替手段として、中国語生成部において冗長表現などを排除するようにした。以上のように、記述書式などは日英などと同一であるが、いわゆる日本語文法は異なっている。

本レポートは以下のような構成を取る。2 章においては、翻訳知識作成に直接、あるいは間接的に使用する言語資源である言語コーパス、辞書などに関して説明する。3 章では、解析知識についての定義ならびに書式を、4 章では変換知識についての定義と書式を説明する。5 章では、実際の翻訳訓練においてこれらの解析知識や変換知識をどのように作成すればよいのかについて、実際の例文とその翻訳過程を示すことで説明を行なう。また同時に翻訳訓練の際に使用するツールについても解説を行ない、現時点での翻訳知識の統計データを示す。

また、付録としてマーカの一覧と定義、カテゴリの一覧と定義、(現時点での) 原言語パタンの一覧を添付する。これらはいずれも翻訳知識作成の際に参考となるデータであり、これによって現時点で作成された翻訳知識の概要を捉えることができる。

なお、本レポートに記述されている仕様、統計データ、付録に示した翻訳知識などはすべて 2000 年 1 月 1 日現在のものである。

# 第 2 章

# 言語コーパスおよび辞書

本節では、TDMT において解析知識、変換知識作成に関係する言語資源、すなわち、テキスト、タグ、辞書に関して説明する。なお、本節以降の記述に関しては以下の書式を使用した。

- [ ] ... 省略可能を示す

  例 (((TARGET-PATTERN) [(GENERATION-INFO)])

- 英小文字 ... そのまま記述することを示す
- 英大文字 ... 該当する内容を記述することを示す

  例
```
(define-transfer-dic :j-c t
    ((POS JAPANESE-STRING) TARGET)
    )
```

## 2.1 翻訳訓練対象テキスト (JTEXT, EJTEXT, JEKTEXT)

〔内容〕 翻訳訓練の対象となる文を集めたもの。対象は話し言葉で、話題は旅行に関するものである。翻訳訓練はこれらの中から選ばれる。テキストは以下の２種類に大別される。

1. 二者対話（モノリンガル会話、バイリンガル会話）
2. 基本表現集

前者は、ある場面（例えば、ホテルのフロント）においてある話題（例えば、今晩の部屋の予約）について、二人（例えば、フロントと客）で話された内容をテキストにしたものである。これは、日本語話者と英語話者の二人が通訳を介して収録され、その英語部分が日本語に翻訳されて全体として日本語の会話としたもの（バイリンガル会話）と、二人とも日本語の話者が会話したものをそのままテキストとしたもの（モノリンガル会話）の２種類に分類できる。

後者の基本表現集は、多種多様な旅行会話表現を網羅するために作成された、日本語、英語、韓国語の対訳付き基本表現集である。このため全体として一連の会話になっているわけではなく、例えば類似した表現がいくつも収められている。様々な表現を翻訳可能にするため、これらの一部も翻訳訓練の対象となることがある。

詳細は「言語データベースの概要」[古瀬 95] を参照のこと。

〔管理者〕 データベース管理者およびシステム管理者

〔格納場所〕

- 二者対話： /DB/LDB/JE/EJTEXT+ 以下の *.EJTEXT 名のファイル
- 基本表現集： /DB/LDB/JEK/JEKTEXT 以下の *.JEKTEXT 名のファイル

〔例１：日英バイリンガル会話〕 /DB/LDB/JE/EJTEXT+/FURU94B/AF430011.EJTEXT

J| いらっしゃいませ。
E|Hello.

E|Hello, [uh] I'd like to get a room, please.
J| こんにちは、［あ］一つ部屋が、欲しいんですけど。
：（中略）
J| ［えー］通常チェックインは十時からなんですが、（少し）今十時前ですから、部屋ができてるか確認して参ります。
E-|[well] Normally, check-in time is usually after ten a.m., so I would like to go and check if a room is ready.
E+|Normally, check-in time is ten o'clock, but since it is before ten right now, I will go and check if a room is ready.

※E- は会話収録時の通訳者の訳または話者の発話。
※E+ は会話収録後のネイティブの proofreading。

〔例２：基本表現集〕 /DB/LDB/JEK/JEKTEXT/S9503/A4JP.JEKTEXT

J|XX|AU4JP001| 今晩のホテルを予約したいのですが。
E|XX|AU4JP001|I would like to make a reservation for tonight.
K|XX|AU4JP001| 오늘 밤 호텔을 예약하고 싶은데요.

J|XX|AU4JP002| 何名さまですか。
E|XX|AU4JP002|For how many people?
K|XX|AU4JP002| 몇 분이십니까?

## 2.2 日本語形態素タグ (JMOR)

〔目的〕 べた書き文字列の文を単語単位（形態素）に分解し、品詞や活用形などの情報を付与したデータ。形態素情報、タギングデータとも言う。翻訳訓練の際は形態素解析[1]を行なわず、あらかじめ解析済みのこのタグをもとに各解析変換知識を作成する。タグは、訓練対象テキストそれぞれに対して、システム管理者が /usr/local/TDMT/tdmt-multi-dev/build/text-definitions.lisp にて指定する。

詳細は「TDMT 用日本語形態素仕様明細」[溝口 99a] を参照のこと。

〔管理者〕 タグ管理者およびシステム管理者

〔格納場所〕

/DB/LDB/TDMT/JMOR /LDB-JE/DEN94B/*.JMOR
/LDB-JE/FURU94B/*.JMOR
/LDB-JE/SOBA93/*.JMOR
/LDB-JE/SOBA94A/*.JMOR
/LDB-JEK/C9608/*.JMOR
/LDB-JEK/S9503/*.JMOR
/LDB-JEK/S9510/*.JMOR
/SLDB/*.JMOR

〔書式〕 発話 ID| 発声 ID| 文節 ID| 単語 ID| 表記形 | 読み | 正規形 | 品詞 | 活用型 | 活用形 | 音便 |

〔例〕 /FURU94B/AF430011.JMOR

10|0010|10|10| いらっしゃい | イラッシャイ | いらっしゃる | 本動詞 | 特殊ラ | 連用 | |
10|0010|10|20| ませ | マセ | ます | 助動詞 | 特殊サ | 命令 | |
10|0010|10|30|。 | |。 | 記号 | | | |
20|0020|20|40| こんにちは | コンニチハ | こんにちは | 感動詞 | | | |
20|0020|30|50| 一つ | ヒトツ | 一つ | 普通名詞 | | | |
20|0020|40|60| 部屋 | ヘヤ | 部屋 | 普通名詞 | | | |
20|0020|40|70| が | ガ | が | 格助詞 | | | |
20|0020|50|80| 欲しい | ホシイ | 欲しい | 形容詞 | 形容詞 | 基本 | |
20|0020|50|90| ん | ン | ん | 準体助詞 | | | |
20|0020|50|100| です | デス | です | 判定詞 | 特殊サ | 基本 | |
20|0020|50|110| けど | ケド | けど | 接続助詞 | | | |
20|0020|50|120|。 | |。 | 記号 | | | |

[1] 任意の文を自動的に形態素に分割して、品詞を決める処理。

## 2.3　日本語意味コード辞書 (sem-code dic)

〔目的〕 角川書店「類語新辞典」の三桁の数字をもとに、形態素ごとに意味コードを付与した辞書。システムが意味距離計算により、類似用例を検索する際に参照する。また、翻訳訓練の際に意味距離計算を行ないながら知識を作成することもできる（5節「解析変換知識作成手順」参照）。

〔管理者〕 意味コード辞書管理者

〔格納場所〕 /usr/local/TDMT/tdmt-multi-dev/j-morph/dic/jma-atr-sem-code.text

〔書式〕

(REG-EXP POS SEM-CODE)

- REG-EXP ... 日本語形態素の正規形
- POS ... 日本語品詞シンボル
- SEM-CODE ... 意味コード（複数指定可）

〔例〕

("扱い" 普通名詞 "382")
("扱う" 本動詞 "380" "382" "489" "787")
("宛" 普通名詞 "109c")
("宛先" 普通名詞 "109c")
("安い" 形容詞 "171a" "693")
("安まる" 本動詞 "405")
("安心" サ変形容名詞 "494" "694")
("安心する" 本動詞 "494" "694")

※詳細は「日本語意味コード辞書作成の手引」[溝口 98] を参照のこと。

## 2.4 日中変換辞書 (transfer dic)

〔目的〕 日本語形態素に対応する中国語訳を記述する辞書。(日中) 対訳辞書とも言う。一語につき、一対訳が登録できる。原言語に対しては (POS JAPANESE-STRING) が一単位となる。目的言語 (TARGET) に対しては、品詞情報を伴った形態素列を記述するのが望ましい。原言語の品詞は、「TDMT 用日本語形態素仕様明細」[溝口 99a] に従うものとする。ここで記述された目的言語は、変換知識のパタン中で別の対訳に変更することができる。詳細は 4.7節「ローカル辞書」を参照のこと。

〔管理者〕 解析変換知識作成者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/dic/japanese-to-chinese.lisp

〔書式〕

```
(define-transfer-dic :j-c t
  ((POS JAPANESE-STRING) TARGET))
```

- define-transfer-dic ... 変換辞書の定義を宣言
- :j-c ... 翻訳の方向。この場合は日中翻訳を示す。
- t ... 辞書データをロードする際にハッシュテーブルを作り直す
- POS ... 日本語品詞シンボル
- JAPANESE-STRING ... 日本語正規形または表層形
- TARGET ... 目的言語の対訳と品詞情報のリスト

〔例〕

```
(define-transfer-dic :j-c t
  ((普通名詞 "アミューズメント") (名 "娱乐")) ;;; "yu2 le4"
  ((サ変名詞 "チェックイン") (动 "办理住宿登记")) ;;; "ban4 li3 zhu4 su4 deng1 ji4"
  ((本動詞 "着ける") (动 "能到达")) ;;; "neng2 dao4 da2"
  ((形容詞 "あやしい") (形 "可疑")) ;;; "ke3 yi2"
  ((副詞 "念のため") (副 "为了慎重起见")) ;;; "wei4 le0 shen4 zhong4 qi3 jian4"
  ((接続詞 "すると") (连 "那么")) ;;; "na4 me0"
  ((感動詞 "こんにちは") (叹 "你好")) ;;; "ni2 hao3"
```

上の例中の ;;; に続く記述は、発音の情報である。TDMT の解析、変換処理とは直接の関係はないが、翻訳結果を音声出力する際に必要な情報であるため、本辞書から抽出、加工を可能なように付加している。ピンインと声調で示されるが、3 声の連続で 2 声に変調するものや、"一 yi1"の声調変化などは、実際に発音される通りに記述する。

## 2.5　中国語生成規則 (gen-rule.lisp)

〔目的〕 生成処理において、正しい中国語を生成するために、中国語の要素を挿入或いは削除する規則。

〔管理者〕 解析変換知識作成者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/c-generation/gen-rule.lisp

〔書式〕

(BEFORE "-)" AFTER (OPERATION))

- BEFORE ... (要素 1 要素 2 ... 要素 m)
- AFTER ... (要素 1' 要素 2' ... 要素 m')
- OPERATION ... (:I i "文字列") または (:D i "NIL")

要素 ... 品詞 または 単語
I ... 挿入時
D ... 削除時
i ... 挿入 または 削除

〔例〕

((套间 房间) "-)" (套间) ((:D 1 NIL)))

「スイートのお部屋」という入力文は、生成処理前には、"套间房间 "と訳出されるが、
この生成処理を行うことによって、語の重複を避けている。

((下午 八) "-)" (晚上 八) ((:I -1 "晚上") (:D 0 NIL)))

「午後八時」という入力文は、生成処理前には、"下午八点 "と訳出されるが、
この生成処理を行うことによって、中国語として自然な "晚上八点 "と訳出される。

# 第 3 章

# 解析知識 (a-data)

解析処理では、後述の変換知識による処理がしやすいように、入力文の原言語形態素列を加工する。解析知識には、以下の 4 種類に分類される。

1. 表記のゆれの吸収 (replace-word)
2. 目的言語に応じた形態素情報の修正 (lexical-transformation)
3. マーカの挿入 (local-transformation)
4. (total-transformation)

実際の処理は上記の順で行なわれる。但し、日中翻訳では total-transformation は使用していない。各種解析知識の詳細については、以下の項目ごとに説明する。

## 3.1 形態素表記のゆれの吸収 (replace-word)

〔目的〕 タグおよび形態素解析結果の表記の揺れを統一し、変換知識のパタン数および用例数を削減するための正規形置換情報。日本語を原言語とする翻訳処理に共通の知識であるため、追加や修正を行なう際は、関係者で協議の上、決定する。

〔管理者〕 正規形置換情報管理者

〔格納場所〕 /usr/local/TDMT/tdmt-multi-dev/j-morph/dic/replace-word-info.lisp

〔書式〕

((POS REG-EXP) (REVISED-REG-EXP))

- POS ... 日本語品詞シンボル
- REG-EXP ... 日本語正規形
- REVISED-REG-EXP ... 置換後の正規形

〔例〕

((サ変名詞 "素泊り") ("素泊まり"))
((サ変名詞 "手続") ("手続き")
((普通名詞 "サンドウィッチ") ("サンドイッチ"))
((普通名詞 "シャンペン") ("シャンパン"))
((本動詞 "くりかえす") ("繰り返す"))
((本動詞 "したがう") ("従う"))
((補助動詞 "出来る") ("できる"))
((補助動詞 "下さる") ("くださる"))

## 3.2 形態素情報の修正 (lexical-transformation)

〔目的〕 変換処理で形態素を扱いやすくするために、形態素情報の修正（正規形や品詞の変更および形態素の合成・分割など）を行なう知識。

〔管理者〕 解析変換知識作成者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/a-data/*.lisp

〔書式〕

以下のように lexical-transformation は、矢印 (=⟩) の左辺が形態素パタンの条件式、右辺が処理の定義式という構成になっており、それらは 1 つのルールの中でリスト構造をとる。

```
(define-lexical-transformation RULE-NAME :j-c PRIORITY
  (((MORPH-INFO-LIST-1) (MORPH-INFO-LIST-2) ... (MORPH-INFO-LIST-N)) =⟩
   (lex ((ARGUMENT) REVISED-MORPH-INFO-LIST)))
   .........
  )
```

- define-lexical-transformation ... 解析知識の定義を宣言
- RULE-NAME ... 任意のルール名。ただし、解析知識内でユニークとする。
- :j-c ... 翻訳の方向。この場合は日中翻訳を示す。
- PRIORITY ... 処理の実行優先番号
- MORPH-INFO-LIST-1, MORPH-INFO-LIST-2,...MORPH-INFO-LIST-N ... 形態素情報のリスト（記述要素 =:word(表層形), :reg-exp(正規形), :pos(品詞), :conj-form(活用形) :pron(読み) :start(発声の開始), :end(発声の終了)）
- lex ... 形態素情報修正の宣言
- ARGUMENT(引数) ... 左辺の形態素列の何番目かを示す番号と任意の文字列の並び。
- REVISED-MORPH-INFO-LIST ... 修正後の形態素情報のリスト（記述要素 =:word(表層形), :reg-exp(正規形), :pos(品詞), :conj-form(活用形)）。ただし、
  :word と :reg-exp の値は文字列やシンボルで指定するか、引数の形態素列の番号で指定する。
  :pos と :conj-form の値はシンボルで指定するか、引数の形態素列の番号で指定する。引数の番号で指定した場合は、該当する形態素が持つ値をコピーする。:all-copyは該当する形態素が持つ:pos, :conj-form, :pron の値をコピーする。:all-copy と :pos, :conj-form の両方を指定した場合は、後の指定が優先される。後述する terminal rule用に合成する場合は、:compoundの指定をこの revised-morph-info-list に入れる。

右辺の処理定義式は、普通は lex で宣言するが、左辺で指定した形態素列に修正を加えない形態素が含まれる場合や、形態素の分割のように新しい形態素を加える場合などは、以下の形態素定義式 define-morphで形態素情報を定義して呼び出す。また、引数の形態素列の番号を指定する際に、異なるフィールドの情報を用いる場合には、以下の文字列定義式 define-stringで文字列としての定義を行なってから呼び出す。

```
(define-morph @RULE-NAME :j-c (ARGUMENT)
  REVISED-MORPH-INFO-LIST)

(define-string $RULE-NAME :j-c (ARGUMENT)
  STRING-INFO)
```

- define-morph ... 形態素の定義を宣言
- define-string ... 文字列の定義を宣言
- @RULE-NAME, $RULE-NAME ... 任意の定義式名。形態素の定義と文字列の定義を区別するため、先頭に @ か $ が必要。解析知識内でユニークとする。
- :j-c ... 翻訳の方向。この場合は日中翻訳を示す。
- ARGUMENT(引数) ... 左辺の形態素列の何番目かを示す番号と任意の文字列の並び。
- REVISED-MORPH-INFO-LIST ... 修正後の形態素情報のリスト（記述要素 =:word(表層形), :reg-exp(正規形), :pos(品詞) , :conj-form(活用形))。記述方法は上に同じ。
- STRING-INFO ... 定義に用いる文字列情報（記述要素 =:word(表層形), :reg-exp(正規形))。記述方法は上に同じ。

〔例〕

1. 形態素の合成

```
(define-lexical-transformation compound-ki-ni-iru :j-c 3
  ((((:word . "気")) ((:word . "に")) ((:reg-exp . "入る"))) =>
   (lex ((1 2 3) (:pos . 本動詞) (:reg-exp . "気に入る") (:conj-form . 3)))
   ))
```

分割されている日本語形態素を合成して、一形態素にするルール。例の「気に入る」は、「気」の対訳 "气 "や、「入る」の対訳 "加 "の意味を持たず、三語を合成して、初めて "称心 "の対訳を決定することができる。このルールでは、入力文の形態素列を部分的に照合し、1 番めの形態素の表層形が「気」、2 番めの形態素の表層形が「に」、3 番めの形態素の表層形が「入る」であることを形態素パタンの条件としている。パタンの条件が合えば、三語は「本動詞」として合成され、正規形は「気に入る」、活用形は引数で指定した 3 番めの形態素（「入る」）の値を用いる。

2．形態素の除去

```
(define-lexical-transformation delete-ne :j-c 2
  ((((:pos . 連体助詞) (:word . "の")) ((:pos . 終助詞) (:reg-exp . "ね")))
   (1)
   ))
```

不必要な日本語形態素を除去するルール。例のように、発声の途中に出現する冗長な「ね」は、格関係の構造を容易に導くために除去したい。このルールでは、1番めの形態素が連体助詞「の」、2番めの形態素が終助詞「ね」である場合に、2番めの「ね」を除去している。

3．正規形の修正

```
(define-lexical-transformation revised-reg-nn :j-c 1
  ((((:word . "ん") (:pos . 準体助詞))) =⟩
   (lex ((1) (:all-copy . 1) (:reg-exp . "の") (:word . "の")))
   ))
```

日本語形態素の正規形を修正するルール。変換知識でのルールや用例を集約するために、機能語はなるべく正規化するのが望ましい。このルールでは、形態素1語を照合し、表層形が「ん」、品詞が準体助詞である場合に、まず、:all-copyでその形態素情報の:posと:conj-formの値をコピーし、その後で正規形と表層形を「の」に修正している。

```
(define-lexical-transformation replace-okosu :j-c 1
  ((((:pos . 本動詞) (:reg-exp . "お越す")) ((:pos . 格助詞))) =⟩
   (@revised-okoshi 2)
   ))
(define-morph @revised-okoshi :j-c (1)
  (:all-copy . 1) (:pos . 普通名詞) (:reg-exp . "お越し"))
```

正規形を修正する際に、別の形態素定義式を呼び出すルール。このルールでは、1番めの形態素の品詞が「本動詞」、正規形が「お越す」、2番めの形態素の品詞が「格助詞」であることを条件としている。処理は、1番めの形態素に対して定義式@revised-okoshiを呼び出し、その形態素情報を用いて、品詞を「普通名詞」、正規形を「お越し」に修正している。2番めの形態素に対しては、修正を行なわない。

4．読み情報を利用した正規形の修正

```
(define-lexical-transformation revised-verb-haireru-ireru :j-c 1
  ((((:pos . 本動詞) (:word . "入れる") (:pron . "ハイレル"))) =⟩
   (lex ((1)(:all-copy . 1) (:reg-exp . "入れる＜ハイレル＞")))
   ))
```

日本語形態素の読み情報を参照して、正規形を修正するルール。例の「入れる」には、「イレル」と「ハイレル」の二つの読みがあり、それぞれ意味が異なるため、変換処理で曖昧性が出るのを抑えるべく、読み情報を参照して区別することができる。このルールでは、本動詞「入れる」の読みが「ハイレル」の場合、:all-copy でその形態素情報の:pos と:conj-form の値をコピーし、その後で正規形を「入れる＜ハイレル＞」と修正している。

現在、読み情報を指定できるのは、以下の 13 件である。

普通名詞　「方＜ホウ＞」「方＜カタ＞」「日＜タチ＞」「一＜ヒト＞」「二＜フタ＞」「三＜ミ＞」「風＜フウ＞」

本動詞　「入れる＜ハイレル＞」「入れ＜ハイレ＞」「入れれ＜ハイレレ＞」「入れる＜イレル＞」「入れ＜イレ＞」「入れれ＜イレレ＞」

5．活用形を利用した正規形の修正

```
(define-lexical-transformation revised-reg-iru-ichidan :j-c 1
  ((((:word . "い") (:reg-exp . "いる") (:conj-form . ない))) =⟩
   (lex ((1) (:all-copy . 1)(:reg-exp . "いる一")))
   ))
```

活用形を参照して、日本語形態素の正規形を修正するルール。例の「いる」には、上一段活用の「居る」と五段活用の「要る・射る」などの意味があるため、区別するために正規形を修正する。このルールでは、形態素を 1 語だけ照合し、表層形が「い」、正規形が「いる」、活用形が「ない」形である場合に、:all-copyでその形態素情報の:pos と:conj-form の値をコピーし、その後で正規形を「いる一」に修正している。

6．terminal rule 用の合成

```
(define-lexical-transformation num-hon :j-c 6
  ((((:pos . 数詞)) ((:word . "本"))) =⟩
   (lex ((1 2) (:pos . 普通名詞) (:reg-exp . "N本") (:compound 1 2)))
   ))
```

日本語形態素を合成して、変項を持つ形態素を作成するルール。「一本」「二本」など、数詞を含む表現は、単に形態素を合成すると、無限にパタンを用意しなくてはならない。そのため、形態素の合成を行なうが、数量表現部分は内部処理として変換を行ないたい。このような場合には terminal rule と呼ばれる変換処理を用い、そのための特別な合成であることを:compoundで指定する。このルールでは、1 番めの形態素の品詞が「数詞」、2 番めの形態素の表層形が「本」である場合に、terminal rule 用に二語を普通名詞として合成している。合成後の正規形は「N本」となる。

```
(define-lexical-transformation gousei-adj-me :j-c 5
  ((((:pos . 形容詞))((:pos . 接尾辞)(:word . "め"))) =>
   (lex ((1 2)(:pos . 普通名詞) (:reg-exp $word-1 2) (:compound 1 2)))
    ))
(define-string $word-1 :j-c (1)
  (:word . 1))
```

terminal rule用の合成を行なう際に、別の文字列定義式を呼び出すルール。形容詞＋接尾辞「め」は変換の都合で合成する必要がある。そのため、第１項の形態素の品詞が「形容詞」、第２項の形態素の表層形が「め」、品詞が「接尾辞」である場合に、terminal rule用に二語を「普通名詞」として合成（:compound）している。合成後の正規形には、引数で指定した第１項の形態素の表層形の値と第２項の形態素の正規形の値を使うため、第１項の形態素の表層形を文字列として定義した$word-1を呼び出している。

7．その他

日中翻訳では現在まだ出現していないが、上記の他に、形態素の分割、発声の終了情報を利用した形態素の除去なども可能である。

〔優先順位別処理一覧〕

※優先順位は、数字の小さい方が先に処理が行われる。優先度に揺れがある場合は、括弧内に揺れの範囲を示す。

＜優先順位 1 ＞

- タグの誤り修正
- 変換の都合のための品詞変更
- 変換の都合のための表層形変更
- 数詞の terminal rule 用合成

＜優先順位 2 ＞

- 不必要な形態素の除去
- 変換の都合のための正規形変更（揺れ 1 － 3）

＜優先順位 3 ＞

- 形態素の合成（揺れ 2 － 4）

＜優先順位 4 ＞

＜優先順位 5 ＞

- 一般的な terminal rule 用合成（揺れ 4 － 6）

＜優先順位 6 ＞

＜優先順位 7 ＞

## 3.3 マーカの挿入 (local-transformation)

〔目的〕 変換処理で構造を導きやすくするために、特定の形態素間の境界を表すマーカの挿入を行なう。

〔管理者〕 解析変換知識作成者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/a-data/local*.lisp

〔書式〕

以下のように local-transformation は、矢印 (=⟩) の左辺が形態素パタンの条件式、右辺が処理の定義式という構成になっており、それらは 1 つのルールの中でリスト構造をとる。

```
(define-local-transformation RULE-NAME :j-c PRIORITY
  (((MORPH-INFO-LIST-1) (MORPH-INFO-LIST-2) ... (MORPH-INFO-LIST-N)) =⟩
   (ARGUMENT & MARKER)
   )
  .........
  )
```

- define-local-transformation ... 解析知識のパタン定義の宣言
- RULE-NAME ... 任意のルール名。ただし、解析知識内でユニークとする。
- :j-c ... 翻訳の方向。この場合は日中翻訳を示す。
- PRIORITY ... 処理の実行優先番号
- MORPH-INFO-LIST-1, MORPH-INFO-LIST-2,...MORPH-INFO-LIST-N ... 形態素情報のリスト (記述要素 =:word(表層形), :reg-exp(正規形), :pos(品詞), :conj-form(活用形), :pron(読み), :start(発声の開始), :end(発声の終了))
  or/nor で複数指定可。
- ARGUMENT & MARKER(引数とマーカ) ... マーカの挿入位置を示した、左辺の形態素列。

〔例〕

日中で使用されるマーカの一覧を付録 A節に示す。

1．前後に連接する品詞情報を示すマーカ

```
(define-local-transformation sn-v :j-c 3
  ((((:pos . サ変名詞)) ((:pos . 本動詞))) =⟩
   (1 ⟨sn-v⟩ 2)
   ))
```

品詞の連接情報を示す一般的なマーカを挿入するルール。内容語である「サ変名詞」と「本動詞」が連接する場合には、変換処理を行なう際に境界を示すマーカが必要になる。このルールでは、1番めの形態素の品詞が「サ変名詞」、2番めの形態素の品詞が「本動詞」である場合、二語の形態素の間にマーカ ⟨sn-v⟩ を挿入している。

2．前の品詞情報のみを示す統一的なマーカ

```
(define-local-transformation interj- :j-c 3
  ((((:pos . 感動詞))
    ((:pos . (or 代名詞 普通名詞 サ変名詞 サ変形容名詞 形式名詞 形容名詞 形容詞
                 連体形容詞 本動詞 感動詞 ローマ字 副詞 接続副詞 数詞 準数詞
                 接頭辞 接続詞 連体詞)))) =⟩
   (1 ⟨i-⟩ 2)
   ))
```

後続する品詞を問わずに、特定の品詞＋後続表現という形の統一的なマーカを挿入するルール。このルールでは、1番めの形態素の品詞が「感動詞」、2番めの形態素の品詞が「代名詞」以下の 18 種類の品詞である場合に、統一的なマーカ⟨i-⟩ を挿入している。後接する形態素の品詞を特定しなくても原言語構文解析に曖昧性が出ない場合は、後接する形態素の情報を省略したマーカを用いることができる。このマーカを利用することにより、変換知識のパタンや用例の集約化がはかれる。

3．活用形など個別な情報を示すマーカ

```
(define-local-transformation /renyou- :j-c 1
  ((((:conj-form . 連用))
    ((:pos . (or 代名詞 普通名詞 サ変名詞 サ変形容名詞 形式名詞 形容名詞 形容詞
                 連体形容 本動詞 感動詞 ローマ字 副詞 接続副詞 数詞 準数詞
                 接頭辞 接続詞 連体詞)))) =⟩
   (1 ⟨/renyou-⟩ 2)
   ))
```

個別な情報を示すマーカを挿入するルール。

形容詞、本動詞、助動詞等の全活用語が、文中で「連用」形で使われ、後続する品詞が「代名詞」以下の 18 種類の品詞である場合に、マーカ⟨/renyou-⟩ を挿入している。品詞情報を示すマーカに先行して、こうした活用形を示すマーカを挿入することによって、文の構造を導きやすくしている。

```
(define-local-transformation sou-denbun :j-c 3
  ((((:conj-form . 基本)) ((:word . "そう")(:pos . (or 準体助動詞 助動詞)))) =⟩
   (1 ⟨sou/denbun⟩ 2)
   ))
```

伝聞を示す「そう」と、様態を示す「そう」の区別をするためのマーカ。

伝聞を示す「そう」（準体助動詞）は、本動詞、形容詞、助動詞、判定詞等活用する品詞全ての基本形に後接する。様態を示す「そう」（助動詞）は、本動詞の連用形、形容詞のそう形、或いは、形容名詞にそのまま後接する。この情報をマーカとして挿入して、構造の決定や訳し分けに利用することができる。このルールでは、１番めの形態素の活用形が「基本」形、２番めの形態素の表層形が「そう」、品詞が「準体助動詞」か「助動詞」である場合に、マーカ ⟨sou/denbun⟩ を挿入している。品詞を「準体助動詞」か「助動詞」に指定しているのは、副詞の「そう」と区別したいためであり、また「準体助動詞」に限定していないのは、音声による入力文等で、日本語形態素解析に誤りがあっても対応できるように考慮したためである。

## 3.4　マーカ挿入の基本方針

マーカ挿入の基本方針は、内容語か機能語かによって異なる。

内容語とは、単独で意味を持つ語で、通常は意味コード辞書に意味コードが登録されている。変換知識においては、用例として扱われることが多い。一方、機能語は、単独では意味を持たない語であり、通常は意味コードがふられていない。以下に内容語と機能語の品詞シンボルを示す。

＜内容語＞

代名詞／普通名詞／サ変名詞／サ変形容名詞／形式名詞／形容名詞／形容詞／連体形容詞／本動詞／感動詞／ローマ字／副詞／接続副詞／数詞／準数詞

＜機能語＞

接頭辞／接続詞／連体詞／副助詞／接尾辞／人称接尾辞／助動詞／準体助動詞／係助詞／格助詞／連体助詞／接続助詞／並立助詞／準体助詞／終助詞／判定詞／補助動詞

内容語同士が連接する場合は、形態素間に品詞情報を示すマーカを必ず挿入する。機能語の場合は、品詞の組合せや条件によってマーカを挿入するかどうかが異なる。活用形等の個別な情報を示すマーカはこれに先行して挿入する。

〔優先順位別処理一覧〕

＜優先順位 1 ＞

- 基本形の情報を示すマーカの挿入
- 連用形の情報を示すマーカの挿入

＜優先順位 2 ＞

＜優先順位 3 ＞

- 一般的な品詞情報を示すマーカの挿入
- 様態の「そう」と 伝聞の「そう」を区別するためのマーカの挿入

＜優先順位 4 ＞

＜優先順位 5 ＞

# 第 4 章

# 変換知識 (p-data)

## 4.1 概要

〔目的〕 この翻訳システムの中心モジュールである変換処理を行なうために参照される。解析処理の済んだ入力文は、変換パタンの用例との意味距離を構成要素ごとに計算され、最小距離のパタンを用いて目的言語への構造変換が行なわれる。変換の際には、生成処理用の情報が付加され、出力データは生成処理へと渡される。

〔管理者〕 解析変換知識作成者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/p-data/*.lisp

〔書式〕

```
(define-pattern RULE-NAME :j-c CATEGORY
  (SOURCE-PATTERN) (:head N)[(CATEGORY-CONTROL)] =⟩
  (((TARGET-PATTERN) ([GENERATION-INFO])
   (
    (EXAMPLES-LIST)
    )
   [LOCAL-DIC-LIST]
   ))
```

- define-pattern ... 変換パタンを定義するための宣言。他に、define-regular-patternがある。
- RULE-NAME ... 任意のルール名。ただし、変換知識内でユニークとする。
- :j-c ... 翻訳の方向。この場合は日中翻訳を示す。
- CATEGORY ... カテゴリ（原言語構造の分類名）
- SOURCE-PATTERN ... 原言語パタン（記述要素：変項・マーカ・日本語形態素）
- :head N ... ヘッドの指定（記述要素：数値）
- CATEGORY-CONTROL ... 下部構造の制限（記述要素：変項・カテゴリ名・品詞・構造分類用の集合体名）。省略可。

- TARGET-PATTERN ...　目的言語パタン（記述要素：変項・中国語形態素）
- GENERATION-INFO ...　生成情報（記述要素：中国語句型・品詞・数値）。省略可。
- EXAMPLES-LIST ...　用例（記述要素：日本語形態素）
- LOCAL-DIC-LIST ...　ローカル辞書（記述要素：変項・日中品詞・日中形態素）。省略可。

※詳細は、次節以下で説明する。

## 4.2 原言語パタン / 目的言語パタン / 用例

変換知識は、原言語表現から目的言語表現への変換をパタン形式で記述する。下の変換パタンの例では、(?x "に" ?y) が原言語パタン ("に"が固定項、?x と ?y が変項) であり、(!y !x) や (!y "在" !x)... は目的言語パタンである。原言語パタンと目的言語パタンの対応関係は、必ずしも1対1ではなく、原言語パタン1つに対して、複数の目的言語パタンが考えられることが多い。そのパタンを選択する際の基準となるのが用例である。入力文は、意味コード辞書 (2.3節「意味コード辞書」参照) を用いて用例との類似度が計算され、意味的に最も近い用例を持つ目的言語パタンが対訳として選ばれる。

```
(define-pattern particle-ni-c :j-c c
  (?x "に" ?y) (:head 3) =⟩
  (((!y !x)(動短 1 2))
   (
    (("郵便局") ("行く"))
    (("ツアー") ("参加"))
   )
  )
  (((!y "在" !x) (動短 1 (介 2) 3))
   (
    (("ホテル") ("滞在"))
    (("ここ") ("座る"))
   )
  )
 )
```

例えば、ここで、「旅館に泊まる」という入力文があったとする。変項? x を「旅館」、?y を「泊まる」として、各用例との意味距離を計算する。その結果、入力文「旅館に泊まる」は (("ホテル") ("滞在")) の用例と距離が近いということが導き出されたとすると、目的言語パタン (!y "在" !x) が選ばれる。(!y "在" !x) は、原言語の ?x の部分と ?y の部分を入れ替え、固定項の "に"を "在"に変換することを意味する。「旅館」「泊まる」という形態素単位の変換は、日中変換辞書 (2.4節「変換辞書」参照) から対訳を得て行なわれる。(!y "在" !x) の横に記述されている (動短 1 (介 2) 3) は生成情報で、(!y "在" !x) が動詞短語、1番めの形態素が !y、2番めの形態素が介詞、3番めの形態素が !x であることを表している。

## 4.3 パタンの種類

変換パタンの定義には 2 種類あり、原言語側の形態素やマーカをすべて表層形のままでマッチングする場合は define-pattern、すべて正規形でマッチングする場合は define-regular-patternを用いる。以下の例のように、define-regular-pattern で定義すると、正規形が「ていただく」である活用形(「ていただか」「ていただき」..)のすべてに対応できる。また、解析知識で正規形を「ていただく」に修正している形態素があれば、それについても同様の処理が行なえる。

```
(define-regular-pattern modal-teitadaku-reg :j-c cx
  (?x "ていただく") (:head 1)  =⟩ ←「ていただか」「ていただき」.. とマッチングする
  (((!x))
   (
    (("教える"))
    (("紹介"))
   )))
```

また、マーカについても正規形 ⟨⟩が用意されており、前後の形態素の種類に関わらず、共通に用いることができる。

```
(define-regular-pattern tsuki-reg :j-c nx
  (?x ⟨⟩ "付き") (:head 1)  =⟩  ←マーカの正規形 ⟨⟩
  ((("包括" !x) (動短 (動 1) 2))
   (
    (("ワンドリンク"))  ←普通名詞 ⟨n-n⟩
    (("食事"))          ←サ変名詞 ⟨sn-n⟩
    )))
```

## 4.4 カテゴリ

各パタンの定義で指定するカテゴリは、原言語構文解析を行なう際の構造分類名であり、以下の 14 種類がある。

| 名称 | 用途 |
|---|---|
| u | 接続詞・感動詞・呼びかけなど、文全体を修飾する語／重文 |
| sx | 終助詞・文末表現 |
| s | 主語、時間表現を含む文 |
| cx | c ＋助動詞類 |
| c | 主語、時間表現以外の格関係 |
| px | p ＋助動詞類 |
| p | 合成用言 |
| a | 副詞＋形容語 |
| na | 形容詞＋名詞類 |
| nnx | nn ＋接辞 |
| nn | 助詞またはマーカを伴う名詞句 |
| nx | 接頭辞・連体詞・接尾辞・副助詞など名詞に前接或いは後接する語 |
| n | 複合名詞 |
| terminal | 数詞を伴う頻出表現／派生形 |

注) カテゴリ terminalのパタンは、解析処理において: compoundの指定をして合成された語句のみに適用される。既に一つの形態素として合成されているため、構造を作る際には形態素としての扱いを受ける。

＜原言語構文解析の例＞

以下のように原言語構文解析が行なわれ、適用されたパタンの木構造が得られる。

```
"お土産はどこで売っていますか"
TOP [(?X か) --- SX]
 |--?X [(?X は ?Y) --- S]
     |--?X [(お ?X) --- N]
     |   |--?X [(土産)]
     |
     |--?Y [(?X ます) --- CX]
         |--?X [(?X ている) --- CX]
             |--?X [(?X で ?Y) --- C]
                 |--?X [(どこ)]
                 |
                 |--?Y [(売っ)]
```

```
"できれば同じ部屋がいいんですけど"
TOP [(?X けど) --- SX]
 |--?X [(?X のです) --- SX]
     |--?X [(?X 〈ADV-〉 ?Y) --- S]
         |--?X [(できれば)]
         |
         |--?Y [(?X が ?Y) --- C]
             |--?X [(?X 〈AN-〉 ?Y) --- NN]
             |   |--?X [(同じ)]
             |   |
             |   |--?Y [(部屋)]
             |
             |--?Y [(いい)]
```

```
"電話番号は零三四四四三一七零零です"
TOP [(?X です) --- SX]
 |--?X [(?X は ?Y) --- S]
     |--?X [(電話番号)]
     |
     |--?Y [(0344431700) --- compound]
```

※4.9節にて、カテゴリ別変換パタン定義の基本方針とパタン例を述べる

## 4.5 下部構造の制限

原言語構造を決定する際、ありえない構造を排除して構造の曖昧性の数を抑制するため、パタンごとに下部構造を制限することができる。下部構造の制限は、カテゴリや個々の形態素の品詞名で指定するが、それらを新たにグループ化した集合体（以下の4種類）を用いると簡潔に記述できる。

word-n(体言性の形態素)

group-n(体言性の形態素および構造)

word-p(動詞類の形態素)

group-a(形容詞類の形態素および構造)

この集合体のメンバは /usr/local/TDMT/tdmt-multi-dev/transfer/j-c/dic/ category-jc.lispにて定義されている。

〔書式〕

```
(define-category-set GROUP-NAME :j-c MEMBER)
```

- define-category-set ... 集合体を定義する
- GROUP-NAME ... 集合体の名前
- :j-c ... 翻訳の方向を示す。この場合は日中翻訳。
- MEMBER ... 集合体のメンバ（記述要素：カテゴリ・品詞シンボル）

〔リスト〕

```
(define-category-set word-n  :j-c 普通名詞 固有名詞 代名詞 サ変名詞 形容名詞
                                  サ変形容名詞 形式名詞 数詞 準数詞 ローマ字)

(define-category-set group-n :j-c 普通名詞 固有名詞 代名詞 サ変名詞 形容名詞
                                  サ変形容名詞 形式名詞 数詞 準数詞 ローマ字
                                  副詞 接続副詞 n nx na nn nnx)

(define-category-set word-p  :j-c 本動詞 形容詞 補助動詞 連体形容詞)

(define-category-set group-a :j-c 形容詞 形容名詞 a)
```

各カテゴリが取り得る下部構造は、同じく category-jc.lispに以下のように定義されている。

〔書式〕

```
(define-category CATEGORY-NAME :j-c MEMBER)
```

- define-category ... カテゴリの定義を宣言する
- CATEGORY-NAME ... カテゴリの名前
- :j-c ... 翻訳の方向を示す。この場合は日中翻訳。
- MEMBER ... カテゴリのメンバ(記述要素：カテゴリ・品詞シンボル・集合体の名前)

〔例〕

```
(define-category c  :j-c group-n word-p p px c
                        group-a 感動詞)
```

例に上げたカテゴリcは、下部構造としてp, px, cの構造の他、group-n, word-p, group-aで指定された構造と感動詞を取り得る。変換パタンでは、これに対して制限を加え、構造の曖昧性を排除する。以下のパタンでは、(:x group-n) の指定を入れることにより、変項xが取り得る構造や語句をgroup-n のメンバに制限している。

```
(define-pattern particle-wo-c :j-c c
  (?x "を" ?y) (:head 3) (:x group-n) =>
  (((!y !x) (動短 1 2))
   (
    (("領収書") ("頂ける"))
     (("テニスコート") ("借りる"))
   )))
```

## 4.6 ヘッド (head)

原言語構造を決定する際、各構造において中心となる構成要素をヘッドと言い、その構造を上部の構造へ伝える重要な働きをする。指定するヘッドは、変項であっても固定項であっても構わないが、機能語はヘッドになり得ない。

〔例〕

```
"場所がわかりません"
TOP [(?X が ?Y) --- C]
 |--?X [(場所)]
 |
 |--?Y [(?X ません) --- PX]
      |--?X [(わかり)]
```

```
(define-regular-pattern auxv-masen :j-c px
  (?x "ません") (:head 1) =⟩
  ((("不" !x) (動短 (副 1) 2))
   (
    (("分かる"))
    )))
```

```
(define-pattern particle-ga-c :j-c c
  (?x "が" ?y) (:head 3) =⟩
  ((((!y !x) (動短 1 2))
   (
    (("場所") ("分かる"))
   )))
```

それぞれ、「分かりません」では「分かる」、「場所が分かる」では「分かる」がヘッドとなる。

日英、日独、日韓翻訳では、格を決定したり、デフォルトの主語を決定したり、或いは対訳を決定したりするために、機能語部分を含めた複数ヘッドの指定をすることがあるが、日中翻訳では現在使用していない。

## 4.7 ローカル辞書 (local dic)

2.4節「日中変換辞書」で定義された対訳はデフォルト値であり、変換知識のパタン中でローカルに別の対訳を与えることができる。例えば、「薬を飲む」の場合は、変換辞書の中で定義されている本動詞「飲む」の対訳"喝"ではなく、"吃"を用いたい。このような場合は、以下のようにパタンごとの訳し分けを行なう。このローカル辞書は、上部構造で記述されるものほど優先される。

〔書式〕

```
((VARIABLE ((POS REG-EXP) TARGET)))
```

- VARIABLE ... 変数
- POS ... 日本語品詞シンボル
- REG-EXP ... 日本語正規形または表層形
- TARGET ... 中国語品詞と対訳のリスト

〔例〕

1. 一般的なローカル辞書

```
(define-pattern particle-wo-c :j-c c
  (?x "を" ?y) (:head 3) (:x group-n) =>
  (((!y !x) (动短 1 2))
   (
    (("薬") ("飲む"))
    )
  ((y ((本動詞 "飲む")(动 "吃"))))
   ))
```

2. terminal として合成した語を変更するためのローカル辞書

```
(define-pattern particle-ga-n :j-c n
  (?x "が" ?y) (:head 3)  =>
  (((!y !x) (名短 1 2))
   (
    (("ベッド") ("N台"))
    (("友達") ("N人"))
    )
   ((y ((普通名詞 "N台")(num-dai ((!x "张")((量 2))))))
    (y ((普通名詞 "N人")(persons ((!x "个")((量 2))))))
   ))
```

TARGET の部分に terminal として合成した時のルール名と対訳、品詞を記述する。

## 4.8 生成情報

各構造ごとに目的言語パタンが選ばれ、形態素レベルでの変換処理が済んだ入力文は、生成処理に回されるが、その際、必要に応じて生成情報が付与される。生成処理では、その情報を用いて、文法的に正しく、かつ、より自然な中国語を出力するため、語の挿入、削除などを行なう。例えば、「お客様のお名前をお願いいたします」という入力文は、変換処理の後、以下のような形態素列データとして生成処理へ渡され、最終的に "请告诉我你的姓名。"という翻訳結果が得られる。

```
*** Generation ***
 ((动短（动 "请 "）（动 "告诉 "）（代 "我 "）（名短（名 "客人 "）（助 "的 "）（名短（代 "你 "）（助 "的
 "）（名 "姓名 "）)))))
 =⟩
 请告诉我你的姓名。
```

形態素列データの中には、中国語句型と品詞、対訳の文字列が記述されている。この文では、以下の２つの生成規則が使用されたことになる。

```
((客人 的 你 的) "-⟩" (你 的) ((:D 0 NIL)(:D 1 NIL)))
((名 後) "-⟩" (名 。 後) ((:I 0 "。")))
```

※2.5節「中国語生成規則」参照

〔中国語品詞一覧〕

1 名词
2 处所词（地方词）
3 方位词
4 时间词
5 数词
6 量词
7 代词
8 动词
9 助动词
10 判断词
11 形容词
12 副词
13 介词
14 连词
15 助词
16 语气词
17 拟声词
18 习惯用语（感叹词）
19 标点符号

〔中国語句型一覧〕

1 名词短语
2 动词短语
3 副词短语
4 介词短语
5 形容词短语
6 时间词短语
7 数词短语
8 数量词短语
9 代词短语
10 地点名短语
11 判断词短语
12 助动词短语
13 拟声词短语
14 方位词短语
15 子句
16 句子

※詳細は「The Part-of-speech and Parsing Rules for the Chinese Language」[宗 99] を参照のこと。

## 4.9 カテゴリ別変換パタン定義の基本方針とパタン例

※元にした解析変換知識は、2000年1月1日時点のものである。
※各パタン例の目的言語パタンおよび用例は、格納場所を参照のこと。

### 4.9.1 カテゴリ u(接続詞・感動詞・呼びかけなど、文全体を修飾する語 / 重文)

パタン種類：通常 define-pattern
ヘッド：後の文の述語の項
変項制限：通常なし。ただし変項 x に制限をつけることがある。

```
(define-pattern conjunction-dewa :j-c u
("では" ?x) (:head 2)
```
格納場所：/usr/local/TDMT/tdmt-multi-dev/transfer/j-c/p-data/conjunction.lisp
訓練文例："では確認させていただきます"

```
(define-pattern marker-cadv-u :j-c u
(?x ⟨cadv-⟩ ?y) (:head 3)
```
格納場所：p-data/marker-cadv.lisp(p-data までのパスは上に同じ。以下同)
訓練文例："また⟨cadv-⟩シャンペンとお花を御用意できますが"

```
(define-pattern marker-interj-u :j-c u
(?x ⟨i-⟩ ?y) (:head 3)
```
格納場所：p-data/marker-interj.lisp
訓練文例："ありがとうございます⟨i-⟩ニューヨークシティホテルでございます"

```
(define-pattern marker-nn-u1 :j-c u
(?x ⟨n-n⟩ ?y) (:head 3)
```
格納場所：p-data/marker-nn.lisp
訓練文例："田中様⟨n-n⟩いつのご予約ですか"

```
(define-regular-pattern conj-node :j-c u
(?x "ので" ?y) (:head 3) (:x group-n word-p p px c cx s sx group-a)
```
格納場所：p-data/particle-node.lisp
訓練文例："スケジュールが書いてあるので急いで見つけたいんですが"

```
(define-pattern conj-ba-u :j-c u
(?x "ば" ?y) (:head 3)
```
格納場所：p-data/particle-conj.lisp
訓練文例："フロントの方に来ていただければすぐ手続さしていただけます

```
(define-pattern marker-endp-u2 :j-c u
(?x ⟨endp-⟩ ?y) (:head 3) (:x group-n word-p p px c cx s sx group-a 感動詞)
```
格納場所：p-data/marker-endp.lisp

訓練文例："そうですか ⟨endp-⟩ では八時三十分にリムジンを手配いたします

(define-pattern marker-/renyou-u :j-c u
(?x ⟨/renyou-⟩ ?y) (:head 3) (:x group-n word-p p px c cx s sx group-a 感動詞)
格納場所：p-data/marker-renyou.lisp
訓練文例："はいニューヨークシティホテルをご利用いただき ⟨/renyou-⟩ ありがとうございました"

(define-pattern /kihon-u :j-c u
(?x ⟨/kihon-⟩ ?y) (:head 3)
格納場所：p-data/marker-kihon.lisp
訓練文例："はい大丈夫です ⟨/kihon-⟩ こちらで誰が運んだかを調べますので

### 4.9.2　カテゴリ sx (終助詞・文末表現)

パタン種類：define-regular-pattern または define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-regular-pattern auxv-desu-sx :j-c sx
(?x "です") (:head 1)
格納場所：p-data/auxv-desu.lisp
訓練文例："入場料はいくらですか"

(define-pattern ending-ka :j-c sx
(?x "か") (:head 1)
格納場所：p-data/particle-final.lisp
訓練文例："入場料はいくらですか"

(define-regular-pattern ending-kedo :j-c sx
(?x "けど") (:head 1)
格納場所：p-data/particle-final.lisp
訓練文例："できれば日本語の話せるお医者さん呼んでいただきたいんですけど"

(define-regular-pattern ending-ne :j-c sx
(?x "ね") (:head 1)
格納場所：p-data/particle-final.lisp
訓練文例："はいではパンプキンポタージュを二名様でございますね"

(define-pattern tense-ta-sx :j-c sx
(?x "た") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："はい昨日電話で予約しましたた"

(define-pattern tekudasai-sx :j-c sx
(?x "てください") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："このフィルムを現像してください"

### 4.9.3 カテゴリ s (主語、時間表現を含む文)

パタン種類：通常 define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-pattern particle-wa-s :j-c s
(?x "は" ?y) (:head 3)
格納場所：p-data/particle-wa.lisp
訓練文例："私は菜食主義者なんですがメニューはありますか"

(define-pattern particle-ni-s :j-c s
(?x "に" ?y) (:head 3)
格納場所：p-data/particle-ni.lisp
訓練文例："七時にテーブルは空いていますか"

(define-pattern particle-dewa-s :j-c s
(?x "では" ?y) (:head 3)
格納場所：p-data/particle-wa.lisp
訓練文例："当ホテルでは一万七千円のお部屋からのご利用となっております"

(define-pattern marker-nn-s :j-c s
(?x ⟨n-n⟩ ?y) (:head 3)
格納場所：p-data/marker-nn.lisp
訓練文例："今⟨n-n⟩どちらが勝っているのですか"

(define-pattern marker-adv-s :j-c s
(?x ⟨adv-⟩ ?y) (:head 3)
格納場所：p-data/marker-adv.lisp
訓練文例："料理はまだ⟨adv-⟩予約していません"

(define-pattern particle-desuga-s :j-c s
(?x "です" "が" ?y) (:head 4)
格納場所：p-data/particle-ga.lisp
訓練文例："シングルルームのお値段ですが一泊九十ドルとなっております"

(define-pattern particle-no-ken-de-s :j-c s

(?x "の" "件" "で" ?y) (:head 5)
格納場所：p-data/particle-de.lisp
訓練文例："送迎サービス の件で 二三教えていただけますか"

### 4.9.4 カテゴリ cx (c +助動詞類)

パタン種類：define-regular-pattern または define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-pattern auxv-masu :j-c cx
(?x "ます") (:head 1)
格納場所：p-data/auxv-masu.lisp
訓練文例："では窓際の席が空いたら移らせてもらえ ます か"

(define-regular-pattern modal-teiru-reg-cx :j-c cx
(?x "ている") (:head 1)
格納場所：p-data/auxv-te.lisp
訓練文例："深夜遅くまでやっ ている カラオケ屋はありますか"

(define-regular-pattern modal-teitadaku-reg :j-c cx
(?x "ていただく") (:head 1)
格納場所：p-data/auxv-te.lisp
訓練文例："いろいろと教え ていただい てありがとう"

(define-regular-pattern modal-dekiru-reg-cx :j-c cx
(?x "できる") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："旅をしているお客さんとおんなじようにいろんな施設利用 できる わけですね"

(define-regular-pattern koto-ga-dekiru-reg-cx :j-c cx
(?x ⟨⟩ "こと" "が" "できる") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："ちょうど船旅のような気分を味わっていただく ⟨⟩ ことができ ます"

### 4.9.5 カテゴリ c (主語、時間表現以外の格関係)

パタン種類：通常 define-regular-pattern または define-pattern
ヘッド：述語の項
変項制限：通常なし。(:x group-n) をつけることがある。

(define-pattern particle-wo-c :j-c c
(?x "を" ?y) (:head 3) (:x group-n)
格納場所：p-data/particle-wo.lisp
訓練文例：" 申し訳ありませんが予約を キャンセルしたいです "

(define-regular-pattern particle-temo-c :j-c c
(?x "ても" ?y) (:head 3)
格納場所：p-data/particle-mo.lisp
訓練文例："遅くなっても 構いません "

(define-regular-pattern particle-de-c :j-c c
(?x "で" ?y) (:head 3)
格納場所：p-data/particle-de.lisp
訓練文例："そのかわりにシアターバーの方で ビデオをお楽しみいただけます "

(define-pattern marker-/renyou-c :j-c c
(?x ⟨/renyou-⟩ ?y) (:head 3)
格納場所：p-data/marker-renyou.lisp
訓練文例："ほかにもう少し早く ⟨renyou-⟩ 着く方法はありませんか "

(define-pattern marker-an-c :j-c c
(?x ⟨an-⟩ ?y) (:head 3)
格納場所：p-data/marker-an.lisp
訓練文例："七十六丁目通りをまっすぐ ⟨an-⟩ セントラルパークへ向かって歩いてください "

(define-pattern marker-adv-c :j-c c
(?x ⟨adv-⟩ ?y) (:head 3)
格納場所：p-data/marker-adv.lisp
訓練文例："わたしも妻もずっと ⟨adv-⟩ 京都にあこがれてまして今回新婚旅行で行くことになりました
"

### 4.9.6 カテゴリ px (p +助動詞類)

パタン種類：define-regular-pattern または define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-regular-pattern modal-itasu-reg :j-c px
(?x "いたす") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："すぐに手配 いたし ますのでしばらくお待ちください "

(define-regular-pattern modal-temorau-reg :j-c px
(?x "てもらう") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："有料テレビから一般放送へスイッチを換えてもらわないといけないんです"

(define-regular-pattern modal-temiru-reg :j-c px
(?x "てみる") (:head 1)
格納場所：p-data/auxv-te.lisp
訓練文例："その日の状況を調べてみることができますけれども"

(define-regular-pattern teoku-reg-px :j-c px
(?x "ておく") (:head 1)
格納場所：p-data/auxv-te.lisp
訓練文例："ではお客様のご予約のほうにメモを付けておきますので"

(define-regular-pattern modal-seru-reg :j-c px
(?x "せる") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："お手伝いさせていただきます"

(define-regular-pattern modal-reru-reg :j-c px
(?x "れる") (:head 1)
格納場所：p-data/auxv.lisp
訓練文例："どちらのほうで結婚式がとり行なわれますでしょうか"

### 4.9.7　カテゴリ p (合成用言)

パタン種類：通常 define-regular-pattern
ヘッド：述語の項
変項制限：通常なし

(define-regular-pattern o-itasu-reg :j-c p
("お" ?x "いたす") (:head 2)
格納場所：p-data/prefix-verb.lisp
訓練文例："ですのでリムジンの方をお勧めいたします"

(define-regular-pattern go-ninareru :j-c p
("ご" ?x "になれる") (:head 2)
格納場所：p-data/prefix-verb.lisp
訓練文例："お一人二十五ドルでプールジャグジーサウナを御利用になれます"

(define-pattern /renyou-aru :j-c p

(?x ⟨/renyou-⟩ "ある" ) (:head 1)
格納場所：p-data/marker-renyou.lisp
訓練文例："結構いろいろと細かく ⟨/renyou-⟩ ある わけですねよく分かりました"

### 4.9.8 カテゴリ a (副詞+形容語)

パタン種類：通常 define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-pattern marker-/renyou-a :j-c a
(?x ⟨/renyou-⟩ ?y) (:head 3)
格納場所：p-data/marker-renyou.lisp
訓練文例："その日ちょっと仕事がすごく ⟨/renyou-⟩ 忙しいので四時半はちょっと無理だと思うんだけれども"

(define-pattern marker-an-jyubun-a :j-c a
("十分" ⟨an-⟩ ?x) (:head 3)
格納場所：p-data/marker-an.lisp
訓練文例："そのままでも 十分 ⟨an-⟩ 広いと思いますが"

### 4.9.9 カテゴリ na (形容詞+名詞類)

パタン種類：通常 define-pattern
ヘッド：述語の項
変項制限：通常なし

(define-pattern /kihon-na :j-c na
(?x ⟨/kihon-⟩ ?y) (:head 3)
格納場所：p-data/marker-kihon.lisp
訓練文例：""非常に広い ⟨/kihon-⟩ お部屋で豪華なお部屋でございましてリビングと広いベッドルームがございます"

### 4.9.10 カテゴリ nnx (nn +接辞)

パタン種類：通常 define-pattern
ヘッド：係り先の項
変項制限：通常なし

(define-pattern particle-made-nnx :j-c nnx
(?x "まで" "の" ?y) (:head 4)
格納場所：p-data/particle-fromto.lisp
訓練文例："できましたら二階か三階までの部屋にしていただきたいんですけど"

(define-regular-pattern deixis-ano-nnx :j-c nnx
("あの" ?x) (:head 2)
格納場所：p-data/determiner.lisp
訓練文例："あの ピアノ奏者はうまいですね"

### 4.9.11　カテゴリ nn (助詞またはマーカを伴う名詞句)

パタン種類：通常 define-pattern
ヘッド：係り先の項
変項制限：通常なし。(:y word-n n nx na nn) をつけることがある。

(define-pattern heijo-to-nn :j-c nn
(?x "と" ?y) (:head 3)
格納場所：p-data/particle-and.lisp
訓練文例："ではお名前 と クレジットカードの番号をお願いいたします"

(define-pattern particle-no-nn :j-c nn
(?x "の" ?y) (:head 3) (:y word-n n nx na)
格納場所：p-data/particle-no.lisp
訓練文例："ええ実は家族と一緒に白浜 の リゾートマンションに泊まってるんです"

(define-pattern particle-de-no-nn :j-c nn
(?x "で" "の" ?y) (:head 4)
格納場所：p-data/particle-no.lisp
訓練文例："現時点 での ご予約は十三日から十七日となっております"

(define-pattern marker-nn-nn :j-c nn
(?x ⟨n-n⟩ ?y) (:head 3) (:y word-n n nx na nn)
格納場所：p-data/marker-nn.lisp
訓練文例："ではお部屋の空き ⟨n-n⟩ 状況を確認いたしますので少々お待ちください"

(define-pattern markar-n-n-xyz-nn :j-c nn
(?x ⟨n-n⟩ ?y ⟨n-n⟩ ?z) (:head 5)
格納場所：p-data/xyz-pattern.lisp
訓練文例："お一人二十五ドルでプール ⟨n-n⟩ ジャグジー ⟨n-n⟩ サウナを御利用になれます"

### 4.9.12 カテゴリ nx (接頭辞・連体詞・接尾辞・副助詞など名詞に前接或いは後接する語)

パタン種類：define-regular-pattern または define-pattern
ヘッド：係り先の項
変項制限：通常なし

(define-pattern yaku-nx :j-c nx
("約" ?x) (:head 2)
格納場所：p-data/suffix.lisp
訓練文例："はい当方の施設では 約 四十種類のアクティビティーを楽しんでいただけます"

(define-regular-pattern deixis-douitta :j-c nx
("どういった" ?x) (:head 2)
格納場所：p-data/determiner.lisp
訓練文例："失礼ですが どういった 雑貨品をお探しでしょうか"

(define-regular-pattern suffix-toiuno-nx :j-c nx
(?x "というの") (:head 1)
格納場所：p-data/suffix.lisp
訓練文例："ええ保険には日本で旅行者用 っていうん ですかその保険に入ってきました"

(define-regular-pattern suffix-sama :j-c nx
(?x "様") (:head 1)
格納場所：p-data/suffix.lisp
訓練文例："鈴木 様 ですねご用件を承ります"

(define-pattern marker-n-n-gimi :j-c nx
(?x ⟨n-n⟩ "気味") (:head 1)
格納場所：p-data/marker-nn.lisp
訓練文例："ここ二三日ずっと風邪 ⟨n-n⟩ 気味 だったもんですから"

### 4.9.13 カテゴリ n (複合名詞)

パタン種類：通常 define-pattern
ヘッド：係り先の項 (例外あり)
変項制限：通常なし

(define-pattern marker-nn-n :j-c n
(?x ⟨n-n⟩ ?y) (:head 3)
格納場所：p-data/marker-nn.lisp
訓練文例："鈴木 ⟨n-n⟩ 直子です"

(define-pattern marker-nn-n2 :j-c n

(?x ⟨n-n⟩ ?y) (:head 1)
格納場所：p-data/marker-nn.lisp
訓練文例："お客様のお名前がメアリ ⟨n-n⟩ フィリップス様"
※性別を区別するため、ファーストネームをヘッドとする。

(define-pattern particle-no-n :j-c n
(?x "の" ?y) (:head 3)
格納場所：p-data/particle-no.lisp
訓練文例："九月三日 の 土曜日とその次の日もお願いしたいんですが"

(define-pattern marker-sn-n :j-c n
(?x ⟨sn-n⟩ ?y) (:head 3)
格納場所：p-data/marker-sn.lisp
訓練文例："ホテルの担当 ⟨sn-n⟩ スタッフが同席するようにいたしますのでご心配いりません"

(define-pattern marker-roman-roman :j-c n
(?x ⟨r-r⟩ ?y) (:head 3) (:x ローマ字)
格納場所：p-data/marker-roman.lisp
訓練文例："はいティー ⟨r-r⟩ エー ⟨r-r⟩ エヌ ⟨r-r⟩ エー ⟨r-r⟩ ケイ ⟨r-r⟩ エーです"

### 4.9.14　カテゴリ terminal (数詞を伴う頻出表現 / 派生形)

※解析で terminal rule 用の形態素合成を行なうこと
パタン種類：通常 define-pattern
ヘッド：通常なし
変項制限：通常なし

(define-pattern n-oclock :j-c terminal
(?x "時")
格納場所：p-data/suffix-counter.lisp
訓練文例："たぶん六 時 ぐらいになると思います"

(define-pattern adj-me :j-c terminal
(?x "め")
格納場所：p-data/suffix.lisp
訓練文例："リビングの奥には広 め のテラスがございます"

# 第 5 章

# 解析変換知識の作成

本節では、解析変換知識の作成手順を、意味距離計算を使って訓練する場合の方法で、例文を用いて解説する。また、作成時に使用する補助ツール類についても説明を行なう。さらに、2001年1月1日現在の解析変換知識の規模について、その統計値を本節最後に添付する。

システムのデフォルトは、2000年1月1日時点で、以下のような値になっている。

- 訓練の単位 ... 発声単位
- 意味距離計算 ... ON
- 新部分翻訳 ... ON

新部分翻訳がONになっているため、翻訳訓練に際しては、exact match用例（入力文と完全に一致する用例）をすべて追加し、望ましい翻訳結果が得られた時点で、距離が0になっていることを訓練終了の条件として確認するものとする。

作業にあたってはTDMT翻訳知識作成ツールを利用するため、前もって「TDMT翻訳知識作成ツールの利用方法」[河井94]を参照のこと。

## 5.1　作成手順の例

意味距離計算を行なう場合の解析変換知識作成手順を説明する。作業は以下の流れで行ない、既に訓練済みの知識を利用できる場合は、省略して次のステップに進む。

1. 前準備
2. 翻訳実行
3. 形態素解析結果の確認
4. 形態素情報の修正 ( lexical-transformation の作成 )
5. 挿入マーカの確認 ( local-transformation の確認 )
6. 日中変換辞書の作成
7. 変換パタンの新規作成 および 目的言語パタンと用例の追加
8. 中国語生成規則の作成
9. 訓練終了

但し、実際にはこの説明通りの順序とならない場合もある。特に、「7．変換パタンの新規作成 および目的言語パタンと用例の追加」の段階では、出力させたい訳文を念頭に置きつつ、形態素情報の確認や部分的な翻訳、構文解析の確認を繰り返し行ないながら、p-data の整備を進めていくことになる。時には前の段階（a-data）に戻って調整することも必要になる。

### 5.1.1 例１："十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります"

1．前準備

(a) allegro common lisp の起動

(b) システムの起動ならびにルールとタギングデータのロード。
setup-jc-tag.lispをロードすることにより、まとめて処理できる。

(c) パッケージを翻訳モードに切替え

(d) TDMT 翻訳知識作成ツールの起動

2．翻訳実行

TDMT ツールのメニューで翻訳を実行する。
意味距離計算を ON にしているため、望ましい翻訳ではないが出力結果が得られる。

*** Input ***
十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります
*** Output ***
十八号在单人房有变化房间，一百美元的费用。（5.0555906）

3．形態素解析結果の確認

TDMT ツールのメニューで、Morph Analysis Only を選択すると以下のように解析知識適用前の形態素情報が表示される。

sentence: 十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります

| WORD | REG-EXP | POS | CONJ-FORM | PRON | SEM-CODES |
|---|---|---|---|---|---|
| 18 | ００ | 数詞 | --- | --- | 120 |
| 日 | 日 | 普通名詞 | --- | --- | --- |
| は | は | 係助詞 | --- | --- | --- |
| シングル | シングル | 普通名詞 | --- | --- | --- |
| に | に | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 部屋 | 部屋 | 普通名詞 | --- | --- | --- |
| を | を | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 替わり | 替わる | 本動詞 | 連用 | --- | --- |
| いただき | いただく | 補助動詞 | 連用 | --- | --- |
| まし | ます | 助動詞 | た | --- | --- |
| て | て | 接続助詞 | --- | --- | --- |
| 料金 | 料金 | 普通名詞 | --- | --- | --- |
| が | が | 格助詞 | --- | --- | --- |
| 100 | ００ | 数詞 | --- | --- | 120 |
| ドル | ドル | 普通名詞 | --- | --- | --- |
| と | と | 格助詞 | --- | --- | --- |
| なり | なる | 本動詞 | 連用 | --- | --- |
| ます | ます | 助動詞 | 基本 | --- | --- |

次に Morph Analysis を選択して比較し、適用されている解析知識を確認する。

sentence:　十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります

| WORD | REG-EXP | POS | CONJ-FORM | PRON | SEM-CODES |
|---|---|---|---|---|---|
| 18 日 | N日 | 普通名詞 | --- | --- | 19 153 828 |
| は | は | 係助詞 | --- | --- | --- |
| シングル | シングル | 普通名詞 | --- | --- | 128 428 941 |
| に | に | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 部屋 | 部屋 | 普通名詞 | --- | --- | 941 |
| を | を | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 替わり | 替わる | 本動詞 | 連用 | --- | 376 |
| いただき | いただく | 補助動詞 | 連用 | --- | --- |
| まし | ます | 助動詞 | た | --- | --- |
| て | て | 接続助詞 | --- | --- | --- |
| 料金 | 料金 | 普通名詞 | --- | --- | 748 |
| が | が | 格助詞 | --- | --- | --- |
| 100 | 0 0 | 数詞 | --- | --- | 120 |
| ドル | ドル | 普通名詞 | --- | --- | 828 |
| となる | となる | 助動詞 | 連用 | --- | --- |
| ます | ます | 助動詞 | 基本 | --- | --- |

この時、Output バッファには、a-data での処理の結果が以下のように表示され、Morph Analysis と併せて、解析処理の結果を確認することができる。

```
*** Distance calculation in analysis ***
LEXICAL-TRANSFORMATION ...
    0.000000 :: (COMPOUND-TO-NARU) 2 : (と なり) =⟩ (となる)
    0.000000 :: (NUM-DATE) 6 : (18 日) =⟩ (18 日)
TOTAL DISTANCE =  0.000000
```

ここでは、既に「N日」が terminal rule 用に合成され、また、「と＋なる」が形態素合成されていることがわかる。さらに、「100 ドル」を、正規形「N円」として、terminal rule 用に合成することにする。

### 4. lexical-transformationの作成

方針に従って、以下のルールを作成する。

```
(define-lexical-transformation num-dollers :j-c 5
 ((((:pos . 数詞)) ((:word . "ドル"))) =⟩
  (lex ((1 2) (:pos . 普通名詞) (:reg-exp . "N円") (:compound 1 2)))
  ))
```

作成したルールをコンパイルし、再度翻訳を実行して形態素情報を確認すると、以下のように表示される。

sentence: 十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります

| WORD | REG-EXP | POS | CONJ-FORM | PRON | SEM-CODES |
|---|---|---|---|---|---|
| 18日 | N日 | 普通名詞 | --- | --- | 19 153 828 |
| は | は | 係助詞 | --- | --- | --- |
| シングル | シングル | 普通名詞 | --- | --- | 128 428 941 |
| に | に | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 部屋 | 部屋 | 普通名詞 | --- | --- | 941 |
| を | を | 格助詞 | --- | --- | --- |
| お | お | 接頭辞 | --- | --- | --- |
| 替わり | 替わる | 本動詞 | 連用 | --- | 376 |
| いただき | いただく | 補助動詞 | 連用 | --- | --- |
| まし | ます | 助動詞 | た | --- | --- |
| て | て | 接続助詞 | --- | --- | --- |
| 料金 | 料金 | 普通名詞 | --- | --- | 748 |
| が | が | 格助詞 | --- | --- | --- |
| 100ドル | N円 | 普通名詞 | --- | --- | 121 828 |
| となる | となる | 助動詞 | 連用 | --- | --- |
| ます | ます | 助動詞 | 基本 | --- | --- |

**5．挿入マーカの確認**

この文の場合は、助詞、助動詞類が内容語の境界を示すので、マーカの挿入は行なわれていない。

**6．日中変換辞書の作成**

この文で対訳が必要な語は、「十八日」「シングル」「部屋」「替わる」「料金」「百ドル」。
部分的に翻訳を行なうと、

```
*** Input ***
（18日）
*** Output ***
十八号。（0.0）

*** Input ***
（シングル）
*** Output ***
单人房。（0.0）

*** Input ***
（部屋）
*** Output ***
房间。（0.0）

*** Input ***
（替わり）
*** Output ***
有变化。（0.0）

*** Input ***
（料金）
*** Output ***
费用。（0.0）

*** Input ***
（100ドル）
*** Output ***
（0.0）
```

となり、「100ドル」の対訳が不足していることがわかるが、この語はterminal rule用に合成しているので、変換パタンを作ることにする。

また、「部屋を替わる」の場合の「替わる」の対訳は、"搬"にしたいが、これは変換パタンにローカル辞書で追加することにする。

### 7. 変換パタンの新規作成および目的言語パタンと用例の追加

部分的な翻訳を行ない、翻訳結果が得られない場合は翻訳メニューの Search Rule Name で変換パタンを検索する。該当する変換パタンが存在しない場合は、変換パタンを新規作成し、存在する場合は、正しい翻訳結果が得られるように必要に応じて目的言語パタンを追加し、不足用例を補う。

先に terminal rule 用に合成した「100 ドル」の変換パタンを以下のように作成する。

```
(define-pattern n-dollers :j-c terminal
  (?x "ドル") (:head 2) =⟩
  (((!x "美元") (量 2))
   (("0 0 "))))
```

再度部分的に翻訳すると、以下のように対訳が得られる。

```
*** Input ***
 (100 ドル)
*** Output ***
 一百美元。(0.0)
```

部分的な翻訳は、各形態素の係り受けを考えながら取り得る下部構造の数が少ない構造から行なう。この文の場合、以下の順となる。

(a) (お 部屋)

(b) (お 替わり いただき)

(c) (お 替わり いただき まし)

(d) (お 部屋 を お 替わり いただき まし)

(e) (シングル に お 部屋 を お 替わり いただき まし)

(f) (十八日 は シングル に お 部屋 を お 替わり いただき まし)

(g) (百ドル となり)

(h) (百ドル となり ます)

(i) (料金 が 百ドル となり ます)

(j) (十八日 は シングル に お 部屋 を お 替わり いただき まし て 料金 が 百ドル となり ます)

```
*** Input ***
 (お) (部屋)
*** Output ***
```

房间。（0.0）

*** Input ***
（お）（替わり）（いただき）
*** Output ***
有变化。（0.6666667）

*** Input ***
（お）（替わり）（いただき）（まし）
*** Output ***
有变化。（0.6666667）

*** Input ***
（お）（部屋）（を）（お）（替わり）（いただき）（まし）
*** Output ***
有变化房间。（0.6666667）

*** Input ***
（シングル）（に）（お）（部屋）（を）（お）（替わり）（いただき）（まし）
*** Output ***
在单人房有变化房间。（1.6666667）

*** Input ***
（18 日）（は）（シングル）（に）（お）（部屋）（を）（お）（替わり）（いただき）（まし）
*** Output ***
十八号在单人房有变化房间。（2.3333335）

*** Input ***
（100 ドル）（となる）
*** Output ***
一百美元。（0.4666707）

*** Input ***
（100 ドル）（となる）（ます）
*** Output ***
一百美元。（0.4666707）

*** Input ***
（料金）（が）（100 ドル）（となる）（ます）
*** Output ***
一百美元的费用。（0.4666807）

*** Input ***

（18 日）（は）（シングル）（に）（お）（部屋）（を）（お）（替わり）（いただき）（まし）（て）（料金）（が）（100 ドル）（となる）（ます）

```
*** Output ***
 十八号在单人房有变化房间，一百美元的费用。（3.3555698）
```

結果、("お" ?x "いただく")、(?x "に" ?y)、(?x "は" ?y)、(?x "となる")、(?x "が" ?y)、(?x "て" ?y) の変換パタンと用例が不足していることがわかる。ここで、変換パタンの新規作成と用例の追加を行なう。

```
(define-regular-pattern o-itadaku-reg :j-c p
  ("お" ?x "いただく") (:head 2)  =⟩
     :
  ((("需要" !x) (动短 (助动 1) 2))      ←目的言語パタンの追加
   (
    (("替わる"))                          ←用例の追加
   ))
     :

(define-pattern particle-ni-c :j-c c
  (?x "に" ?y) (:head 3)  =⟩
     :
  (((!y "到" !x) (动短 1 (介 2) 3))     ←目的言語パタンの追加
   (
    (("シングル") ("替わる"))           ←用例の追加
    )
   ((y ((本動詞 "替わる") (动 "搬"))))  ←ローカル辞書の追加
   )
     :

(define-pattern particle-wa-s :j-c s
  (?x "は" ?y) (:head 3)   =⟩
  (((!x !y) (子句 1 2))
   (
    (("N日") ("替わる"))                ←用例の追加
   ))
     :

(define-regular-pattern modal-tonaru-reg :j-c p
  (?x "となる") (:head 1) =⟩
   (((!x))
   (
    (("N円"))                           ←用例の追加
   ))
     :
```

```
(define-pattern particle-ga-s :j-c s
  (?x "が" ?y) (:head 3) =⟩
     :
  (((!x "为" !y) (子句 1 (判 2) 3))     ←目的言語パタンの追加
   (
    (("料金") ("N円."))                  ←用例の追加
   ))
     :

(define-pattern conj-te-u :j-c u
  (?x "て" ?y) (:head 3) =⟩
  (((!x ", " !y) (句子 1 (符 2) 3))
   (
    (("替わる") ("N円"))                 ←用例の追加
   ))
     :
```

不足している変換パタンを作成し、かつ、すべての不足用例を補うと、以下のように全文の翻訳結果が得られる。「部屋」の対訳"房间"は余剰となるが、後ほど生成の段階で削除することとし、これを生成処理前の最終結果とする。

```
*** Input ***
 十八日はシングルにお部屋をお替わりいただきまして料金が百ドルとなります
*** Output ***
 十八号需要搬房间到单人房，费用为一百美元。（0.0）
```

**8．中国語生成規則の作成**

中国語として、"搬房间到单人房"という表現はあり得ない。そこで、以下の生成規則を追加する。

```
((搬 房间 到 单人房) "-⟩" (搬 到 单人房) ((:D 1 NIL)))
```

再度翻訳を実行すると、翻訳結果は以下のようになる。

```
*** Generation ***
 ((句子（子句（数 "18"）（名 "号"）（动短（动短（动短（助动 "需要"）（动 "搬"））（名 "房
间"））（介 "到"）（名 "单人房"））)（符 ", "）（子句（名 "费用"）（判 "为"）（数 "100"）（量 "美
元"））））
 =⟩
 十八号需要搬到单人房，费用为一百美元。
```

**9．訓練終了**

TDMT ツールの画面では、上の Output を得るために、以下の変換パタンが適用されたことが表示される。

```
*** Distance calculation in Transfer ***
    0.000000 :: (CONJ-TE-U) U :(?X て ?Y) =⟩ (!X ， !Y) .... ((替わる) (N円))
    0.000000 :: (PARTICLE-WA-S) S :(?X は ?Y) =⟩ (!X !Y) .... ((N日) (替わる))
    0.000000 :: (N-DAY) TERMINAL :(?X 日) =⟩ (!X 号) .... ((0 0))
    0.000000 :: (~DIC) : (数詞 0 0) =⟩ ((数 18))
    0.000000 :: (PARTICLE-NI-C) C :(?X に ?Y) =⟩ (!Y 到 !X) .... ((シングル) (替
わる))
    0.000000 :: (~DIC) : (普通名詞 シングル) =⟩ ((名 単人房))
    0.000000 :: (PARTICLE-WO-C) C :(?X を ?Y) =⟩ (!Y !X) .... ((部屋) (替わる))
    0.000000 :: (PREFIX-O) N :(お ?X) =⟩ (!X) .... ((部屋))
    0.000000 :: (~DIC) : (普通名詞 部屋) =⟩ ((名 房间))
    0.000000 :: (AUXV-MASHI) PX :(?X まし) =⟩ (!X) .... ((替わる))
    0.000000 :: (O-ITADAKU-REG) P :(お ?X いただく) =⟩ (需要 !X) .... ((替わる))
    0.000000 :: (~LOCAL) : (本動詞 替わる) =⟩ ((动 搬))
    0.000000 :: (PARTICLE-GA-S) S :(?X が ?Y) =⟩ (!X 为 !Y) .... ((料金) (N円))
    0.000000 :: (~DIC) : (普通名詞 料金) =⟩ ((名 费用))
    0.000000 :: (AUXV-MASU) CX :(?X ます) =⟩ (!X) .... ((N円))
    0.000000 :: (MODAL-TONARU-REG) P :(?X となる) =⟩ (!X) .... ((N円))
    0.000000 :: (N-DOLLERS) TERMINAL :(?X ドル) =⟩ (!X 美元) .... ((0 0))
    0.000000 :: (~DIC) : (数詞 0 0) =⟩ ((数 100))
TOTAL DISTANCE =  0.000000

*** Distance calculation in analysis ***
LEXICAL-TRANSFORMATION ...
    0.000000 :: (COMPOUND-TO-NARU) 2 : (と なり) =⟩ (となる)
    0.000000 :: (NUM-DOLLERS) 5 : (100 ドル) =⟩ (100 ドル)
    0.000000 :: (NUM-DATE) 6 : (18 日) =⟩ (18 日)
TOTAL DISTANCE =  0.000000
```

・構文解析の確認

原言語構文解析結果を確認する。この構文解析の図は、訓練途中においても随時部分的に出力させ、不足している知識を検討したり、係り受けの正しさを確認するのに活用するとよい。

```
TOP [(?X て ?Y) --- U]
 |--?X [(?X は ?Y) --- S]
 |   |--?X [(18 日) --- compound]
 |   |
 |   |--?Y [(?X に ?Y) --- C]
 |       |--?X [(シングル)]
 |       |
 |       |--?Y [(?X を ?Y) --- C]
 |           |--?X [(お ?X) --- N]
 |           |   |--?X [(部屋)]
```

```
|           |
|           |--?Y [(?X まし) --- PX]
|               |--?X [(お ?X いただく) --- P]
|                   |--?X [(替わり)]
|
|--?Y [(?X が ?Y) --- S]
    |--?X [(料金)]
    |
    |--?Y [(?X ます) --- CX]
        |--?X [(?X となる) --- P]
            |--?X [(100 ドル) --- compound]
```

翻訳結果が正しく出力されていること、構文解析が正しいこと、意味距離が 0 になっていること、訳文が複数出ていないことを確認し、訓練終了とする。

### 5.1.2　例２：”説明書ですか見当たらなかった気がするんですけれども”

1．前準備

例１に同じ

2．翻訳実行

TDMT ツールのメニューで翻訳を実行する。望ましい翻訳ではないが出力結果が得られる。

```
*** Input ***
 説明書ですか見当たらなかった気がするんですけれども
*** Output ***
 说明书? 没有找到了，气，干。 (5.3214316)
```

3．形態素解析結果の確認

TDMT ツールのメニューで、Morph Analysis Only を選択すると以下のように解析知識適用前の形態素情報が表示される。

sentence:　説明書ですか見当たらなかった気がするんですけれども

| WORD | REG-EXP | POS | CONJ-FORM | PRON | SEM-CODES |
|---|---|---|---|---|---|
| 説明書 | 説明書 | 普通名詞 | --- | --- | --- |
| です | です | 判定詞 | 基本 | --- | --- |
| か | か | 終助詞 | --- | --- | --- |
| 見当たら | 見当たる | 本動詞 | ない | --- | --- |
| なかっ | ない | 助動詞 | た | --- | --- |
| た | た | 助動詞 | 基本 | --- | --- |
| 気 | 気 | 普通名詞 | --- | --- | --- |
| が | が | 格助詞 | --- | --- | --- |
| する | する | 本動詞 | 基本 | --- | --- |
| ん | ん | 準体助詞 | --- | --- | --- |
| です | です | 判定詞 | 基本 | --- | --- |
| けれども | けれども | 接続助詞 | --- | --- | --- |

次に Morph Analysis を選択して比較し、適用されている解析知識を確認する。

sentence:　説明書ですか見当たらなかった気がするんですけれども

| WORD | REG-EXP | POS | CONJ-FORM | PRON | SEM-CODES |
|---|---|---|---|---|---|
| 説明書 | 説明書 | 普通名詞 | --- | --- | 845 |
| です | です | 判定詞 | 基本 | --- | --- |
| か | か | 終助詞 | --- | --- | --- |
| 見当たら | 見当たる | 本動詞 | ない | --- | 426 |
| なかっ | ない | 助動詞 | た | --- | --- |
| た | た | 助動詞 | 基本 | --- | --- |
| 気 | 気 | 普通名詞 | --- | --- | 160 834 |
| が | が | 格助詞 | --- | --- | --- |
| する | する | 本動詞 | 基本 | --- | 250 361 400 |
| のです | のです | 助動詞 | 基本 | --- | --- |
| けれども | けれども | 接続助詞 | --- | --- | --- |

ここでは、準体助詞「ん」が「の」に修正され、さらに、「の＋です」が形態素合成されている。

**4．lexical-transformationの作成**

さらなる形態素情報の修正は必要がないので、行なわない。

**5．挿入マーカの確認**

以下のように形態素の境界を示すマーカが挿入されている。

```
LOCAL-TRANSFORMATION ...
(説明書 です か ⟨ENDP-⟩ 見当たら なかっ た ⟨/KIHON-⟩ 気 が する のです けれども)
```

**6．日中変換辞書の作成**

この文で対訳が必要な語は、「説明書」「見当たる」。
他に、「気」「する」も内容語であるが、対訳の都合上、変換パタンに組み込むことにする。
部分的に翻訳を行なうと、

```
*** Input ***                    *** Input ***
(説明書)                         (見当たら)
*** Output ***                   *** Output ***
 说明书。 (0.0)                   找到。 (0.0)
```

となる。この文では、「見当たる」の対訳を "看见 "としたいが、これは変換パタンにローカル辞書で追加することにする。

**7．変換パタンの新規作成 および 目的言語パタンと用例の追加**

・構文解析の確認

原言語構文解析結果を確認し、不足している知識を検討する。

```
TOP [(?X ⟨ENDP-⟩ ?Y) --- U]
 |--?X [(?X か) --- SX]
 |    |--?X [(?X です) --- SX]
 |         |--?X [(説明書)]
 |
 |--?Y [(?X ⟨/KIHON-⟩ ?Y) --- U]
      |--?X [(?X た) --- SX]
      |    |--?X [(?X ない) --- CX]
      |         |--?X [(見当たら)]
      |
      |--?Y [(?X が ?Y) --- U]
           |--?X [(気)]
           |
           |--?Y [(?X けれども) --- SX]
```

```
        |--?X [(?X のです) --- SX]
            |--?X [(する)]
```

(?X た) がカテゴリ SX となっているため、その外側の構造に、カテゴリ U のパタンが選択されている。これでは、「見当たら なかっ た」が「...⟨/KIHON-⟩ 気 が する」の下部構造になり得ない。そこで、「見当たら なかっ た」をカテゴリ CX に、「...⟨/KIHON-⟩ 気 が する」をカテゴリ S にする方針で以下の変換パタンを作成する。

```
(define-pattern nakatta-cx :j-c cx    ←変換パタンの新規作成
  (?x "なかっ" "た") (:head 1) =⟩
  ((("没有" !x) (动短 (动 1) 2))       ←用例の登録
   (
    (("见当たる"))
   )))

(define-pattern /kihon-ki-ga-suru :j-c s    ←変換パタンの新規作成
  (?x ⟨/kihon-⟩ "気" "が" "する") (:head 1) =⟩
  ((("好像" !x) (动短 (动 1) 2))
   (
    (("见当たる"))                          ←用例の登録
    )
   ((x ((本動詞 "見当たる")(动 "看见"))))    ←ローカル辞書の登録
   ))
```

再度翻訳すると、以下のように正しい構文解析が得られる。

```
TOP [(?X ⟨ENDP-⟩ ?Y) --- U]
 |--?X [(?X か) --- SX]
 |    |--?X [(?X です) --- SX]
 |         |--?X [(説明書)]
 |
 |--?Y [(?X けれども) --- SX]
      |--?X [(?X のです) --- SX]
           |--?X [(?X ⟨/KIHON-⟩ 気 が する) --- S]
                |--?X [(?X なかっ た) --- CX]
                     |--?X [(見当たら)]
```

ここで翻訳結果は以下のようになる。

```
*** Input ***
 説明書ですか見当たらなかった気がするんですけれども
*** Output ***
 说明书? 好像没有看见。 (3.00002)
```

「説明書ですか」を "吗"を使った疑問文にしたいので、以下の用例を追加する。

```
(define-pattern ending-ka :j-c sx
  (?x "か") (:head 1)  =⟩
  (((!x "吗" "？ ") (子句 1 (语气 2)(符 3)))
   (
    (("説明書"))     ←用例の追加
        :
   )))
```

以下のような翻訳結果が得られるので、これを最終結果とする。

```
*** Input ***
 説明書ですか見当たらなかった気がするんですけれども
*** Output ***
 说明书吗？好像没有看见。 (2.3333535)
```

・exact match 用例の追加

翻訳結果が決まったので、以下の適用パタン一覧を見て exact match 用例の不足を補う。

```
*** Distance calculation in Transfer ***
    1.000010 :: (MARKER-ENDP-U2) U :(?X ⟨ENDP-⟩ ?Y) =⟩ (!X 。 !Y) .... ((いる一) (捜
す))
    0.000000 :: (ENDING-KA) SX :(?X か) =⟩ (!X 吗 ？ ) .... ((説明書))
    0.666667 :: (AUXV-DESU-SX) SX :(?X です) =⟩ (!X) .... ((こと))
    0.000000 :: (~DIC) : (普通名詞 説明書) =⟩ ((名 说明书))
    0.666667 :: (ENDING-KEREDOMO) SX :(?X けれども) =⟩ (!X 。) .... ((存知る))
    0.000010 :: (AUXV-NODESU) SX :(?X のです) =⟩ (!X) .... ((捜す))
    0.000000 :: (/KIHON-KI-GA-SURU) S :(?X ⟨/KIHON-⟩ 気 が する) =⟩ (好像 !X) .... ((見
当たる))
    0.000000 :: (NAKATTA-CX) CX :(?X なかっ た) =⟩ (没有 !X) .... ((見当たる))
    0.000000 :: (~LOCAL) : (本動詞 見当たる) =⟩ ((动 看见))
TOTAL DISTANCE =  2.333354
```

用例が不足しているのは、距離が "0.000000"でない箇所である。
(?x "です")(?x "のです")(?x "けれども")(?x ⟨endp-⟩ ?y) のパタンにそれぞれ用例を追加する。

**8. 中国語生成規則の作成**

この文では、以下の既存の句点の削除の規則が使われている。

```
((？ 。 动) "-⟩" (？ 动) ((:D 1 NIL)))
```

```
*** Generation ***
 ((句子 (子句 (名 "说明书") (语气 "吗") (符 "？ ")) (符 "。") (动短 (动 "好像") (动
短 (动 "没有") (动 "看见"))) (符 "。")))
 =⟩
 说明书吗？好像没有看见。
```

9．訓練終了

翻訳結果が正しく出力されていること、構文解析が正しいこと、意味距離が 0 になっていること、訳文が複数出ていないことを確認し、訓練終了とする。

## 5.2 精度を確認するためのツール類

〔目的〕 新規の訓練を続けていく間に、訓練済みの言語表現に対して副作用を起こしてしまうことがあるため、200 文程度の訓練が終了した時点を目安に、定期的に各種のツールを用いて解析変換知識の精度を確認する作業を行なうのが望ましい。

〔管理者〕 システム管理者

〔格納場所〕

/usr/local/TDMT/tdmt-multi-dev/transfer/code/compare-results.lisp
/usr/local/TDMT/tdmt-multi-dev/misc/*.lisp
※misc に格納されているプログラムは実行前にロードすること。

〔書式および例〕

1. 翻訳結果の作成

素読み用ファイル、また差分確認用のマスタファイルおよび比較ファイルを作成するのに用いる。指定する対象テキストは、text-definitions.lispの階層構造に従うものとする。

書式：(make-result 〈翻訳結果出力ファイル〉 '(((〈対象テキスト〉))))

例：(make-result "~/JC-closed.text" '(("JC-closed"))

2. 翻訳結果の差分の作成

1 で作成したマスタファイルと比較ファイルとの差分を作成するのに用いる。

書式：(compare-results 〈マスタファイル〉 〈比較ファイル〉 〈差分ファイル〉)

例：(compare-results "~/JC-closed.master" "~/JC-closed.text"
"~/JC-closed.dif")

3. 重複用例の確認

同一原言語パタン内で、重複して出現する用例を出力するのに用いる。

書式：(find-same-examples 〈p-data の場所〉 〈出力ファイル〉)
あるいは、(check-example-duplicates 〈p-data の場所〉 〈出力ファイル〉)

例：(find-same-examples "~/j-c/p-data/*.lisp" "~/output")
あるいは、
(check-example-duplicates "~/j-c/p-data/*.lisp" "~/output")

4．過剰訓練の確認

訓練文に使用していない、過剰な解析知識、および変換知識の原言語パタン・目的言語パタン・用例を出力するのに用いる。指定する対象テキストは、text-definitions.lisp の階層構造に従うものとする。

書式：(count-unused-patterns '((⟨対照テキスト⟩)) ⟨出力ファイル⟩)

例：(count-unused-patterns '(("JC-closed")) "~/output")

## 5.3 各種知識の統計データ

※2000年1月1日現在

### 辞書

1. 日本語意味コード辞書 (jma-atr-sem-code.text)
   - 単語数 ... 21,045
   - バイト数 ... 801,051
2. 日中変換辞書 (japanese-to-chinese.lisp)
   - 単語数 ... 10,676
   - バイト数 ... 697,053
3. 中国語生成規則 (gen-rule.lisp)
   - 規則数 ... 148
   - バイト数 ... 9,842

### 解析知識

1. replace-word-info.lisp
   - 単語数 ... 1,699
   - バイト数 ... 106,612
2. lexical-transformation
   - ルール数 ... 156
   - バイト数 ... 42,616

目的別内訳

| 項目 | ルール数 |
|---|---|
| 形態素の合成 | 51 |
| 形態素の除去 | 1 |
| 品詞の変更 | 12 |
| 正規形の修正 | 33 |
| 表層形の修正 | 4 |
| 誤りの修正 | 1 |
| terminal rule | 54 |
| 計 | 156 |

3. local-transformation

- ルール数 ... 33
- バイト数 ... 7,732

## 変換知識

- ルール数 ... 745
- バイト数 ... 400,535

1. パタンの種類別内訳

| パタンの種類 | 個数 |
|---|---|
| define-pattern | 488 |
| define-regular-pattern | 257 |
| 計 | 745 |

2. カテゴリ別内訳

| カテゴリ | 個数 |
|---|---|
| u | 99 |
| sx | 87 |
| s | 70 |
| cx | 67 |
| c | 101 |
| px | 66 |
| p | 42 |
| a | 4 |
| na | 2 |
| nnx | 3 |
| nn | 74 |
| nx | 58 |
| n | 20 |
| terminal | 52 |
| 計 | 745 |

# 参考文献

[古瀬 95] 古瀬蔵，増井淳子：言語データベースの概要．TR-IT-0136(1995)

[河井 94] 河井淳：TDMT 翻訳知識作成ツールの利用方法．TR-IT-0088(1994)

[溝口 98] 溝口淳子：日本語意味コード辞書作成の手引き (1998)

[溝口 99a] 溝口淳子：TDMT 用日本語形態素仕様明細 (1999)

[溝口 99b] 溝口淳子：replace-word の基本原則 (1999)

[宗 99] 宗 成 庆，山 本 和 英：The Part-of-speech and Parsing Rules for the Chinese Language．TR-IT-0307(1999)

# 付録 A

# マーカ一覧

2000年1月1日現在で、定義されているマーカの一覧及びその定義を示す。

## 体言に続くマーカ

| 名称 | 前 | 後 |
|---|---|---|
| ⟨n-n⟩ | 名詞[1]形式名詞 接尾辞 人称接尾辞 | 名詞 連体形容詞 ローマ字 接頭辞 連体詞 |
| ⟨n-adv⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 副詞 接続副詞 |
| ⟨n-i⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 感動詞 |
| ⟨n-v⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 本動詞 |
| ⟨n-a⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 形容詞 |
| ⟨n-an⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 形容名詞 |
| ⟨n-san⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | サ変形容名詞 |
| ⟨n-sn⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | サ変名詞 |
| ⟨n-fn⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 形式名詞 |
| ⟨n-c⟩ | 名詞 形式名詞 接尾辞 人称接尾辞 | 接続詞 |
| ⟨sn-n⟩ | サ変名詞 | 名詞 連体形容詞 ローマ字 接頭辞 連体詞 |
| ⟨sn-adv⟩ | サ変名詞 | 副詞 接続副詞 |
| ⟨sn-i⟩ | サ変名詞 | 感動詞 |
| ⟨sn-v⟩ | サ変名詞 | 本動詞 |
| ⟨sn-a⟩ | サ変名詞 | 形容詞 |
| ⟨sn-an⟩ | サ変名詞 | 形容名詞 |
| ⟨sn-san⟩ | サ変名詞 | サ変形容名詞 |
| ⟨sn-sn⟩ | サ変名詞 | サ変名詞 |
| ⟨sn-fn⟩ | サ変名詞 | 形式名詞 |
| ⟨sn-c⟩ | サ変名詞 | 接続詞 |

[1] ここでは普通名詞、代名詞、数詞、準数詞のこと。

## その他のマーカ

| 名称 | 前 | 後 |
|---|---|---|
| ⟨sou/denbun⟩ | 全活用語の基本形 | 準体助動詞或いは助動詞の「そう」 |
| ⟨/kihon-⟩ | 全活用語の基本形 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨/renyou-⟩ | 全活用語の連用形 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨adv-⟩ | 副詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨cadv-⟩ | 接続副詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨an-⟩ | 形容名詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨san-⟩ | サ変形容名詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨da-⟩ | 連体形容詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨i-⟩ | 感動詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨subp-⟩ | 副助詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨endp-⟩ | 本動詞 助動詞 補助動詞 判定詞[2] 終助詞 準体助動詞 | 内容語 接頭辞 接続詞 連体詞 |
| ⟨r-⟩ | ローマ字 | 内容語[3] 接頭辞 接続詞 連体詞 |
| ⟨r-r⟩ | ローマ字 | ローマ字 |

[2] ただし基本形、連用形の場合を除く。
[3] ただしローマ字の場合を除く。

# 付録 B

## category-jc.lisp

2000年1月1日現在で、category-jc.lisp の全文を以下に示す。

```
;
; category-jc.lisp
;

(in-package "TRANSFER")
(clear-categories :j-c)

(define-category-set word-n  :j-c 普通名詞 固有名詞 代名詞 サ変名詞 形容名詞
                                  サ変形容名詞 形式名詞 数詞 準数詞 ローマ字)
(define-category-set group-n :j-c 普通名詞 固有名詞 代名詞 サ変名詞 形容名詞
                                  サ変形容名詞 形式名詞 数詞 準数詞 ローマ字
                                  副詞 接続副詞 n nx na nn nnx)
(define-category-set word-p  :j-c 本動詞 形容詞 補助動詞 連体形容詞)
(define-category-set group-a :j-c 形容詞 形容名詞 a)

;;; terminal
(define-category terminal :j-c  word-n word-p terminal)

;;; 名詞関係 noun
(define-category n   :j-c word-n n)           ; 複合名詞、氏名
(define-category nx  :j-c word-n n nx)        ;   +接頭辞、接尾辞
(define-category na  :j-c 形容詞 a           ; 形容詞
                          word-n n nx    nn) ;   +名詞類
(define-category nn  :j-c word-n n nx na nn  ; 名詞類+の+名詞類
                          word-p p px c cx)  ; 連体修飾語+名詞類
(define-category nnx :j-c word-n n nx na nn nnx) ; nn + 接辞

;;; 形容詞関係 adjective
(define-category a  :j-c group-a 副詞 接続副詞)  ; 副詞+形容語
```

```
;;; 文関係  predicate -〉 case element -〉 sentence -〉 utterance
(define-category p  :j-c group-n word-p p                ; 複合用言
                        group-a)                         ; 連用修飾語
(define-category px :j-c group-n word-p p px             ;    +助動詞類
                        group-a)                         ;
(define-category c  :j-c group-n word-p p px c           ; 主語以外の格要素
                        group-a 感動詞)
(define-category cx :j-c group-n word-p p px c cx        ;    +助動詞、副詞
                        group-a 感動詞)
(define-category s  :j-c group-n word-p p px c cx s      ; 主語含む文
                        group-a 感動詞)
(define-category sx :j-c group-n word-p p px c cx s sx   ;    +文末語尾
                        group-a 感動詞)
(define-category u  :j-c group-n word-p p px c cx s sx u ; 発話
                        group-a 感動詞)

;
; 更新履歴
; 1999. 2. 8 新体系に移行
; 1999. 2.19 group-n に nnx を追加
; 1999. 3. 5 nn の下位 に cx を追加
; 1999. 3.31 接続副詞を追加
;
```

# 付録 C

# 原言語パタン一覧

2000年1月1日現在で、日中変換知識として使用されている原言語パタンの一覧をカテゴリ別に示す。現在のパタン数は合計で 745 である。

各行に、パタンのカテゴリ、原言語パタンの二項目を記述し、対応する目的言語パタンは省略した。パタンに含まれる変項のうち、標準的でない下部構造の制限がある場合は、それも記述した。パタンの書式には define-regular-pattern と define-pattern の２種類があるが、この区別は省略し、実際にパタンとして利用される表現 (define-pattern なら表層形、define-regular-pattern なら正規形) のまま列挙した。以上のような理由により、リスト上では同一に見える項目もあるが、それらも重複させたまま列挙した。

なお、以下では文字列に対する引用符 "" も省略した。

## C.1 カテゴリ u (99)

u (いずれ にしろ ⟨subp-⟩ ?x)
u (こちら こそ ⟨subp-⟩ ?x)
u (しかし ?x)
u (じゃ ?x)
u (そういたしましたら ?x)
u (そうしましたら ?x)
u (そうしまして ?x)
u (そうしますと ?x)
u (そうすると ?x)
u (そこで ?x)
u (そしたら ?x)
u (そして ?x)
u (その 分 ⟨n-san⟩ ?x)
u (それから ?x)
u (それじゃ ?x)
u (それでしたら ?x)
u (それでは です ね ⟨endp-⟩ ?x)
u (それでは ?x)
u (それで ?x)
u (それと ?x)
u (それに ?x)
u (だから ?x)
u (ですから ?x)
u (ですが ?x)
u (ですけれども ?x)
u (ですので ?x)
u (では ?x)
u (でも ?x)
u (で ?x)
u (ということは ?x)
u (ところで ?x)
u (なん か ⟨subp-⟩ もう ⟨adv-⟩ ?x)
u (実は ?x)
u (?x かどうか ?y)
u (?x から ?y)
u (?x から ?y)
u (?x が ?y)
u (?x けども ?y)
u (?x けど ?y)
u (?x けれども それでも ?y)
u (?x けれども ?y)
u (?x し ?y)
u (?x たらば ?y)
u (?x たら ?y)
u (?x た ⟨/kihon-⟩ ため ⟨n-n⟩ ?y)
u (?x た ⟨/kihon-⟩ まま ⟨n-v⟩ ?y)
u (?x た ⟨/kihon-⟩ 場合 ⟨n-n⟩ ?y)
u (?x だ から ?y)
u (?x って こと は ?y)
u (?x って ?y)
u (?x て それで ?y)
u (?x て ⟨endp-⟩ ?y)
u (?x て ?y)
u (?x でし たら ?y)
u (?x です が ?y)
u (?x では ?y)
u (?x で ⟨/renyou-⟩ ?y)
u (?x で ?y)

```
u (?x で ?y)
  (:x group-n word-p p px c cx s sx group-a)
u (?x と ?y)
u (?x と ?y)
  (:x group-n word-p p px c cx s sx group-a
     感動詞)
u (?x なら ?y)
u (?x な ので ?y)
u (?x には ?y)
u (?x ので ?y)
u (?x ので ?y)
  (:x group-n word-p p px c cx s sx group-a)
u (?x の だ ある ば ?y)
u (?x の 場合 は ?y)
u (?x の ?y)
u (?x は ?y)
u (?x ば ?y)
u (?x まで ?y)
u (?x も ?y)
u (?x よう だ ある ば ?y)
u (?x よう です と ?y)
u (?x ⟨/kihon-⟩ が ?y)
u (?x ⟨/kihon-⟩ ですから ?y)
  (:x group-n word-p p px c cx s sx group-a
     感動詞)
u (?x ⟨/kihon-⟩ で ?y)
u (?x ⟨/kihon-⟩ 場合 ⟨n-n⟩ ?y)
u (?x ⟨/kihon-⟩ ?y)
u (?x ⟨/renyou-⟩ ?y)
  (:x group-n word-p p px c cx s sx group-a
     感動詞)
u (?x ⟨⟩ 次第 ⟨⟩ ?y)
u (?x ⟨⟩ 次第 ?y)
u (?x ⟨adv-⟩ ?y)
u (?x ⟨cadv-⟩ ?y)
u (?x ⟨endp-⟩ だから ?y)
u (?x ⟨endp-⟩ ?y)
u (?x ⟨endp-⟩ ?y)
  (:x group-n word-p p px c cx s sx group-a
     感動詞)
u (?x ⟨i-⟩ ?y)
u (?x ⟨n-adv⟩ ?y)
u (?x ⟨n-c⟩ ?y)
u (?x ⟨n-i⟩ ?y)
u (?x ⟨n-n⟩ ?y)
u (?x ⟨n-n⟩ ?y)
  (:x word-p p px c cx s sx u 感動詞)
u (?x ⟨n-sn⟩ ?y)
u (?x ⟨r-⟩ ?y)
u (?x ⟨sn-c⟩ ?y)
u (?x ⟨sn-n⟩ ?y)
u (?x ⟨subp-⟩ ?y)
```

## C.2 カテゴリ sx (87)

```
sx (お ?x て)
sx (お ?x できる ません か)
sx (もう ⟨adv-⟩ 1度 ⟨n-n⟩ ?x てください)
sx (?x う)
sx (?x かあ)
sx (?x かしら)
sx (?x かどうか)
sx (?x かもしれません)
sx (?x から)
sx (?x か な)
sx (?x か ね)
sx (?x か を ?y)
sx (?x か ⟨endp-⟩ それとも ?y か)
sx (?x か ⟨endp-⟩ ?y)
sx (?x か)
sx (?x が)
sx (?x ください)
sx (?x けど)
sx (?x けれども)
sx (?x し)
sx (?x じゃ ない でしょう か)
sx (?x ずに)
sx (?x た よう)
sx (?x た)
sx (?x だ ある)
sx (?x だ)
sx (?x っけ)
sx (?x ていただける ない でしょう か)
sx (?x ていただける ません か)
sx (?x てください)
sx (?x て は)
sx (?x て)
sx (?x でございます)
sx (?x でした)
sx (?x です)
sx (?x で でし た でしょう か)
sx (?x で なかっ た です か)
sx (?x で ?y)
sx (?x で)
sx (?x という こと)
sx (?x と いい ます と)
sx (?x と 良い)
sx (?x と ?y ⟨an-⟩ どちら に なさい ます か)
sx (?x と ?y)
sx (?x なあ)
sx (?x な のです)
sx (?x な の でしょう か)
sx (?x に も ?y を ください)
sx (?x ね)
sx (?x のです)
sx (?x ので)
sx (?x の かしら)
sx (?x の じゃ ない か なあ)
sx (?x の だ)
sx (?x は どういう ものです か)
sx (?x は)
sx (?x ば よろしい)
sx (?x ば 良い でしょう か)
sx (?x ば 良い わけです か)
sx (?x ば 良い)
sx (?x ば )
sx (?x ましょ う か)
sx (?x ません か)
sx (?x ません でした か)
sx (?x ません でしょう か)
sx (?x ませ)
sx (?x ものです)
sx (?x も ください)
sx (?x も ?y を ください)
sx (?x よう です)
sx (?x よ ね)
sx (?x よ)
sx (?x わ)
sx (?x を ください)
sx (?x を 下さい)
sx (?x ⟨/renyou-⟩ ?y)
```

```
sx (?x ⟨⟩ なる ない でしょう か)
sx (?x ⟨⟩ もの で)
sx (?x ⟨⟩ わけ)
sx (?x ⟨adv-⟩ ?y)
sx (?x ⟨an-⟩ ?y)
sx (?x ⟨i-⟩ ?y へ)
sx (?x ⟨n-c⟩ それとも ?y か)
sx (?x ⟨n-n⟩ それとも ?y か)
sx (?x ⟨n-v⟩ ください)
sx (?x ⟨n-v⟩ 下さい)
sx (?x ⟨sn-a⟩ ?y)
```

## C.3 カテゴリ s (70)

```
s (ちょっと ⟨n-n⟩ ?x)
s (なかなか ⟨an-⟩ ?x が ?y)
s (なん か ⟨subp-⟩ ?x が ?y)
s (まあ ⟨adv-⟩ ?x)
s (もう ⟨adv-⟩ ?x)
s (今 まで ?x)
s (?x かどうか ⟨endp-⟩ ?y)
s (?x か と ?y)
s (?x か は ?y)
s (?x か ⟨endp-⟩ ?y)
s (?x か ⟨subp-⟩ ?y)
s (?x が ございます の は ?y)
s (?x が ?y)
s (?x た という ?y)
s (?x た の は ?y)
s (?x だ と ?y)
s (?x って ?y)
s (?x て から ?y)
s (?x て ?y)
s (?x です が ?y)
s (?x です けれども ?y)
s (?x です ね ⟨endp-⟩ ?y)
s (?x では ?y)
s (?x でも ⟨subp-⟩ ?y)
s (?x でも ?y)
s (?x で です ね ⟨endp-⟩ ?y)
s (?x で ?y)
s (?x という ?y)
s (?x という ?y)
s (?x と ?y)
s (?x ながら ?y)
s (?x なら ?y)
s (?x なんて ⟨subp-⟩ ?y)
s (?x にする と ?y)
s (?x について ?y)
s (?x には ?y)
s (?x によって ?y)
s (?x に つき まし て は ?y)
s (?x に です ね ⟨endp-⟩ ?y)
s (?x にも ?y)
s (?x に 関し て は ?y)
s (?x に 関し まし て は ?y)
s (?x に ?y)
s (?x ので ?y)
s (?x の か ⟨endp-⟩ ?y)
s (?x の ため に ?y)
s (?x の は ?y)
s (?x の 件 で ?y)
s (?x の 件 は ?y)
s (?x の ?y)
s (?x は です ね ⟨endp-⟩ ?y)
s (?x は ?y)
s (?x ばと ?y)
s (?x ば ?y)
s (?x まで ?y)
  (:x word-n n nx na nn)
s (?x も ?y も ?z)
  (:x word-n n nx)
  (:y word-n n nx)
s (?x も ?y)
s (?x よう ⟨/kihon-⟩ ?y)
s (?x を ?y)
s (?x ⟨/kihon-⟩ 気 が する)
s (?x ⟨adv-⟩ ?y)
s (?x ⟨n-adv⟩ ?y)
s (?x ⟨n-an⟩ ?y)
s (?x ⟨n-n⟩ とも ⟨n-an⟩ ?y)
s (?x ⟨n-n⟩ とも ⟨n-n⟩ ?y)
s (?x ⟨n-n⟩ とも ⟨n-n⟩ ?y)
s (?x ⟨n-n⟩ ?y)
s (?x ⟨n-sn⟩ ?y)
s (?x ⟨n-v⟩ ?y)
s (?x ⟨sn-v⟩ ?y)
```

## C.4 カテゴリ cx (67)

```
cx (そんなに ⟨adv-⟩ ?x ない)
cx (もし ⟨⟩ ?x たら と 思う)
cx (?x いただける)
cx (?x う と する)
cx (?x かどうか)
cx (?x が できる)
cx (?x が よろしい)
cx (?x しか ⟨subp-⟩ ?y ません)
cx (?x せる)
cx (?x そう)
cx (?x たい)
cx (?x たら と 思う)
cx (?x たら 良い)
cx (?x た だけ)
cx (?x た)
cx (?x だ ない か と ?y)
cx (?x だ ない)
cx (?x だ は ございません)
cx (?x だ は 無い)
cx (?x てある)
cx (?x ていただく ておる)
cx (?x ていただく)
cx (?x ていただける)
cx (?x ている)
cx (?x ておく)
cx (?x ておる)
cx (?x てください)
cx (?x てもらえる)
cx (?x て よろしい)
cx (?x て 良い)
cx (?x できる)
cx (?x でしょう)
```

cx (?x という こと)
cx (?x ない)
cx (?x なかっ た)
cx (?x なければならない)
cx (?x なければなりません)
cx (?x など)
cx (?x な の でしょう)
cx (?x にする)
cx (?x になる)
cx (?x のみ)
cx (?x の でしょう)
cx (?x は できる)
cx (?x ば 良い)
cx (?x ましょ)
cx (?x まし)
cx (?x ます)
cx (?x ません)
cx (?x も できる)
cx (?x よう にする)
cx (?x れる)
cx (?x わけです)
cx (?x をする)
cx (?x ⟨/kihon-⟩ はず)
cx (?x ⟨/kihon-⟩ 者)
cx (?x ⟨/renyou-⟩ ?y)
cx (?x ⟨⟩ こと が ある)
cx (?x ⟨⟩ こと が できる)
cx (?x ⟨⟩ こと が できる)
cx (?x ⟨⟩ こと にする)
cx (?x ⟨⟩ こと になる)
cx (?x ⟨⟩ こと は できる)
cx (?x ⟨⟩ こと ⟨⟩ できる)
cx (?x ⟨⟩ こと)
cx (?x ⟨⟩ 方法 が ある)
cx (?x ⟨an-⟩ ?y)

## C.5 カテゴリ c (101)

c (あう せる て ?x)
c (ちょっと ⟨n-adv⟩ ?x)
c (ちょっと ⟨n-i⟩ ?x)
c (ちょっと ⟨n-san⟩ ?x)
c (ちょっと ⟨n-sn⟩ ?x)
c (どなた か ⟨subp-⟩ ?x)
c (もう ⟨adv-⟩ 少し ⟨adv-⟩ ?x)
c (もう ⟨adv-⟩ ?x)
c (もし ⟨⟩ ?x さえ ⟨⟩ ?y)
c (よく ⟨/renyou-⟩ ?x ない )
c (誰 か ⟨subp-⟩ ?x に ?y)
c (?x から の ?y)
  (:x word-n n nx na nn)
c (?x から ?y まで ?z)
  (:x word-n n nx na nn)
  (:y word-n n nx na nn)
c (?x から ?y)
  (:x word-n n nx na nn)
c (?x か で ?y)
c (?x か と ?y)
c (?x か ⟨subp-⟩ なに か ⟨subp-⟩ ?y)
c (?x か ⟨subp-⟩ ?y)
c (?x が ?y)
c (?x ごと に ?y)
c (?x たり ⟨subp-⟩ ?y たり)
c (?x だけ で ?y)
c (?x だけ ⟨subp-⟩ ?y)
c (?x ても ?y)
c (?x て から ?y)
c (?x て そして ?y)
c (?x て ?y)
c (?x では ?y)
c (?x でも ⟨subp-⟩ ?y)
c (?x でも ?y)
c (?x で でも ⟨subp-⟩ ?y)
c (?x で は あり ません)
c (?x で ?y を お 願う する)
c (?x で ?y)
c (?x という ふう に ?y)
c (?x という もの)
  (:x word-n n nx na)
c (?x という ?y)
  (:x word-n n nx na)
c (?x として ?y)
c (?x と の ?y)
c (?x と は 別 に ?y)
c (?x と 同じ よう に ?y)
c (?x と 別 に ?y)
c (?x と ?y)
c (?x ながら ?y)
c (?x なら ?y)
c (?x な のです が それ が ?y)
c (?x な ?y)
c (?x につき ?y)
c (?x には ?y)
c (?x に でも ⟨subp-⟩ ?y)
c (?x に ?y)
c (?x の だ は 無い か と ?y)
c (?x の は ?y)
c (?x の よう に ?y)
c (?x の を ?y)
c (?x の 中 でも ?y)
c (?x の ?y を ?z)
  (:x word-n n nx na nn)
  (:y word-n n nx na nn)
c (?x の ?y)
c (?x の ?y)
  (:x word-n n nx na)
c (?x は 別 にする)
c (?x は ?y のみ となる)
c (?x は ?y)
c (?x へ の ?y)
c (?x へ 向かっ て ?y)
c (?x へ ?y)
c (?x まで ある)
c (?x まで に ?y)
c (?x まで は ?y)
c (?x まで ?y)
  (:x word-n n nx na nn)
c (?x も とも に ?y)
c (?x も ?y)
c (?x よう に ?y)
c (?x より も ?y が ある)
c (?x より も ?y )
c (?x より ?y)
c (?x を どうぞ)
c (?x を ?y)
c (?x を ?y)
c (?x ⟨/kihon-⟩ ?y)
c (?x ⟨/renyou-⟩ ?y)
c (?x ⟨⟩ こと も ?y)
c (?x ⟨⟩ 以外 に ?y)

c (?x ⟨adv-⟩ ?y ません)
c (?x ⟨adv-⟩ ?y)
c (?x ⟨adv-⟩ ?y)
c (?x ⟨an-⟩ ?y)
c (?x ⟨cadv-⟩ ?y)
c (?x ⟨i-⟩ ?y)
c (?x ⟨n-a⟩ ?y)
c (?x ⟨n-adv⟩ ?y)
c (?x ⟨n-n⟩ ?y)
c (?x ⟨n-sn⟩ ?y)
c (?x ⟨n-v⟩ ?y)
c (?x ⟨san-⟩ ?y)
c (?x ⟨sn-i⟩ ?y)
c (?x ⟨sn-n⟩ どおり)
c (?x ⟨sn-n⟩ 中)
c (?x ⟨sn-n⟩ ?y)
c (?x ⟨sn-sn⟩ ?y)
c (?x ⟨sn-v⟩ ?y)
c (?x ⟨subp-⟩ ?y)

## C.6　カテゴリ px (66)

px (?x いたす)
px (?x いただく)
px (?x いただける)
px (?x かねる)
px (?x かもしれない)
px (?x かもしれません)
px (?x か という こと)
px (?x が ある)
px (?x が ございます)
px (?x さしあげる)
px (?x じゃう)
px (?x する)
px (?x す)
px (?x せる)
px (?x たい と 思う)
px (?x たい)
px (?x た)
px (?x ちゃう)
px (?x ていく)
px (?x ていただける)
px (?x ている た ⟨⟩ こと が ある)
px (?x ている)
px (?x ておく)
px (?x ておる)
px (?x てください)
px (?x てくださる)
px (?x てくる ている)
px (?x てくる)
px (?x てくれる)
px (?x てしまう)
px (?x てはいない)
px (?x てまいる)
px (?x てまわる)
px (?x てみる)
px (?x てもらう)
px (?x てらっしゃる)
px (?x て帰る)
px (?x できる)
px (?x というの)
px (?x という こと)
px (?x とく)
px (?x ないでくださる)
px (?x ない で 済む)
px (?x ない)
px (?x なさる)
px (?x にくい)
px (?x になる れる)
px (?x の こと)
px (?x ば な と 思う)
px (?x ましょ)
px (?x まし)
px (?x ません)
px (?x やすい)
px (?x よう です)
px (?x れる)
px (?x を 致す)
px (?x を 頂戴 ておる)
px (?x を ?y せる)
px (?x 願う たい)
px (?x ⟨/kihon-⟩ はず)
px (?x ⟨⟩ こと)
px (?x ⟨⟩ する ていただける)
px (?x ⟨⟩ する)
px (?x ⟨⟩ なる)
px (?x ⟨⟩ やる ていく)
px (?x ⟨san-⟩ いり ません)

## C.7　カテゴリ p (42)

p (お ?x いたす)
p (お ?x いただく)
p (お ?x いただける)
p (お ?x ください)
p (お ?x する)
p (お ?x せる)
p (お ?x だ いらっしゃる)
p (お ?x できる)
p (お ?x でございます)
p (お ?x でしょう)
p (お ?x です)
p (お ?x になる)
p (お ?x ねがえる)
p (お ?x もうしあげる)
p (お ?x 願う)
p (お ?x 申す)
p (ご ?x いただく)
p (ご ?x いただける)
p (ご ?x ください)
p (ご ?x になれる)
p (ご ?x ねがえる)
p (ご ?x ⟨⟩ 無い)
p (ご ?x)
p (?x だ いらっしゃる)
p (?x だ らっしゃる)
p (?x だ)
p (?x でございます う)
p (?x となる)
p (?x ない ⟨⟩ なる)
p (?x な のです)

```
p (?x にする)
p (?x になる)
p (?x の ものです)
p (?x は する)
p (?x をする)
p (?x を なさる)
p (?x ⟨/renyou-⟩ ある )
p (?x ⟨/renyou-⟩ ございます )
p (?x ⟨⟩ こと になる ている)
p (?x ⟨⟩ こと になる)
p (?x ⟨⟩ こと になる)
p (?x ⟨n-sn⟩ ?y)
```

## C.8 カテゴリ a (4)

```
a (その ?x)
a (十分 ⟨an-⟩ ?x)
a (?x に ?y)
a (?x ⟨/renyou-⟩ ?y)
```

## C.9 カテゴリ nnx (3)

```
nnx (あの ?x)
nnx (?x まで の ?y)
nnx (?x ⟨/kihon-⟩ 方)
```

## C.10 カテゴリ nn (74)

```
nn (その ?x)
nn (ただ の ?x)
nn (どこ か ⟨⟩ ?x)
nn (また の ?x)
nn (?x から ?y まで)
   (:x word-n n nx na nn)
   (:y word-n n nx na nn)
nn (?x から ?y)
   (:x word-n n nx na nn)
   (:y word-n n nx na nn)
nn (?x か ⟨subp-⟩ ?y)
nn (?x か ⟨subp-⟩ ?y)
nn (?x か ⟨subp-⟩ ?y)
   (:y word-n n nx na)
nn (?x が ?y)
nn (?x が ?y)
nn (?x た ⟨/kihon-⟩ 方)
nn (?x た ⟨/kihon-⟩ ?y)
nn (?x た ⟨/kihon-⟩ ?y)
   (:y word-n n nx na)
nn (?x た ⟨⟩ もの)
nn (?x て ?y)
nn (?x で それと ?y)
nn (?x で の ?y)
nn (?x で ?y)
nn (?x という もの)
   (:x word-n n nx na)
nn (?x といった もの)
   (:x word-n n nx na)
nn (?x とか ね ⟨endp-⟩ ?y とか
    ⟨subp-⟩ ?z とか)
nn (?x とか ⟨subp-⟩ ?y とか)
nn (?x とか ⟨subp-⟩ ?y)
nn (?x と それから ?y)
nn (?x と です ね ⟨endp-⟩ ?y)
   (:y word-n n nx na)
nn (?x と ?y と ?z)
nn (?x と ?y)
nn (?x な ?y)
nn (?x な ?y)
   (:y word-n n nx na nn)
nn (?x には ?y)
nn (?x に ?y)
nn (?x の よう な ?y)
nn (?x の 間)
nn (?x の ?y)
nn (?x の ?y)
   (:y word-n n nx na)
nn (?x は ?y ⟨⟩ もの)
nn (?x は ?y)
nn (?x は ?y)
   (:x word-n n nx na)
nn (?x へ の ?y)
nn (?x みたい な ?y)
nn (?x や ?y)
nn (?x よう な ?y)
nn (?x より ?y)
   (:x word-n n nx na nn)
   (:y word-n n nx na nn)
nn (?x らしい ⟨/kihon-⟩ ?y)
nn (?x を ?y)
nn (?x を ?y)
nn (?x ⟨/kihon-⟩ 分)
nn (?x ⟨/kihon-⟩ 方)
nn (?x ⟨/kihon-⟩ ?y)
nn (?x ⟨/kihon-⟩ ?y)
nn (?x ⟨/kihon-⟩ ?y)
   (:y word-n n nx na)
nn (?x ⟨⟩ あるいは ?y)
nn (?x ⟨⟩ 当たり ⟨⟩ ?y)
nn (?x ⟨an-⟩ ?y)
   (:y word-n n nx na)
nn (?x ⟨da-⟩ ?y)
nn (?x ⟨endp-⟩ ?y)
nn (?x ⟨n-a⟩ ?y)
nn (?x ⟨n-an⟩ ?y)
nn (?x ⟨n-c⟩ そして ?y)
nn (?x ⟨n-c⟩ もしくは ?y)
nn (?x ⟨n-n⟩ そして ?y)
nn (?x ⟨n-n⟩ 前 の ?y)
nn (?x ⟨n-n⟩ ?y ⟨n-n⟩ ?z)
nn (?x ⟨n-n⟩ ?y)
   (:y word-n n nx na nn)
nn (?x ⟨n-sn⟩ 関係)
nn (?x ⟨n-sn⟩ ?y ⟨sn-n⟩ ?z)
nn (?x ⟨n-sn⟩ ?y)
nn (?x ⟨sn-fn⟩ ?y)
nn (?x ⟨sn-n⟩ ?y)
nn (?x ⟨sn-sn⟩ ?y)
```

```
nn (?x ⟨subp-⟩ あと は ?y)
nn (?x ⟨subp-⟩ し た⟨/kihon-⟩ ?y)
```

```
nn (?x ⟨subp-⟩ ?y)
```

## C.11 カテゴリ na (2)

```
na (?x で ⟨/renyou-⟩ ?y)
```

```
na (?x ⟨/kihon-⟩ ?y)
```

## C.12 カテゴリ nx (58)

```
nx (あの ?x)
nx (あらゆる ?x)
nx (いろんな ?x)
nx (おもだった ?x)
nx (この ?x)
nx (こんな ?x)
nx (そういう ?x)
nx (そういった ?x)
nx (そのような ?x)
nx (その ?x)
nx (ちょっとした ?x)
nx (どういう ?x)
nx (どういった ?x)
nx (どのような ?x)
nx (どの ?x)
nx (どんな ?x)
nx (ぼんの ?x)
nx (また ?x)
nx (約 ?x)
nx (?x から)
nx (?x か)
nx (?x くらい)
nx (?x さん)
nx (?x しか)
nx (?x ずつ)
nx (?x だけ)
nx (?x というの)
nx (?x という こと)
nx (?x という ところ)
```

```
nx (?x という よう な もの)
nx (?x とう)
nx (?x とか)
nx (?x など)
nx (?x なんか)
nx (?x について)
nx (?x のみ)
nx (?x の こと)
nx (?x の ため)
nx (?x の ほう)
nx (?x の もの)
nx (?x の 者)
nx (?x の 場合)
nx (?x の 分)
nx (?x の 方)
nx (?x のところ)
nx (?x の)
nx (?x ほど)
nx (?x まで)
nx (?x 的)
nx (?x 等)
nx (?x 様)
nx (?x ⟨⟩ 頃)
nx (?x ⟨⟩ 付き)
nx (?x ⟨n-n⟩ 気味)
nx (?x ⟨n-n⟩ 建て)
nx (?x ⟨n-n⟩ 向け)
nx (?x ⟨n-n⟩ 程度)
nx (?x ⟨n-sn⟩ ?y)
```

## C.13 カテゴリ n (20)

```
n (お ?x)
n (ご ?x)
n (?x が ?y)
n (?x の これ は ?y)
n (?x の ?y)
n (?x の ?y)
n (?x を ?y)
n (?x 達)
n (?x ⟨⟩ 当たり)
n (?x ⟨n-n⟩ あたり)
n (?x ⟨n-n⟩ 以上)
```

```
n (?x ⟨n-n⟩ 分)
n (?x ⟨n-n⟩ ?y)
n (?x ⟨n-n⟩ ?y)
n (?x ⟨n-sn⟩ ?y)
n (?x ⟨r-⟩ ?y)
n (?x ⟨r-r⟩ ?y)
  (:x ローマ字)
n (?x ⟨sn-n⟩ 分)
n (?x ⟨sn-n⟩ ?y)
n (?x ⟨sn-sn⟩ ?y)
```

## C.14 カテゴリ terminal (52)

```
terminal (第 ?x)
terminal (?x さ)
terminal (?x め)
terminal (?x カ月)
terminal (?x キロ)
terminal (?x セント)
terminal (?x タイプ)
terminal (?x ドル)
terminal (?x パーセント)
```

```
terminal (?x ポンド)
terminal (?x 円)
terminal (?x 回)
terminal (?x 階)
terminal (?x 曲)
terminal (?x 月)
terminal (?x 個)
terminal (?x 号 室)
terminal (?x 才)
```

```
terminal (?x 歳)
terminal (?x 皿)
terminal (?x 時間)
terminal (?x 時)
terminal (?x 種類)
terminal (?x 週間)
terminal (?x 食)
terminal (?x 人 様)
terminal (?x 人)
terminal (?x 台)
terminal (?x 点 ?y パーセント)
terminal (?x 点 ?y 倍)
terminal (?x 度)
terminal (?x 棟)
terminal (?x 日 間)
terminal (?x 日 分)
terminal (?x 日 目)
terminal (?x 日)
terminal (?x 年)
terminal (?x 杯)
terminal (?x 倍)
terminal (?x 泊)
terminal (?x 箱)
terminal (?x 晩)
terminal (?x 番)
terminal (?x 品)
terminal (?x 部屋)
terminal (?x 部)
terminal (?x 分)
terminal (?x 本)
terminal (?x 枚)
terminal (?x 名 様)
terminal (?x 名)
terminal (?x ⟨n-n⟩ ?y)
         (:x 数詞)
```

# 索引