

TR-I-0307

日英変換主導型機械翻訳システム
Japanese-to-English
Transfer-Driven Machine Translation System

古瀬 蔵 河井 淳* 栗原 英男**
Osamu FURUSE Jun KAWAI Hideo KURIHARA

1993.3

概要

ATRでは、話し言葉の新しい翻訳手法として、変換主導型機械翻訳(Transfer-Driven Machine Translation)の研究を行なっている。変換主導型機械翻訳では、翻訳のための経験的知識である変換知識を活用して、入力文の性質に応じた柔軟な翻訳処理を行なう。

日英変換主導型機械翻訳システムは、現在、日本語1500語彙のレベル、国際会議に関する問い合わせというドメインで、効率的な処理を実現している。このシステムはシンボリック・リスプマシン上で開発されているが、ユーザインターフェース部以外はCOMMON LISPで記述されているため、UNIXマシン上でも翻訳実行できる。本資料では、日英変換主導型機械翻訳システムの概要およびプログラムの実行方法について説明する。

ATR自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

© (株) ATR自動翻訳電話研究所 1993
© 1993 ATR Interpreting Telephony Research Laboratories

*株式会社東洋情報システム

**ニチメングラフィックス株式会社

TDMT システム

目次

I. TDMT システム動作原理

1. TDMT概要	1
2. TDMTシステムの枠組み	2
3. 用例距離計算について	3
3-1. 意味コードを用いた距離計算	3
3-2. 変換、解析知識の記述と用例距離計算	5
3-3. 構造的曖昧性などの解消	7
4. TDMT による日英翻訳システム	8
4-1. 翻訳処理の制御	8
4-2. TDMT での変換、解析知識について	9
4-3. プライオリティー制御	10
5. TDMT システムによる翻訳のしくみ	11
5-1. スtringマッチング	11
5-2. パターンマッチング	11
5-3. 文法マッチング	12
5-4. 変換モジュールにおける変換知識の適用	12
5-5. 解析処理	13

II. システム機能仕様

1. システム概要	14
2. 全体の構成	15
3. 翻訳処理部	16
3-1. 形態素処理部	16
3-2. 解析部	16
3-3. 変換部	17
3-4. 生成部	20
4. ユーザー・インターフェース	22
4-1. ウィンドウ	22
4-2. コマンド詳細	23

5. ファイル構成	25
5-1. 関数ファイル群	25
5-2. 辞書ファイル群	26
5-3. 変換知識ルールファイル群	26
5-4. 解析知識ルールファイル群	26

III. プログラム説明

1. 主要な関数	27
2. モジュール説明	28
2-1. 翻訳処理部	28
2-2. ユーザ・インターフェース部	41

IV. データ仕様

1. 知識ルール	44
1-1. 解析知識ルールの仕様	44
1-2. 変換知識ルールの仕様	48
2. 辞書	51
2-1. 日英変換辞書	51
2-2. 英語属性辞書	52
2-3. 不規則変化動詞辞書	54

V. 操作説明

1. インストールおよびシステム起動方法	55
1-1. インストール	55
1-2. システム起動	56
2. ウィンドウおよびコマンド説明	57
3. 知識ルールについて	58
3-1. 知識ルールの概要	58
3-2. 知識ルールの変更	58
4. 辞書について	59
4-1. 辞書の概要	59
4-2. 辞書の変更	59

5. UNIX 版について	60
5-1. システム起動	60
5-2. 使用方法	61

付録 A. MORPH 62

日本語形態素解析システム

I. システム機能説明

1. システムの概要	63
2. システムの動作原理	64
3. システムの構成	68
4. システムの機能	69

II. プログラム説明

1. 主要な関数	70
2. モジュール説明	71

III. 辞書説明

1. 品詞、活用、活用形コード表	76
2. モノグラム・データ	77
3. バイグラム・データ	78
4. 意味コード辞書	79
5. 形態素辞書	80

IV. 操作説明

1. インストール	81
2. プログラムのコンパイルとロード	82
3. 辞書およびデータのロード	82
4. 実行関数の使用方法	83
5. 辞書の変更	83

付録 B. TDMT-COLOR 84

カラー・デモンストレーション・システム

I. システム機能仕様

1. システム概要	85
2. ユーザー・インターフェース	86
2-1. ウィンドウ	86
2-2. コマンド	87
3. ファイル構成	88
3-1. 関数ファイル群	88
3-2. イメージデータファイル群	88

II. プログラム説明

1. 主要な大域変数	89
2. 主要な関数	90

III. 操作説明

1. インストールおよびシステム起動方法	91
1-1. インストール	91
1-2. システム起動	91
2. ウィンドウおよびコマンド説明	92

付録 C. 翻訳実行結果 94

I. TDMT システム動作原理

1. TDMT概要

変換主導型機械翻訳 (Transfer-Driven Machine Translation 以下 TDMT と呼ぶ) は、原言語表現と目的言語表現の対応関係を示す変換知識を、翻訳という問題を解く鍵として位置づけ、その適用を制御する変換モジュールを中心として効率的な機械翻訳を実現するための翻訳手法である。

人間は翻訳する際に、様々な戦略と様々なレベルの知識を使い分けられていると考えられる。すなわち、単純な文に対しては浅いレベルの知識のみを使って即座に翻訳結果を返し、複雑な文に対しては構文、意味、文脈などの知識を駆使して翻訳を行なっている。

これと同様に TDMT では入力文の性質に応じて必要な処理を駆動し、効率よく翻訳を行なう。中心となる変換モジュールは入力文を翻訳するために必要な変換知識を次々と選び出し、適用していく。変換知識には文の基本構成を捉えるグローバルな単位のものから、語句を捉えるローカルな単位のものまであり、そのうちのどの知識をどういった順序で適用すべきかを決定しなければならない。このために TDMT では用例主導の枠組みを取り入れ、用例距離計算を用いてそれぞれの知識の適用が妥当であるかどうかを判定するようになっている。用例距離計算は明瞭で単純なため、高速で処理できる。また、この結果は複数の翻訳結果からもっともらしいものを選ぶ際にも役に立つ。さらに、文の構造が複雑で変換知識だけでは翻訳が不可能な場合は、解析機構が駆動され解析知識が適用される。ここでも、用例距離計算により知識の適用が制御される。生成機構では変換知識から与えられた生成マーカーと格情報、および前文参照による文脈情報によって必要な処理が駆動される。

この章では TDMT による翻訳処理の原理を説明しているので、実験システムの詳細についてのみ知りたい方はとばしていただけてください。

実験システムは Symbolics Machine 上の Genera 8.0.1 の環境で作成されていますが、翻訳処理部は Common Lisp の仕様にもとづいているため UNIX の Lisp 上でも動作可能です。詳しくは操作説明の章を参照のこと。

2. TDMTシステムの枠組み

TDMTシステムは図1に示すようなモジュールの組み合わせにより概念的に表現できる。TDMTでは変換モジュールが中心となり、解析や生成のモジュールがその働きを助けるしくみになっている。入力文が単純な場合はほとんど変換モジュールだけで翻訳を行ない、複雑になるに従って解析、生成へ負荷が分散される。変換、解析では用例距離計算による評価が用いられ、効率的に知識の適用が行なわれる。

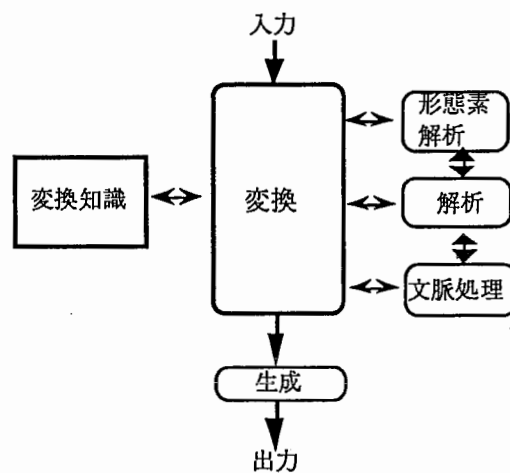


図1 変換主導翻訳の概念

3. 用例距離計算について

TDMT では変換、解析の知識に対して用例距離計算を行ない、その適用が妥当であるかどうかを判断している。ここでは、TDMT システムで用いられている用例距離計算について説明する。

3-1. 意味コードを用いた距離計算

変換、解析知識の適用を制御するために用例距離計算を行なう。用例距離計算とは入力と用例の重みづけを動的に行なう方法で、シソーラス上での意味属性の距離を求めることである。

それぞれの語は特定のシソーラスコード（以下意味コードという）を持ち、それはある意味属性に対応する。二つの意味属性間の距離はシソーラス階層での位置関係によって決められ、図2に示すように0から1までの値を与える。d(A, B)は意味属性AとBの間の距離を表す。

値0は二つの意味属性が同じカテゴリーに属することを意味し、値1は関係づけられないことを意味する。例えば、意味属性 'writing' と 'book' はすぐ上位の意味属性 'document' で一般化できるので、意味属性 'writing' を持つ語「論文」と、意味属性 'book' を持つ語「予稿集」の距離は1/3となる。

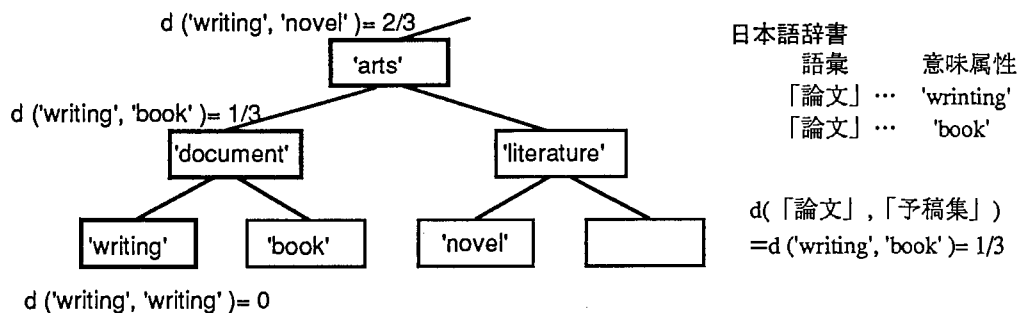


図2 シソーラスの階層と距離

なお、TDMT 日英翻訳システムでは用例距離計算には正規形を使い、正規形が一致していれば” exact matching ”という状態として、意味コードを用いた距離計算は行なわず値0を与えている。また、この場合と、通常の距離計算の結果、値が0になった場合とを区別するために、距離計算を行なうときにわずかな値を加えている。その他の場合で、” 語 ”が意味コードを持たないものの場合は” 意味コード無し ”という状態で、現在のところ値1/2を与えている。

以上は対象の語が唯一の意味属性に属する場合であるが、より一般的には特定の意味属性に属さず複数の意味コードを持つ語が考えられる。(例えば文脈などにより全く違う意味に使われる語などがこれにあたる。) こうした語に対応するために次のように語の距離を拡張定義する。いま、二つの対象となる語を W_a, W_b とし、それらの持つ意味コードをそれぞれ S_a, S_b で表す。また、その個数をそれぞれ m, n 個とする。このとき、 W_a, W_b のどちらが入力でどちらが用例であるかは問わない。

$$\begin{aligned} W_a(S_{a_1}, S_{a_2}, S_{a_3}, \dots, S_{a_m}) & \dots (3-1-1) \\ W_b(S_{b_1}, S_{b_2}, S_{b_3}, \dots, S_{b_n}) & \dots (3-1-2) \end{aligned}$$

ここで次の手順を用いて距離計算を行なう。

[1] W_a から i 番目の意味コードを一つ取りだし、 W_b のそれぞれの意味コードと距離計算を行ない、その最小の値を da_i とする。

$$da_i = \min(d(S_{a_i}, S_{b_1}), d(S_{a_i}, S_{b_2}), \dots, d(S_{a_i}, S_{b_n})) \dots (3-1-3)$$

[2] W_b から j 番目の意味コードを一つ取りだし、 W_a のそれぞれの意味コードと距離計算を行ない、その最小の値を db_j とする。

$$db_j = \min(d(S_{a_1}, S_{b_j}), d(S_{a_2}, S_{b_j}), \dots, d(S_{a_m}, S_{b_j})) \dots (3-1-3)'$$

[3] i, j を 1 から m, n まで走査し、すべての d_i に対する平均を W_a, W_b に対する語の距離 D_{ab} とする。

$$D_{ab} = (\sum_{i=1}^m d_i + \sum_{j=1}^n d_j) / (m + n) \dots (3-1-4)$$

この定義によれば $m=n=1$ とした場合、 $D_{ab} = d(S_{a_1}, S_{b_1})$ となり元の定義に一致する。また、 W_a, W_b が共に全く同じ意味コードの集合を持つ場合 (例えば $W_a = W_b$ の場合)、 $D_{ab} = 0$ となる。

3-2. 変換、解析知識の記述と用例距離計算

後述のように、パターンレベルの変換知識は原言語表現から目的言語表現への写像として、また、解析知識は原言語表現から修正原言語表現への写像として記述される。その対応は一般に一对多の形をしており、解としてはその内の一つを選択する必要がある。その際に、選択条件として記述された用例と、原言語表現にマッチした入力との距離計算を行ない最小の値を示したものを選択する。

原言語表現=>

目的言語表現₁ (用例₁₁ 用例₁₂ . . .)

:

目的言語表現_n (用例_{n1} 用例_{n2} . . .)

. . . (3-2-1)

原言語表現=>

修正原言語表現₁ (用例₁₁ 用例₁₂ . . .)

:

修正原言語表現_n (用例_{n1} 用例_{n2} . . .)

. . . (3-2-2)

ここで原言語表現中の入力（ここでは変項に結び付けられた語、および語句、のよう
に用例距離計算の対象となる部分）の個数を m とすると、用例_{ij} は $(E_{ij1}, E_{ij2}, \dots, E_{ijm})$ といった k 個の語（または語句）の組として表され、同様に入力は $(l_1, l_2, \dots, l_m)$ で表される。このとき、入力と用例の距離を次のように定義する。

$$d_{ij}(\text{用例}_{ij}, \text{入力}) = d_{ij}((E_{ij1}, E_{ij2}, \dots, E_{ijm}), (l_1, l_2, \dots, l_m))$$

$$= (1/m) \sum_{k=1}^m d(E_{ijk}, l_k)$$

. . . (3-2-3)

なお、この E_{ijk}, l_k は一語とは限らず、一連の複数語からなる語句の場合もあるが、このときは次のように d_{ij} を求める。 i, j, k の表示を省略して、 E_{ijk}, l_k をそれぞれ u 個、 v 個の一連の語から成っているとする。

$$E = (e_1, e_2, \dots, e_u) \quad \dots (3-2-4)$$

$$l = (l_1, l_2, \dots, l_v) \quad \dots (3-2-5)$$

i) $u \leq v$ の場合

$$d(E, l) = (1/u) \sum_{i=1}^u d(e_i, i_{v-u+i}) \quad \dots (3-2-6)$$

これを (3-2-3) の $d(E_{ijk}, l_k)$ として用いる。

ii) $u > v$ の場合

この場合、 $d(E, l)$ は未定義。従って、(3-2-3) の d_{ij} は”意味コード無し”の状態と等しく扱われる。

上記のようにして求めた d_{ij} を用いて、入力と目的言語表現_i、あるいは入力と修正原言語表現_iとの距離 d_i は次のように定義する。

$$d_i = \min_j (d_{ij}(\text{用例}_{ij}, \text{入力})) \quad \dots (3-2-7)$$

この計算の結果、最小の d_i を持つ目的言語表現_i、あるいは修正原言語表現_iが選択される。また、このとき最小の d_i が複数ある場合は解（または解候補）が分裂する。

変換モジュールはパターン変換知識に対するパターン・マッチングと用例距離計算により変換木構造を組み立てるが、このとき、パターン・マッチングと用例距離計算の両方でその曖昧性により解（または解候補）に複数の可能性を生じることがある。この場合、距離の総和により最も適切な構造を選択したり、しきい値を設けて無駄な枝を刈ることができる。

(会議の(登録費の割引)) . . . (3-2-8)
((会議の登録費) の割引) . . . (3-2-9)

(X の Y) =>

(Y' of X')	(用例 _{1 1} 用例 _{1 2} . . .)
(Y' for X')	(用例 _{2 1} 用例 _{2 2} . . .)
(Y' at X')	(用例 _{3 1} 用例 _{3 2} . . .)
:	
	. . . (3-2-1 4)

Figure 1 illustrates the hierarchical structure of the sentence "the conference registration fee discount". The diagram shows two possible interpretations of the sentence structure, both resulting in a total distance of 0.34.

The top structure shows the sentence as (X of Y), where X is "the conference" and Y is "the registration fee discount". The distance between the root node (X of Y) and the child node X is 0.17. The distance between the root node (X of Y) and the child node Y is 0.17. The total distance is 0.34.

The bottom structure shows the sentence as (X of Y), where X is "the conference" and Y is "the registration fee". The distance between the root node (X of Y) and the child node X is 0.17. The distance between the root node (X of Y) and the child node Y is 0.17. The total distance is 0.34.

図4 (（会議の登録費）の割引)

4. TDMT による日英翻訳システム

TDMT による日英翻訳システムは、TDMT の原理の有効性を確かめる目的で作られたものである。システムの翻訳処理部分はすべて Common Lisp の仕様にもとづいて書かれている。現在、“国際会議に関する問い合わせ”の入力に対応した知識を持っており、高速な翻訳を行なうことができる。

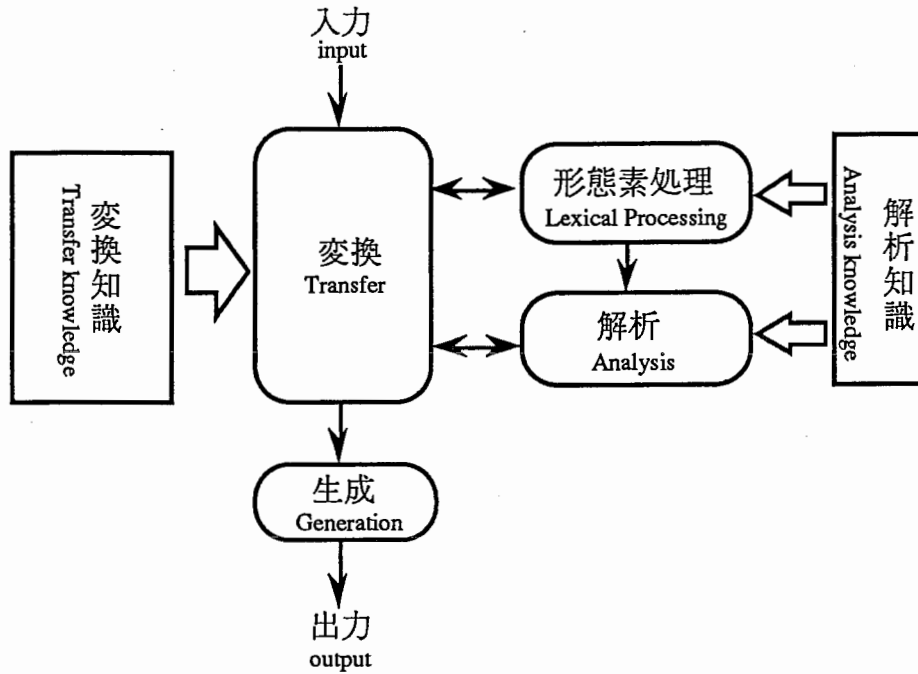


図5 TDMT 日英翻訳システム

4-1. 翻訳処理の制御

本システムは逐次計算機上で稼働するように作られているので、本来 TDMT がめざしている“必要な処理のみを駆動する”といった動作を必ずしも行なうことができない。そこで、できるだけ効率的に処理を行なうために次のようなメカニズムを採用することにした。

- ・ 先ず、変換知識のみを適用して翻訳を試みる。これに失敗した場合は解析知識を適用して再度、翻訳を試みる。
- ・ できるだけ浅い知識レベルの知識から適用される。すなわち、ストリングレベル、パターンレベル、文法レベルの順に適用を試みる。
- ・ 入力文の最も高いレベルの単位から、より低いレベルの単位へ変換処理が進められる。すなわち、文の基本構成を捉える単位の知識から、順次、節、句の単位へ知識が適用されていく。

4-2. TDMT での変換、解析知識について

変換知識とは原言語表現を目的言語表現に写像する知識のことで、TDMT システムはストリングマッチング、パターンマッチング、文法的マッチングに対応したレベルのものを持っている。本システムではストリングレベルの変換知識は日英辞書、パターンレベルの変換知識はパターンデータ、文法レベルの変換知識は POS データの形で記述されている。変換モジュールはこれらの変換知識を用いて変換木構造を作成しどのような変換を行なうべきかを決定する。また、パターンレベルおよび文法レベルの変換知識にはプライオリティーがついており、その適用が制限されるようになっている。

これに対し、解析知識は原言語表現を修正原言語表現に写像する知識で、形態素レベル、部分レベル、全体レベルのものがある。形態素レベルの解析知識は Lexical-transformation データ、部分レベルの解析知識は Local-transformation データ、全体レベルの解析知識は Total-transformation データとして記述されている。以下にそれぞれの働きを説明する。

- ・ Lexical-transformation データ
形態素レベルの解析データでおもに形態素の合成、および修正を行なう。
- ・ Local-transformation データ
部分レベルの解析データで連体修飾などの文法マーカーを挿入する。
- ・ Total-transformation データ
全体レベルの解析データで語順の正規化などを行なう。

Lexical-transformation、Local-transformation の解析処理は共に入力データが動的に変更されていき、処理結果はただ一つだけとなる。また、このため解析知識を変更、追加する場合、無限ループに陥るような解析知識を記述しないように注意しなければならない。

Total-transformation の解析処理の結果は複数に分裂する場合がある。また、すべての解析知識にはプライオリティーがついており、その適用が制限されるようになっている。

4-3. プライオリティー制御

パターンレベルの変換知識は文の基本構成を捉えるためのグローバルな単位のものから、語句を捉えるためのローカルな単位のものまでがあるが、それらを適当な順序でマッチングさせ、無駄なマッチングをさけるために、変換モジュールはプライオリティー制御を行なっている。

すべてのパターン変換知識はそれぞれプライオリティーを持っており、変換モジュールが指し示したプライオリティーに該当しない知識は適用されない。変換モジュールは通常、静的なプライオリティー制御を行なっており、優先度の高いプライオリティーから、より低いプライオリティーへと順次マッチング処理を進め、変換木を構成していく。この静的なプライオリティー制御は文から節、句といった各レベルに対応する知識の適用に関しておおまかなガイドの役割を果たす。また、適用された知識によって動的にプライオリティー制御を行なうこともできる。この動的なプライオリティー制御は名詞句レベルの変換に対しての微妙な制御に有効である。

なお、文法レベルの変換知識、解析処理についても静的なプライオリティー制御は行なわれており、ある程度、段階的な解析が行なわれるようになっている。また、これらのプライオリティー制御はそれぞれの知識ごとに独立して行なわれる。例えば、変換モジュールではパターンレベル変換知識と文法レベル変換知識のプライオリティーは互いに関係しない。

5. TDMT システムによる翻訳のしくみ

ここでは、TDMT 日英翻訳システムによる翻訳のしくみを示す。

変換知識は再帰的に適用され、入力はさらに下位の変換知識とマッチングが取られる。こうして、末端部分まですべてマッチングが成功すると翻訳が可能となる。変換知識だけでは翻訳不可能な場合は解析処理を通した後、再び変換処理が行なわれる。

なお、以下に示した知識記述例は TDMT のしくみを説明するための概念的なもので、実際の知識記述とは異なるので注意のこと。

5-1. スtring マッチング

String レベルでは、用例距離計算を行なうことなしに原言語表現から目的言語表現に変換される。String レベルの形態素は表記形と標準形の情報を持っているが、初め表記形によるマッチングが行なわれ、それがなければ標準形によるマッチングが行なわれる。本システムでは日英辞書にこれらの写像関係が記述されている。なお、英語形態素としての属性や文脈条件は、この後に生成処理で参照される英語属性辞書により記述されている。

変換例)

もしもし => hello

5-2. パターンマッチング

この場合のパターンマッチングとは表層上の語のパターンのみでマッチングを行なうことを意味する。パターンレベルの変換知識の原言語表現は1つ以上の変項を持っている。この変項に結びつけられた語（または語句）がこのパターンでの入力とみなされる。入力は用例距離計算に用いられ、目的言語表現が選出される。

知識記述例：

X を お願い し ます	=>	may I speak to X' (様 局)
		please give me X' (番号)
		please X' (確認する)
		could you reserve X' (部屋)

変換例：

会議事務局 を お願い し ます	=>	may I speak to the personal section
題目 を お願い し ます	=>	please give me the title

5-3. 文法マッチング

表層の情報だけでは翻訳が不可能な場合がある。そのため、文法的な属性によるマッチングが行なえる変換知識が用意されている。この知識では変項は存在しないが、文法属性の記述にマッチした語（または語句）が入力として扱われる。入力是用例距離計算に用いられ、目的言語表現が選出される。このとき、ストリングマッチングと同様のの手続きが呼び出され、目的言語に変換される。

知識記述例：

[普名詞 1] [普名詞 2] [普名詞 3] [数詞] =>
[数詞] ' [普名詞 3] ' [普名詞 2] ' [普名詞 1] ' (東京都 港区 新橋 1)

変換例：

大阪市 北区 茶屋町 2 3 => 23 Chaya-machi Kita-ku Osaka-shi

5-4. 変換モジュールにおける変換知識の適用

すでに述べたように、変換知識にはストリングレベル、パターンレベル、文法レベルに対応するものがあり、これらを再帰的に適用することにより内部的に変換木構造を作成する。この変換木構造が完成すれば、すべての部分の変換が可能となり、従って翻訳が可能となる。変換木構造が完全にできあがっても、その中に曖昧性があれば翻訳結果が複数に分裂することになり、また変換木構造が誤ったものであれば翻訳結果も誤ったものとなる。従って、変換モジュールがどのような手順で様々なマッチングを行なうかは翻訳結果に重大な影響をおよぼす。そこで、実際に変換モジュールがどのようにこれを行なっているかを以下に示す。

- [1] 文法マッチングによる変換を試みる。→可能であれば復帰する。
- [2] ストリングマッチングによる変換を試みる。→可能であれば復帰する。
- [3] パターンマッチングを行なう。マッチすれば変項に結び付いた部分を新たな入力として再帰的に変換を試みる。このとき、いくつかの変換知識に対してマッチすることがあるが、その場合はそれぞれについて再帰変換を試みる。→成功すれば復帰する。マッチするパターンを持った知識が1つも存在しない場合は変換（従って翻訳）が不可能で、該当する変換木構造は無効となる。

以上、[1] [2] [3] の手順を再帰的に行ない、完全に復帰すれば変換木構造が完成する。

5-5. 解析処理

構造的に複雑な文は変換知識だけでは翻訳できない。この場合に解析機構が駆動され、変換モジュールに必要な情報が送られる。

ここでいう解析処理とは文法規則による構文解析や意味解析ではなく、変換知識を適用したり、正しい翻訳を導くための情報を付加することである。

例えば日本語では格の順序は比較的自由であるためwh-case成分の位置は固定していないが、英語ではこれが文の様相を決定する重要な成分となり、その位置も制約を受ける。

例えば、wh-case成分”どのように”を含む次のような文がある。

参加料 は どのように お支払い に なり ます か . . . (5-5-1)

これに対しては次の解析知識が適用され、wh-case という解析のマーカースが挿入される。

知識記述：

どのように X => どのように wh-case X' (参加料 . . .)
. . . (5-5-2)

更に語順を正規化する解析知識が適用されると次の結果が返される。

どのように wh-case 参加料 は お支払い に なり ます か
. . . (5-5-1)'

これを再び変換モジュールに渡すと、解析マーカースに関する変換知識が駆動され、適正な翻訳が行なわれる。

how will you pay the attendance fee . . . (5-5-1)''

また、解析処理は上の例のように動的な操作を行なうので、その適用が無限に続くような解析知識は記述しないように注意する必要がある。なお、一つの文中で同じ解析知識がマッチする部分が2箇所以上ある場合は、文頭に近い部分から操作が実行される。また、プライオリティーのより高い解析知識から適用が試みられる。

ただし、文全体レベルのTotal-transformation 解析知識は、適用の仕方に複数の可能性がある場合、それぞれに応じた複数の結果を返す。

II. システム機能仕様

1. システム概要

本システムは Symbolics Lisp Machine 上で稼働する変換主導型機械翻訳 (TDMT: Transfer-Driven Machine Translation) の実験用モジュールである。

プログラムは大きくわけて翻訳処理部とユーザ・インターフェース部からなり、翻訳処理部はさらに形態素処理部、解析部、変換部、生成部に分かれる。

翻訳処理部では初め解析部を通さずに実行し、結果がでなければ解析部を通して再度実行する。なお、解析部、変換部ではそれぞれの知識ルールと意味コード辞書および日英変換辞書を、生成部では英語属性辞書、不規則変化動詞辞書を用いて処理を行なう。

ユーザ・インターフェースはウィンドウにより、マウス操作またはキー入力によりコマンドを選択すると、その実行結果でウィンドウが再表示されるようになっている。

以下に、本システムの動作環境などを示す。なお、下位システムとして MORPH (形態素解析システム) が必要である。

使用機器:	Symbolics Lisp Machine
動作環境:	Genera 8.0.1
使用言語:	Common Lisp (翻訳処理部)、 Symbolics Common Lisp (ユーザ・インターフェース部)
下位システム:	MORPH -pro (形態素解析システム)

2. 全体の構成

本システムはプログラムと知識ルールおよび各種の辞書により構成される。

プログラムは大きくわけて翻訳処理部とユーザ・インターフェース部からなり、翻訳処理部はさらに形態素処理部、解析部、変換部、生成部に分かれる。一方ユーザ・インターフェース部はウィンドウにより入出力を行なうようになっている。

知識ルールには解析部用のものと変換部用のものとがあり、これらが翻訳処理の中心的役割を果たす。

辞書には日英変換辞書、英語属性辞書、不規則変化動詞辞書があり、これらを参照しながら処理が進んでいく。

以下に日英 TDMT システムの構成図を示す。

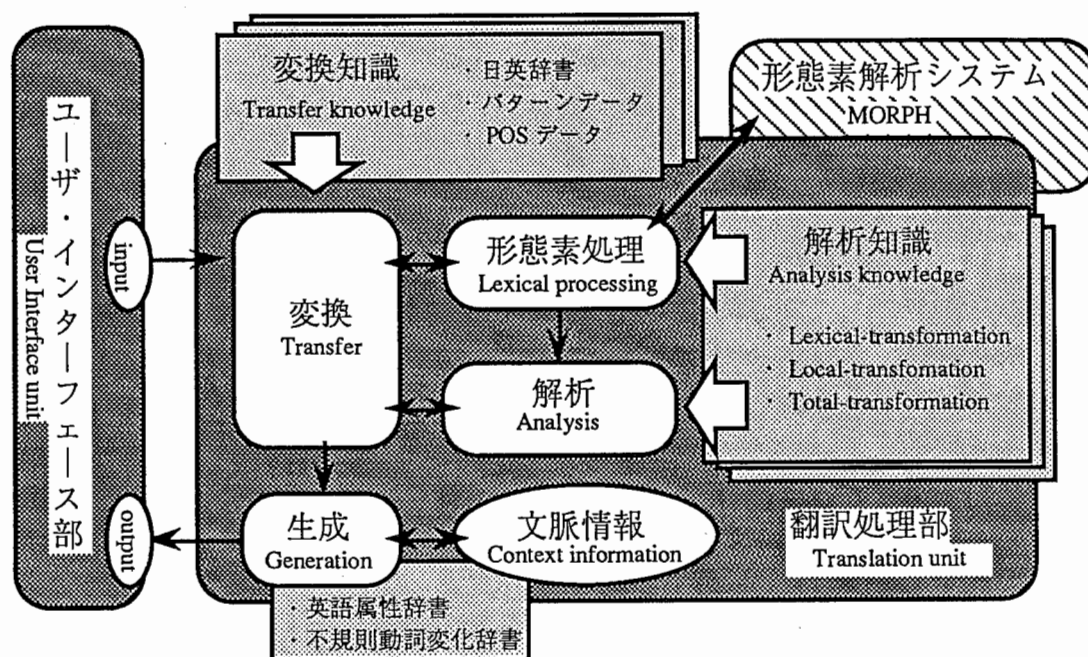


図6 日英 TDMT システムの構成

3. 翻訳処理部

3-1. 形態素処理部

・形態素解析処理

日本語の文字列を形態素に区切り、それぞれの形態素を `japanese-word` の構造体に置き換える。なお、形態素解析処理の前に日英辞書を参照し、文全体が登録されているもの（”もしもし”、”それではよろしくお願いします”等）であれば、解析処理を行わずに変換部にうつる。なお、この処理は形態素解析システム MORPH を使って行なわれる。

・形態素合成処理 (lexical-transformation)

知識ルール (`lexical-transformation.lisp`) を用いて複数の形態素を 1 つにまとめる合成や助詞補完、置換を行なう。

3-2. 解析部

・局所的文構造変換処理 (local-transformation)

知識ルール (`local-transformation.lisp`) を用いて変換駆動マーカー (`rentai`, `wh-case`等) の挿入を行なう。

・大域的文構造変換処理 (total-transformation)

知識ルール (`total-transformation.lisp`) を用いて日本文語順正規化を行なう。

なお、上記の変換駆動用解析マーカーは、生成で用いられる生成マーカーとは違い、日本語形態素と同様に変換処理でのマッチングに使われるものである。

3-3. 変換部

3-3-1. 変換部の概要

変換部では知識ルールおよび辞書を用いて日本語形態素列から英語形態素列への変換を行なう。内部的には Transfer 処理（構造変換）、POS 処理（特種日英変換）、english-string 処理（英語形態素列生成）に分かれる。

生成される英語形態素列は知識ルールから来ている symbol と日英変換によって得られた english-word の構造体とで構成されている。

3-3-2. 変換部の詳細

- ・ Transfer 処理（日本語構造解析）

ストリングマッチング、パターンマッチングによる変換。日本語形態素列から変換構造（transfer-structures）を生成する。変換知識を用いる。

- ・ POS 処理（特種日英変換）

文法マッチングによる変換。日本語構造の終端部（住所など）の処理。変換知識（pos.lisp）を用いる。

- ・ english-string 処理（英語格構造リスト生成）

変換構造から英語形態素の格構造リストへ変換する。

3-3-3. 変換知識のプライオリティ制御

(1) プライオリティ制御の概略

TDMTでは、入力文から次々に変換知識を探索し、トップ・ダウンに日本語木構造をつくる。その際に、ある変換知識が適用されると、その下では、通常、同じプライオリティ・グループか、より低いプライオリティ・グループの中から変換知識が選ばれる。

(プライオリティ・グループとはおなじプライオリティを持つ変換知識群をさす。)

しかし、もう少し細かいプライオリティ制御をしたい場合がある。その場合は、変換知識にプライオリティ制御の記述を加えると、任意のプライオリティに制御を移すことができる。

(2) プライオリティの拡張

細かいプライオリティ制御を可能にするため、プライオリティ・グループを条件により細分化できるように、プライオリティを拡張する。

まず、以下のようにメジャー・グループとマイナー・グループを定義する。

- a.) マイナー・グループはメジャー・グループに含まれる。
- b.) プライオリティについては、メジャーは整数、マイナーは非整数で表わす。

プライオリティ・グループ

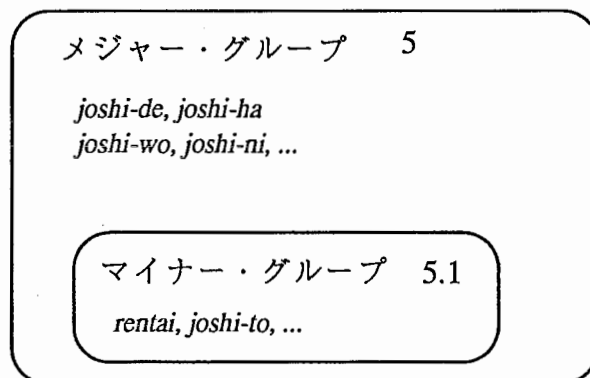


図7 プライオリティ・グループの包含関係

(3) 駆動法

- ・プライオリティ制御記述がない場合

マッチしたルールと同じか、それ以下のプライオリティへ進む。ここで、次のプライオリティに進むときは、必ずメジャー・グループを選ぶ。

- ・プライオリティ制御記述がある場合

そこで記述されているプライオリティ・グループへ進む。

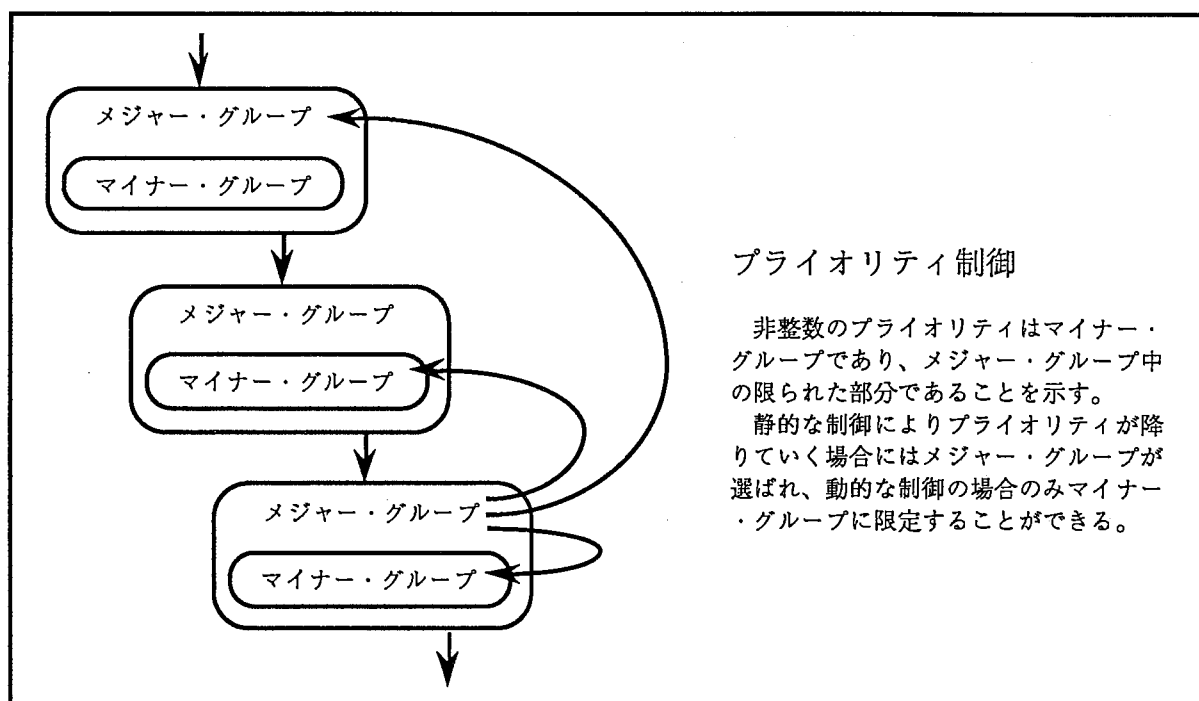


図8 プライオリティ制御の流れ図

3-4. 生成部

3-4-1. 生成部の概要

生成部では変換部からの出力の英語形態素列を変換していく。

この英語形態素列は list になっており、知識ルールから来ている symbol と日英変換によって得られた english-word の構造体とで構成されている。

3-4-2. 生成部の詳細

1) デフォルト格の処理

SUB, OBJ, IND-OBJ 格には、それぞれ *SUB, *OBJ, *IND-OBJ のデフォルト格があり、もし明示された格があればそれを優先する処理。

2) 語順整列処理

格マーカー、品詞情報、様相マーカーから語順を整列する処理。

3) 単語変形処理

変形マーカーによる単語の変形処理。(BE 動詞処理、三単現処理、冠詞処理、名詞句処理にはマーカーはないが、同じように扱う。)

これには以下のものがあり、この順序で適用される。

・ %TO-THATマーカー処理 (%TO-THAT)

英語形態素列の %TO-THATマーカーを "to" または "that" に変形する。

このとき、マーカーの後方に SUB があれば "that" に、なければ "to" に変形される。

・ be動詞処理

be 動詞については SUB 格を参照してその形を決めている。また、このあとの処理では、be 動詞は助動詞として扱われる。

・ 否定処理 (NOT+)

動詞を否定する処理。助動詞があればそれを否定形に変形し、普通動詞の場合は助動詞 do 等の否定形を挿入する。

・ 三単現処理

三人称単数現在形へ助動詞、動詞を変形する。

- ・ 名詞複数形処理 (PL+)
名詞を複数形に変形する処理。
- ・ 名詞句単数形処理 (SING+)
名詞句を単数形に変形する処理。
- ・ 形容詞、副詞変形処理 (EST+, ER+)
形容詞、副詞を最上級、比較級に変形する処理。
- ・ 冠詞処理
明示された冠詞がある場合、デフォルトの冠詞を打ち消す処理。
- ・ 名詞句処理
副詞、形容詞、名詞の相互作用により、デフォルト冠詞、複数形などの属性を伝播させる。
- ・ 所有格処理 (POSS+)
名詞を所有格に変形する。このあと、形容詞扱いにする。
- ・ 動詞変形処理 (PAST+, PP+, ING+)
英語形態素列中の動詞を過去形、過去分詞形、現在分詞形に変換する。
このときPAST+は助動詞に作用しようとするが、なければ動詞に作用する。また、PP+, ING+はこの後方の最も近いところにある動詞に作用する。
変換は、まず英語属性、次に不規則変化動詞辞書(conjugation.text)を調べ、該当するものがなければ語尾を直接変形することにより行なっている。

4. ユーザー・インターフェース

4-1. ウィンドウ

ウィンドウは9つのペインから構成されており、それぞれ次の内容を示している。

1) Title	TDMTの文字
2) Menu	コマンド・メニュー
3) TDMT Listener	TDMT専用のリスナー
4) Input	入力された日本語文
5) Outout	選択された英語文
6) Process	処理の途中経過
7) Distance calculation in Transfer	変換の際の距離計算結果
8) Distance calculation in Analysis	解析の際の距離計算結果
9) Other Candidates Output	他の英語文候補

操作はコマンドをMenu からマウスで選択するか、TDMT Listener にキー入力で直接打ち込むことによって行なう。

以下に実際のウィンドウ表示例を示す。

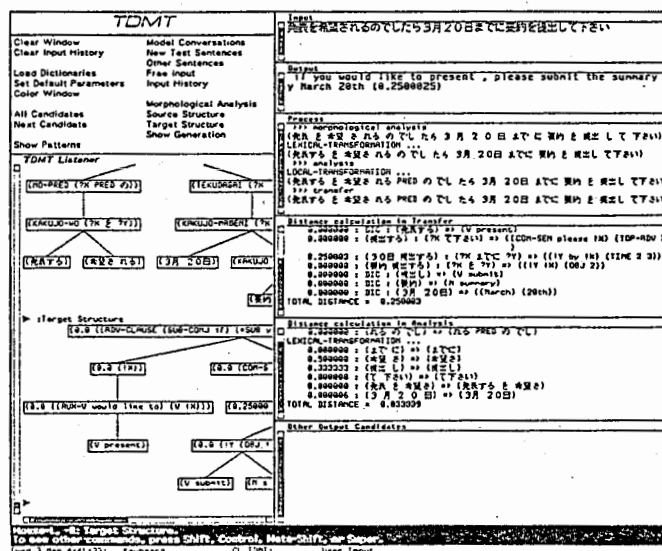


図9 ウィンドウ表示例

4-2. コマンド詳細

TDMT で使用できるコマンドは Menu のペインなどのマウス・クリック、または TDMT Listener からのキー入力により起動される。以下にそれぞれの働きと使用法を示す。

Clear Window

結果表示用のペインをすべて消去する。

Clear Input History

すでに入力された日本語文のストックを消去する。

Load Dictionaries

変換知識ルール及び辞書をロードする。

Set Default Parameters

デフォルトのプライオリティー値や距離計算の上限値といったデフォルトのパラメータをセットする。解析部と変換部は独立にコントロールできる。

ALL Candidates

すべての英語文候補の中から1つを選択する。初めの番号は順位を表す。

Next Candidate

現在選択されている英語文のすぐ次の候補を選択する。現在選択されている英語文が最後の場合は先頭の候補を選択する。

Model Conversations, New Test Sentences

会話例から選択された日本語文に対して翻訳を実行する。会話例はそれぞれモデル会話、新評価文。

Other Sentences

任意のファイルから入力文を選び、その翻訳を実行する。

Free Input

自由入力された日本語文に対して翻訳を実行する。

Input History

すでに入力された日本語文のストックから選択された日本語文に対して翻訳を実行する。(cf. Clear Input History)

Morphological Anarysis

入力された日本語文の形態素解析の結果を表示する。もとのウィンドウに戻るには <select>+0 とする。

Source Structure

入力された日本語文の構造木を表示する。

Target Structure

選択されている英語文の構造木を表示する。

Show Generation

生成部での処理内容を示す。

5. ファイル構成

日英 TDMT システムは関数ファイル群、辞書ファイル群、変換、および解析知識ルールファイル群からなる。Symbolics 上で、関数ファイル群はシステム TDMT の元に、その他は TDMT-DATA の元に管理されている。

以下論理ホスト名を "TDMT" とし、ディレクトリ、ファイルのパス名は論理パス名を使用する。

5-1. 関数ファイル群

(1) 翻訳処理部

ディレクトリ： TDMT:CODE;

ファイル： PACKAGE.LISP

DATA-STRUCTURES.LISP

DICTIONARY-TOOLS.LISP

PATTERN-MATCH.LISP

JAPANESE-STRUCTURE.LISP

ENGLISH-STRUCTURE.LISP

ENGLISH-SENTENCE.LISP

ANALYSIS-UTILS.LISP

ANALYSIS-DATA-STRUCTURES.LISP

ANALYSIS-MACROS.LISP

ANALYSIS-PROCESS.LISP

GENERATION.LISP

CONJUGATION-FOR-VERBS.LISP

INTERACTION-FOR-WORDS.LISP

MAIN.LISP

TEST.LISP

(2) ユーザ・インターフェース部

ディレクトリ： TDMT:WINDOW;

ファイル： GLOBALS-FOR-WINDOW.LISP

WINDOW.LISP

DYNAMIC-LISP-LISTENER.LISP

5-2. 辞書ファイル群

ディレクトリ： TDMT:DIC;

ファイル： CODE-TO-STRING-ALIST.LISP
JAPANESE-TO-ENGLISH.LISP
ENGLISH-ATTRIBUTION.LISP
CONJUGATION.LISP

5-3. 変換知識ルールファイル群

ディレクトリ： TDMT:P-DATA;

ファイル： *.LISP

5-4. 解析知識ルールファイル群

ディレクトリ： TDMT:A-DATA;

ファイル： *.LISP

III. プログラム説明

1. 主要な関数

次に主要な関数の説明を行なう。

`translate japanese-string &optional (analysis-first *analysis-first*)`

日本語文字列 *japanese-string* を英語文字列に変換して、距離計算の値とともにリストにして返す。*analysis-first* を `t` にすると解析部を省かずに処理する。翻訳処理部のトップ関数。

`morph sentence`

日本語文字列 *japanese-string* を形態素解析して、日本語形態素列を返す。ここでは、A-DATA のうち、`define-lexical-transformation` で定義された形態素処理用の解析知識が使われる。

`analysis morph-structures&analysis`

日本語形態素列に対し形態素変形処理を加え、結果の形態素列を返す。ここでは、A-DATA のうち、`define-local-transformation` , `define-tocal-transformation` で定義された解析知識が使われる。

`transfer collect-morph-structures&analysis &optional (clear-db t)`

日本語形態素列から変換構造を作る。このとき、P-DATA として定義された変換知識が使われる。また、A-DATA のうち `define-pos` で定義された変換知識も使われる。

`generation english-sentences`

変換構造から生成処理を行ない、英語文字列を作る。

2. モジュール説明

次に各モジュールの説明を行なう。

2-1. 翻訳処理部

翻訳処理部は Common Lisp の仕様で作成されている。

モジュール:

data-structures

data-structures.lisp

このファイルは大域変数、構造体を定義している。

大域変数:

ウィンドウ用:

message-string

インプット・ヒストリーのデフォルトメッセージ。

stock-input-data

インプット・ヒストリー。

変換部用:

tdmt-patterns

変換知識ルール名のシンボルと変換知識ルール構造の写像を管理するハッシュテーブル。

pattern-db

変換知識ルール内で使用される文字とその文字を使用している変換知識ルールのリストの写像を管理するハッシュテーブル。

default-priority

変換知識ルールのデフォルトのプライオリティー。

default-distance-limit

変換知識ルールにおける距離計算の際の上限値。

penalty-for-not-exact-matching

距離計算においてexact-matchingしなかったときの加算値。

illegal-word-list

"が"、"を"、"の"、"に"などの助詞のリスト。

morphed-sentence

形態素解析した結果の日本語形態素のリスト。

japanese-structures

日本語構造の候補のリスト。

english-structures

英語構造の候補のリスト。

english-sentences

英語文構造の候補のリスト。

morph-analysis

解析処理の全履歴。

debug

デバッグ・モードのスイッチ。

record-process

途中経過を記録するかどうかのスイッチ。

analysis-first

解析部を初めから実行するかどうかのスイッチ。

selected-japanese-structure

候補ごとに分割された日本語構造のリスト。

selected-english-structure

候補ごとに分割された英語構造のリスト。

生成部用：

***comparative-degree-marker* 'ER+**

比較級変化のマーカ。

***superative-degree-marker* 'EST+**

最上級変化のマーカ。

***v-past-t-marker* 'PAST+**

過去形変化のマーカ。

***v-past-part-marker* 'PP+**

過去分詞変化のマーカ。

***v-pres-p-marker* 'ING+**

現在分詞変化のマーカ。

***not-marker* 'NOT**

否定のマーカ。

***plural-marker* 'PL+**

複数形変化のマーカ。

***singular-marker* 'SING+**

単数形変化のマーカ。

***possessive-marker* 'POSS+**

所有格変化のマーカ。

***to-or-that-marker* '%TO-THAT**

TO不定詞、またはTHAT節の未定マーカ。

***normal-modality* 'NORMAL**

平叙文の様相マーカ。

modality-marker-list '(WH-Q YN-Q TAG-Q COM-SEN GREETING NORMAL)

様相マーカーのリスト。他の生成マーカーと区別するために使われる。

前から順に、WH疑問文、Yes-No疑問文、付加疑問文、命令文、挨拶、平叙文。

no-rearrangement-case-list '(NP INF TO AT WITH FOR OF)

内部では語順整列処理を行なわない格のリスト。

generation-marker-list

生成マーカーのリスト。様相マーカーも含む。格マーカーと区別するために使われる。

subject-case 'SUB

主格のマーカー。

case-list

格マーカーのリスト。語順整列に関わる品詞も含む。

default-cases

明示格とデフォルト格とのドット・ペアのリスト。

general-verb 'V

一般動詞の品詞。

auxiliary-verb 'AUX-V

助動詞の品詞。

be-verb

BE動詞の品詞。

pos-list

品詞のリスト。他の格マーカーと区別するために使われる。

sentence-mode-rearrange-list

様相マーカーごとの語順整列を記述したリスト。

出力：

output-nth

出力される候補の番号。

構造体：

pattern

name

priority

source

next-priority

example-to-target-hash

targets

変換知識ルール構造。name は変換知識ルール名。priority は値の小さい方が高い。

source は日本語の変換知識ルールが入る。next-priority はプライオリティ制御で使用される。targets は target 構造のリストが入る。

target

- target
- alternative
- examples
- selections

英語ターゲット構造。target は英語のパターンが入る。alternative は英語パターンを格リストに展開したものである。examples は用例のリストが入る。selections はこのパターンが選択された場合の単語訳し分けのリストが入る。

japanese-structure

- distance
- sub-sentence
- pattern
- form
- substructures
- analysis

日本語構造。distance は距離計算の結果、sub-sentence はこの構造で表現されたものの部分形態素列。pattern は変換のパターン構造。form はパターン構造の source 部分。substructures は変項 ?x, ?y などに対応する部分の日本語構造のリスト。analysis はトップの日本語構造で使われる処理過程履歴の構造。

english-structure

- distance
- matched-example
- from
- target
- form
- substructures

英語構造。日本語構造から作られる。distance はその構造の距離。matched-example はルールが適合した時に最も距離の近かった用例。from はその英語構造が作られた元の日本語構造。target は英語変換パターン。form は英語のパターン。substructures は form の論理変数を埋める部分英語構造。これは、論理変数とそれに対応する英語構造のリストのリストからなる。

english-sentence

- total-distance
- logical-variable
- from

form
substructures
analysis

英語文構造。基本的には english-structure と同じものだが、株分けされて枝がなくなっている。total-distance はこれより以下の構造の距離の合計が入る。logical-variable はその文に対応する論理変数が入る。form は英語のパターン。from はその英語文構造が作られた元の英語構造。substructures は form の論理変数を埋める部分英語文構造。これは、論理変数とそれに対応する英語文構造のリストのリストからなる。

japanese-word

word
reg-exp
pos
conjugated-form
sem-code

日本語の形態素構造。形態素解析された情報。word は入力された語。reg-exp はその正規形。pos はその品詞。conjugated-form は活用形。sem-code は意味コード。

english-word

word
reg-exp
pos
info
rule

英語の形態素構造。日英変換された情報。word は入力された語。reg-exp はその正規形。pos はその品詞。info は属性情報。rule は訳語選択の条件。

conjugation

entry
v-past-t
v-past-part
v-pres-p

動詞活用辞書の内部構造。原形に対する過去形、過去分詞形、現在分詞形のセット。

macros.lisp

このファイルは主に変換部で使用される基本的なマクロ、関数を定義している。

マクロ：

`define-pattern name priority japanese-pattern arrow &rest english-patterns`

変換知識ルールを定義するマクロ。

モジュール：

*dictionary-tools**dictionary-tools.lisp*

このファイルは辞書および知識ルール読み込み、および、辞書引きのための大域変数、関数を定義している。

大域変数：

`*japanese-to-english-dic-file*`

日英変換辞書のファイル。

`*japanese-to-english-dic*`

日英変換辞書を取り込むハッシュテーブル。

`*english-attribution-dic-file*`

英語属性辞書のファイル。

`*english-attribution-dic*`

英語属性辞書を取り込むハッシュテーブル。

`*conjugation-dic-file*`

不規則変化動詞辞書のファイル。

`*conjugation-dic*`

不規則変化動詞辞書を取り込むハッシュテーブル。

`*pattern-data-files*`

変換知識ルールのファイル群。

`*analysis-data-files*`

解析知識ルールのファイル群。

マクロ：

`get-english japanese &optional (selections nil)`

日英辞書を引くためのマクロ。

`define-dic dic-name &optional (remake nil) &rest patterns`

日英辞書、英語属性辞書を定義するためのマクロ。

関数：

`make-japanese-to-english-dic &optional (remake t)`

日英変換辞書を読み込む。

`make-english-attribution-dic &optional (remake nil)`

英語属性辞書を読み込む。

`make-conjugation-dic &optional (remake t)`

不規則変化動詞辞書を読み込む。

`load-all-pattern-data`

変換知識ルールを全て読み込む。

`load-all-analysis-data`

解析知識ルールを全て読み込む。

`load-dictionaries &optional (load-patterns t) (remake nil)`

すべての辞書、ルールを読み込む。

モジュール:

pattern-match

pattern-match.lisp

このファイルはパターンマッチのための関数を定義している。

関数:

pattern-match *pattern sentence*

pattern と *sentence* をパターンマッチさせ、可能な組み合わせをリストにして返す。

例) (pattern-match '(?X の ?Y) '(A の B の C))
->
(((?X A) (?Y B の C)) ((?X A の B) (?Y C)))

モジュール:

japanese-structure

japanese-structure.lisp

このファイルは変換処理のための関数を定義している。

関数:

collect-japanese-structure-candidates *morph-structures&analysis*

日本語形態素列から日本語構造の候補リストを返す。*morph-structures&analysis* はシンボルまたは
japanese-word 構造体からなるリストと解析履歴構造体の対。

モジュール:

english-structure

english-structure.lisp

このファイルは変換処理のための関数を定義している。

関数:

transfer-japanese-to-english-all *japanese-structures*

日本語構造の候補リストから英語構造の候補リストを返す。*japanese-structures* は日本語構造の候補リスト。

モジュール：

english-sentence

english-sentence.lisp

このファイルは変換処理のための関数を定義している。

関数：

make-english-sentence-all *english-structures*

英語構造の候補リストから英語文構造の候補リストを返す。*english-structures* は英語構造の候補リスト。

make-english-case-expression *english-sentences*

英語文構造の候補リストを受け取り、それぞれを格構造リストの形に変換して返す。*english-sentences* は英語文構造の候補リスト。

モジュール:

analysis-utils

analysis-utils.lisp

このファイルは主に解析部で使用する基本的な関数を定義している。

関数:

make-macro-name *type*

type により各解析処理を定義するマクロ名のシンボルを作り、返す。

make-remover-name *type*

type により各解析処理の定義を解除する関数名のシンボルを作り、返す。

モジュール:

analysis

analysis-data-structures.lisp

このファイルは主に解析部で使用する大域変数、構造体を定義している。

大域変数:

default-analysis-priority

解析知識ルールのデフォルトのプライオリティー。

default-analysis-distance-limit

解析知識ルールにおける距離計算の際の上限値。

pos-pattern-table

pos 処理の知識ルール名のシンボルと知識ルール構造の写像を管理するハッシュテーブル。

local-transformation-pattern-table

local-transformation 処理の知識ルール名のシンボルと知識ルール構造の写像を管理するハッシュテーブル。

total-transformation-pattern-table

total-transformation 処理の知識ルール名のシンボルと知識ルール構造の写像を管理するハッシュテーブル。

lexical-transformation-pattern-table

lexical-transformation 処理の知識ルール名のシンボルと知識ルール構造の写像を管理するハッシュテーブル。

synthesized-morph-table

lexical-transformation 処理で形態素合成を行なう合成手続きのルール名シンボルとその構造の写像を管理するハッシュテーブル。

pos-pattern-list

pos 処理の知識ルール名の優先順位リスト。

local-transformation-pattern-list

local-transformation 処理の知識ルール名の優先順位リスト。

total-transformation-pattern-list

total-transformation 処理の知識ルール名の優先順位リスト。

lexical-transformation-pattern-list

lexical-transformation 処理の知識ルール名の優先順位リスト。

pos-db

pos 処理の結果を蓄えるハッシュテーブル。

構造体：

analysis-pattern-group

name

type

priority

patterns

解析知識ルール構造。name は解析知識ルール名。type は解析処理の種別。
priority は値の小さい方が高い。patterns は analysis-pattern 構造のリスト。

analysis-pattern

source

targets

source は日本語の解析知識ルールが入る。targets は target 構造のリストが入る。

analysis-process

process

local-transformation

total-transformation

lexical-transformation

pos

解析処理における履歴。process は翻訳処理の全体的な履歴。以下は各解析処理の部分的な履歴。

analysis-macros.lisp

このファイルは主に解析部で使用されるマクロなどを定義している。

マクロ：

define-analysis-pattern-group type

type により各解析知識ルールを定義するマクロを定義するマクロ。

define-pos name priority &rest patterns

pos 処理用の知識ルールを定義するマクロ。*name* はルール名。*priority* はプライオリティー。*patterns* はルール本体。

define-local-transformation name priority &rest patterns

local-transformation 処理用の知識ルールを定義するマクロ。*name* はルール名。*priority* はプライオリティー。*patterns* はルール本体。

define-total-transformation name priority &rest patterns

total-transformation 処理用の知識ルールを定義するマクロ。*name* はルール名。*priority* はプライオリティー。*patterns* はルール本体。

define-lexical-transformation name priority &rest patterns

lexical-transformation 処理用の知識ルールを定義するマクロ。*name* はルール名。*priority* はプライオリティー。*patterns* はルール本体。

define-morph name arrange &rest morph-attributes

形態素合成の手続きを定義するマクロ。

analysis-process.lisp

このファイルは主に解析処理（*total-transformation* 処理を除く）のための関数を定義している。

関数：

analyze-sentence morph-structures type operated-patterns &optional (priority 0)

type により各解析処理（*total-transformation* 処理を除く）を行なう関数。

morph-structures は日本語形態素列。

local-transformation morph-structures

local-transformation 処理を行なう関数。*morph-structures* は日本語形態素列。

part-of-speech morph-structures

pos 処理を行なう関数。*morph-structures* は日本語形態素列。

lexical-transformation morph-structures

lexical-transformation 処理を行なう関数。*morph-structures* は日本語形態素列。

total-transformation.lisp

このファイルは *total-transformation* 処理のための関数を定義している。

関数：

total-transformation morph-structures

total-transformation 処理を行なう関数。*morph-structures* は日本語形態素列。

モジュール：

generation

generation.lisp

このファイルは生成処理のための関数を定義している。

関数：

generate-english english-strings

格構造リストから英語文生成処理を行なう。

generate-string english-case-list

生成格リストから英語文字列を作る。

conjugation-for-verbs.lisp

このファイルは生成部で使用される関数を定義している。

関数：

conjugation-for-verbs english-strings

動詞の活用、変形を行なう関数。*english-strings* は英語形態素列。

モジュール：

main

main.lisp

このファイルはトップレベルの関数を定義している。主要な関数の項参照。

2-2. ユーザ・インターフェース部

モジュール：

window

globals-for-window.lisp

このファイルは主にウィンドウで使用される大域変数を定義している。

大域変数

example-sentences-1 ~-10*

入力用の会話例サンプル。

command-window

コマンドメニューのウィンドウ。

input-window

入力された日本語文字列の表示用のウィンドウ。

output-window

選択された英語文字列の表示用のウィンドウ。

process-window

解析処理の履歴表示用のウィンドウ。

similarity-for-transfer-window

主に変換部での距離計算の結果表示用のウィンドウ。

similarity-for-analysis-window

主に解析部での距離計算の結果表示用のウィンドウ。

other-candidates-out-put-window

他の英語候補表示用のウィンドウ。

dynamic-lisp-listener.lisp

ここでは形態素解析結果の表示用ウィンドウを作成している。

関数：

make-dynamic-lisp-listener

形態素解析結果用のダイナミック・ウィンドウを作成する。

window.lisp

このファイルはウィンドウおよびその動作を規定するコマンド等が定義されている。

フレームワーク：

TDMT

- title
- command-menu
- command
- input
- output
- process
- similarity-for-transfer
- similarity-for-analysis
- other-candidates-output

ウィンドウを9個のペインに分割する。title はタイトル。command-menu はコマンドメニュー。command はコマンド入力用。input は入力日本語文字列表示用。output は選択された英語文字列表示用。process は処理の履歴表示用。similarity-for-transfer は変換部の距離計算結果表示用。similarity-for-analysis は解析部の距離計算結果表示用。other-candidates-output 他の英語文字列候補表示用。

関数：

clear-all-window

結果表示用のペインを全て消去する。

initialize-panes

各ペインのインスタンスを大域変数に取り込む。

run-tdmt-window *string*

翻訳処理部を起動する。日本語文字列 *string* を翻訳して結果をウィンドウに表示する。

display-output

選択された英語文字列に関する情報を表示する。

コマンド：

コマンドについては操作説明・ウィンドウおよびコマンド説明の項参照。

モジュール：

test

test.lisp

このファイルはテスト用の関数等が定義されている。

関数：

test-all

会話例サンプルを全てテストする。

test-all-file

会話例サンプルを全てテストし、結果をファイルに出力する。

IV. データ仕様

1. 知識ルール

1-1. 解析知識ルールの仕様

解析知識ルールには lexical-transformation, local-transformation, total-transformation, pos 用のものがある。pos 処理は変換部で行なわれるが、pos 用のルールは解析知識ルールの一種とする。

以下に解析知識ルールの仕様を示す。

1. 解析知識ルール外部データ構造

1.) lexical-transformation, local-transformation, total-transformation, pos 用外部データ詳細

```
<rule-group>::=(<define-name> <rule-name> <priority> <rule>*)
  <define-name>::=define-lexical-transformation |
    define-local-transformation |
    define-total-transformation |
    define-pos
  <rule-name>::=any symbol
  <priority>::=<number> (1,2,3,4,5) | ( ) (= *default-analysis-priority* = 3)
  <rule>::=(<source> => <operator-examples>*)
  <source>::=(<attribute-of-morph>*)
    <attribute-of-morph>::=(<item>. <attribute>) | a symbol of the variable such as ?X
    <item>::=a symbol of morph's item
    <attribute>::=a symbol of the value for morph's item
  <operator-examples>::=(<operator> (<example>*))
    <operator>::=a list of numbers or symbols or lists for operation
    <example>::=a list of symbols as reg-exp. for distance calculation
```

- ・データはファイルの中に入っているいくつかの <rule-group> から構成される。
- ・外部データ構造の形式は基本的に共通であるが、プロセスの種類によって rule を追っていく順序が若干異なるので、注意が必要である。
- ・外部データを内部データに取り込む際には、priority, rule-group, rule の順に優先順序の強さを持ち、特に rule-name, rule は配置された順に暗黙の優先順序をもっている。
- ・<source> の中に現われる変項 ?X等は total-transformation の処理では任意の個数の形態素を表すが、その他の処理では 1 個の形態素を表す。
- ・<source> の要素数と <example> の要素数は一致していなければならない。

- ・ <operator> は 1 個の list で内部に許されるものは次の 3 つである。
 - number: マッチングした形態素の並びに対応する番号。
並べかえに用いる。
 - symbol: 合形成態素名(lexical-transformation)、埋め込まれる flag
(local-transformation、total-transformation) などの symbol。
 - list: 形態素合成用の list。この内部では number と symbol
が許される。詳しくは lexical-transformation の項参照。
- ・ <operator> に現われる number は <source> の要素の順に対応している。
- ・ 各プロセスごとの注意事項は 2.) に示す。

2.) 解析知識ルール外部データの具体例と注意事項の詳細

a.) lexical-transformation の処理 (形態素合成、助詞補完、置換)

```
(define-lexical-transformation sahen-1 ()
  (( (:HINSHI . サ変名詞) (:HINSI . 補助動詞) )
    => ( (sahen-1-a) ((参加 する)) )
      ( (sahen-1-b 2) ((参加 できる)) )
  )
)
(define-morph sahen-1-a (1 2)
  (:HINSHI . 1) (:KATSUYOU-TYPE . 2) (:SEM-CODE . (1 2)))
(define-morph sahen-1-b (1 する)
  (:ALL-COPY . 1))
```

または、

```
(define-lexical-transformation sahen-1 ()
  (( (:HINSHI . サ変名詞) (:HINSI . 補助動詞) )
    => ( ((& 1 2 (:HINSHI . 1) (:KATSUYOU-TYPE . 2) (:SEM-CODE . (1 2))))
      ((参加 する)) )
      ( ((& 1 する (:ALL-COPY . 1)) 2) ((参加 できる)) )
  )
)
```

- ・ lexical-transformation の処理ではすべての rule を順に調べていき、逐次変換していく。
- ・ 上記のように <operator> に合形成態素名を埋め込むことで define-morph での定義が展開され、形態素合成が行なわれる。
- ・ 品詞等の形態素情報の要素は number のときはそれが指す形態素からの情報を引用し、symbol のときはそのものが使われる。
- ・ :WORD、:REG-EXP についてはそれぞれの要素をアペンドし自動的に作られる。
- ・ :KATSUYOU-TYPE については list にして書くことにより、複数の形態素からの情報を合わせ持つこともできる。
- ・ :ALL-COPY は :HINSHI、:KATSUYOU-TYPE、:SEM-CODE をまとめてコピーする。
(これらの要素の初期値は nil となっている。)

b.) local-transformation の処理 (各種flag (rentai, wh-case等) 挿入)

```
(define-local-transformation rentai-1 2
  ((((:KATSUYOU-TYPE . 連体形)) (:HINSHI . 普通名詞)))
  => ((1 rentai 2) ((いる 口座番号) (扱う 話題) (れる 論文) (できる ホテル))
    ((1 2) ((言う 内容) (れる 場合) (た 者) (する 予定) (た 場合)))
  )
  ((((:KATSUYOU-TYPE . 連体形)) (:HINSHI . 形容名詞)))
  => ((1 rentai 2) ((為る 広汎)))
  )
)
```

・ local-transformation の処理ではすべての rule を順に調べていき、逐次変換していく。

c.) total-transformation の処理 (語順正規化等)

```
(define-total-transformation taiou-1 2
  ((((:HINSHI . 副詞)) ?X ((:REG-EXP . ない)))
  => ((1 3 2) ((全然 分る ない)))
  )
  ((((:HINSHI . 副詞)) ?X ((:REG-EXP . ない)) ?Y)
  => ((1 3 2 4) ((全然 分る ない が)))
  )
)
```

・ total-transformation の処理ではすべての priority の順に rule-group を調べ逐次変換していく。

・ 1 つの rule-group の中では 1 rule しか採用せず、次の rule-group へ移っていく。

d.) pos の処理 (part-of-speech)

```
(define-pos pos-4 0
  ((((:HINSHI . 接頭辞)) (:HINSHI . 普通名詞)))
  => ((2) ((御 名前) (御 住所)))
  )
)
```

・ pos の処理ではすべての rule を順に調べていき、*japanese-structures* を変換していく。

・ 処理方式は*japanese-structures*の末端部分を調べ、複数の形態素があるものについてpos の処理を行なう。このときpos の処理を通過したにもかかわらず、全く処理がされなかった場合 (処理結果がもとと同じになった場合を除く) は、*japanese-structures*のその部分に関する枝を削除する。

2. 解析知識ルール内部データ構造

```
*lexicaal-transformation-pattern-list*::=<global-variable>      (: for lexical-transformation)
*local-transformation-pattern-list*::=<global-variable>         (: for local-transformation)
*total-transformation-pattern-list*::=<global-variable>         (: for total-transformation)
*pos-pattern-list*::=<global-variable>                           (: for pos)
```

```
<global-variable>::=<(rule-name)>*
```

```
<rule-name>::=<atom> the key of each hash-tables
```

```
*lexicaal-transformation-pattern-table*::=<hash-table-for-analysis> (: for lexical-transformation)
  *morph-pattern-table*::=<hash-table-for-definition-of-morph>      (: for lexical-transformation)
*local-transformation-pattern-table*::=<hash-table-for-analysis>    (: for local-transformation)
*total-transformation-pattern-table*::=<hash-table-for-analysis>    (: for total-transformation)
*pos-pattern-table*::=<hash-table-for-analysis>                      (: for pos)
```

```
<hash-table-for-analysis>::=<(name)=><analysis-pattern-group>*
```

```
<analysis-pattern-group>::=#(<name> <type> <priority> <patterns>) the structure
```

```
<name>::=<rule-name>
```

```
<type>::=:lexical-transformation !
```

```
          :local-transformation !
```

```
          :total-transformation !
```

```
          :pos
```

```
<patterns>::=<analysis-pattern>*
```

```
<analysis-pattern>::=#(<source> <targets>) the structure
```

```
<targets>::=<target>*
```

```
<target>::=#(<alternative> <examples>) the structure
```

```
<alternative>::=<operator>
```

- ・ 内部データはプロセスの種類ごとにグローバル変数に分類されている。
- ・ データの実体は各 <rule-name> を key とした hash-table の中に入っている。
- ・ <priority> は外部データで () となっているものは () のまま取り込んでいる。
実際に参照されるときには *default-analysis-priority* の値が用いられる。
- ・ *morph-pattern-table* には lexical-transformation で使用される合形成態素名の展開形が入っている。

1-2. 変換知識ルールの仕様

以下に変換知識ルールの仕様を示す。

1. 変換知識ルール外部データ構造

1.) 変換知識ルール外部データ詳細

```

<rule-group>::=<define-name> <rule-name> <priority> <rule>
<define-name>::=define-pattern
<rule-name>::=any symbol
<priority>::=<number> (1,2,3,4,5,6) ! ( ) (= *default-pattern-priority* = 4)
<rule>::=<source> => <targets>*
<source>::=<item-of-japanese-morph>*<priority-controls>*<[] means omissible.
<item-of-japanese-morph>::=<item> ! ?<variable>
<item>::=a symbol of japanese morph's word
?<variable>::=a symbol of the variable such as ?X
<priority-controls>::=<variable><control-priority>
:<variable>::=a symbol of the variable such as :X
<control-priority>::=<number> (1,2,3,4,5,6)
<targets>::=<target><case-infomation>* (<example>*) (<selection-je>*)
<target>::=<item-of-english-morph>*
<item-of-english-morph>::=<item> ! !<variable> !<generation-marker>
<item>::=a string of japanese morph's word
!!<variable>::=a symbol of the variable such as !X
<generation-marker>::=a symbol of the generation marker
<case-infomation>::=<case-name> <case-expander>*
<case-name>::=a symbol of case name
<case-expander>::=a number of target term
<example>::=a list of symbols as any japanese reg-exp. for distance calculation
<selection-je>::=a set of local japanese-to-english-dic

```

- ・データはファイルの中に入っているいくつかの <rule-group> から構成される。
- ・<source> の中に現われる変項 ?X等は任意の個数の形態素を表す。
- ・変項 数と <example> の要素数は一致していなければならない。

2.) 変換知識ルール外部データの具体例

```
(define-pattern dekimasu ()
  (?x でき ます)
  =>
  (((("you" "will" "be able to" !x)(*sub 1)(aux-v 2)(be-v 3))
    ((会う) (参加) (利用))
   )
   (((("we" "will" "be able to" !x)(*sub 1)(aux-v 2)(be-v 3))
    ((準備) (送る) (用意))
   )
   (((("you" "can" "make" !x)(*sub 1)(aux-v 2)(v 3))
    ((予約))
   )
  )

(define-pattern kakujo-ni 5
  (?x に ?y) (:x 5.1)
  =>
  . . . . .
  ((((!y !x) (ind-obj 2))                                     ::: 英語パターン
   ( . . . . .                                             ::: 用例リスト
    (其方 伺う)
    . . . . .
   )
   ((y ("伺う" ("visit" "v")))                               ::: 訳し分けリスト
    (x ("彼" ("him" "pro"))
      ("彼女" ("her" "pro"))
      . . . . .
     )
   )
  )
  . . . . .
)
```

2. 変換知識ルール内部データ構造

tdmt-patterns::=<hash-table-for-patterns>

<hash-table-for-patterns>::=(<rule-name>=><pattern>)*

<pattern>::=#(<name> <priority> <source> <next-priority> <targets>) is the structure

<name>::=<rule-name>

<source>::=(<symbol-of-morph>*)

<symbol-of-morph>::=a list of japanese morph's words or variables such as ?X

<next-priority>::=(<priority-controls>*)!nil

<targets>::=(<target>*)

<target>::=#(<target> <alternative> <examples><selections>) is the structure

<target>::=a list of english morph's words or variables such as !X

<alternative>::= a list of expanded target with case-expanders

<examples>::=lists of examples

<selections>::=list of local japanese-to-english-dic

- ・データの実体は各 <rule-name> を key とした hash-table の中に入っている。
- ・<priority> は外部データで () となっているものは () のまま取り込んでいる。
実際に参照されるときには *default-pattrearn-priority* の値が用いられる。

2. 辞書

2-1. 日英変換辞書

日本語形態素を英語形態素に変換する際に用いられる。品詞情報は省略可能（注）。

書式：

```
(define-dic 'japanese-to-english-dic t  
  (japanese-string (english-string pos))  
  ..... )
```

正規形日本語文字列 *japanese-string* は英語文字列 *english-string* に変換される。
pos は品詞。

例：

```
(define-dic 'japanese-to-english-dic t  
  .....  
  ("振込む" ("transfer" "v"))  
  .....  
)
```

注) 日本語形態素から英語形態素の品詞が確定できる場合は、品詞情報を記述したほうがよいが、様相など他の条件によって確定できない場合は、品詞情報をあえて省略して英語属性辞書で訳し分けるほうがよい。

2-2. 英語属性辞書

英語形態素の属性を得るために用いられる。様相などによる訳し分けの情報も記述できる。

書式：

```
(define-info info-name ()
  attribution-list)
.....
(define-dic english-attribution-dic t
  (english-string (pos word reg-exp (attribution-info) condition-list))
  ..... )
```

define-infoはデフォルト属性を定義する。attribution-listはデフォルト属性のリスト。english-stringは属性を引くときのエントリー。posは品詞シンボル。word、reg-expはそれぞれ英語形態素の表記形と標準形。attribution-infoはデフォルト属性名と、個別属性のリスト。condition-listは訳し分けの際の条件リスト（省略可能）。

例：

デフォルト属性の定義

```
(define-info be-v-info ()
  (cond . pres) (per . 3rd) (v-pres-p . "being") (v-past-part . "been"))
.....
```

品詞、表記形、標準形、個別属性の設定

```
(define-dic english-attribution-dic t
  ("am" (be-v "be" "be" (be-v-info (v-past-t . "was"))))
  .....)
```

;;; 訳し分け情報の記述例

```
("any-some"
  (adj "any" "any" (adj-info (det . none)) ((cond . not)))
  (adj "any" "any" (adj-info (det . none)) ((modal . YN-Q)))
  (adj "some" "some" (adj-info (det . none))))
.....
)
```

・属性情報の記述について

属性情報は、(属性名・属性値)のドット・ペアのリストで表される。処理の過程で新たな属性が追加される場合は、必ず先頭に加えられる。また、属性を参照するときも先頭から探索されるので、古い属性情報はシャドウされる。属性には以下のものがある。

det	デフォルトの冠詞。ind-art (不定冠詞)、“the”、none (無冠詞)。
ind-art	とるべき不定冠詞の形。“a”、“an”。
num	単複。sing (単数)。pl (複数)。
pl	複数形の形。任意の文字列。
poss	所有格の形。任意の文字列。
comp	比較級の形。任意の文字列。
super	最上級の形。任意の文字列。
v-past-t	過去形。任意の文字列。
v-past-part	過去分詞形。任意の文字列。
v-pres-p	現在分詞形。任意の文字列。
3rd-sing-pres	三単現の形。任意の文字列。

・訳し分けの手順

- 1) 先頭の候補から順に調べて、条件に適合したものがあればそれが選ばれる。
- 2) 複数の条件が記述されている場合はそのすべてを満たしたときに選ばれる。
- 3) 条件部が省略されたものは無条件に適合する。(従って、もし途中でこれがあればそれ以降のものは無視されてしまう。)
- 4) 条件に適合するものがなければ、先頭のものが選択される。

・訳し分け条件の記述について

訳し分け条件は、(条件名・条件値)のドット・ペアのリストで表される。リストに含まれるすべての条件が満たされたときのみ、適合する。条件には以下のものがある。

cond	現在処理中の文の肯定、否定の別。not (否定)。
modal	現在処理中の文の種類。normal, YN-Q, WH-Q, COM-SEN, ..., etc.
pre-cond	前文の肯定、否定の別。not (否定)。
pre-modal	前文の種類。normal, YN-Q, WH-Q, COM-SEN, ..., etc.

2-3. 不規則変化動詞辞書

生成処理で動詞を活用させるとき用いられる。

書式：

`((english-string conjugated-string)pos sem-code*)`

english-string は英語文字列。*conjugated-string* は変化形。*pos* は品詞。*sem-code** はいくつかの意味コード。なお、ここで使う品詞は次のとおり。

"v past t"	過去形	"v pres p"	現在分詞
"v past part"	過去分詞		

例：

`((("transfer" "transferred") "v past part" "J215" "M52"))`

V. 操作説明

1. インストールおよびシステム起動方法

1-1. インストール

本システムは Symbolics Lisp Machine 上の Genera 8.0.1 環境で開発されたが、翻訳処理部は Common Lisp の仕様で作成されているので、容易に移植可能である（UNIX 上の版は後述）。また、本システムでは形態素解析のためのモジュール MORPH を使用するので、あらかじめこれがインストールされていなければならない。

ここでは、Symbolics Lisp Machine 上の Genera 8.0.1 環境で形態素解析システム MORPH がロードされているものとして説明する。テープはディストリビューション・フォーマットで作成されており、ディスクにインストールするには次のようにする。

(1) SYS:SITE;TDMT.TRANSLATIONS.NEWEST の作成とロード

プログラムおよびデータ・ファイルの物理パスと論理パスを結び付ける変換ファイル SYS:SITE;TDMT-PRODUCT.TRANSLATIONS.NEWEST を作成し（すでに存在すればこの必要はない。）、それをロードする。

その内容は例えば次のようにすればよい。

```
;;; -*- Mode: LISP; Syntax: Common-lisp; Package: USER -*-
```

```
(fs:set-logical-pathname-host "tdmt"
```

```
  :physical-host "lm02"
```

; 物理ホスト名

```
  :translations
```

```
  '(("top;" ">tdmt>")
```

```
    ("**" ">tdmt>**>")))
```

(2) Command: Restore Distribution

テープをセットし Lisp listener から Restore Distribution コマンドをタイプする。インストールに必要なディレクトリが存在しない場合は新たに作成するかどうか確認してくるので Yes として作成する。ファイル構成については I. システム機能仕様 5. ファイル構成の項を参照。

上記の操作により、テープからディスク上に次のシステムがインストールされる。

TDMT

プログラムを管理するシステム

TDMT-DATA

辞書、データを管理するシステム

1-2. システム起動

システムの立ち上げ手順を以下に示す。インストールした直後に初めて起動する場合は、バイナリファイルを作成するためにコンパイルしなければならない。次の操作により、システムはコンパイルおよびロードされる。

Command: Compile System "tdmt" ; システムのコンパイル

2回目以降は（バイナリファイルがすでにあれば）、次のようにシステムをロードするだけでよい。

Command: Load System "tdmt" ; システムのロード

さらに、次の関数を実行すればデフォルトの辞書、および変換、解析知識がすべてロードされる。

Command: (tdmt::load-dictionaries t t) ; 辞書のロード

TDMT のウィンドウを呼び出すには次のようにする。

Command: <select>+0 ; ウィンドウの選択

これによりウィンドウが呼び出され、コマンド待ち状態になる。

なお、ウィンドウ・メニューの Load Dictionaries を用いて辞書をロードしてもよい。

上記のロード関数により必要な全ての知識ルールおよび辞書がロードされるが、個別にロードしたい場合は後で述べるシステムコマンド Load Dictionaries を用いるか、以下のように関数を呼び出せばよい。

解析知識ルール	Command: (load-all-analysis-data)
変換知識ルール	Command: (load-all-pattern-data)
日英変換辞書	Command: (make-japanese-to-english-dic t)
英語属性辞書	Command: (make-english-attribution-dic t)
不規則変化動詞辞書	Command: (make-conjugation-dic t)

2. ウィンドウおよびコマンド説明

ウィンドウは9個のペインに分割されており、それぞれ次の内容を表示している。

1) Title	TDMTの文字
2) Menu	コマンド・メニュー
3) TDMT Listener	TDMT専用のリスナー
4) Input	入力された日本語文
5) Outout	選択された英語文
6) Process	処理の途中経過
7) Distance calculation in Transfer	変換の際の距離計算結果
8) Distance calculation in Analysis	解析の際の距離計算結果
9) Other Candidates Output	他の英語文候補

操作はコマンドをMenu からマウスで選択するか、TDMT Listener にキー入力で直接打ち込むことによって行なう。キー入力でコマンドを打ち込む場合は、"："の後につづけなければならない。コマンドの詳細は1. システム機能仕様、4-2. ウィンドウ詳細の項を参照のこと。

以下に実際のウィンドウ表示例を示す。

図2 ウィンドウ表示例

3. 知識ルールについて

3-1. 知識ルールの概要

TDMT で参照する知識には大きく分けて変換知識と解析知識があるが、特に変換知識はストリングレベル、パターンレベル、文法レベルというように様々な種類のものがあり、その働きばかりでなく形式上もちがう形態をしている。ここでは、データの記述されている形式上の分類により説明する。以下、ストリングレベルの知識は日英変換辞書、パターンレベルの知識は変換知識ルール（単に変換知識と呼ぶ場合と区別する。）、文法レベルの知識は POS データ（解析知識に分類）として記述されているものとする。

知識ルールには解析部用のものと変換部用のものがある。

変換部用の知識ルールのファイルは p-data> の下にあり、*.lisp の形式の名前がついている。

解析部用の知識ルールのファイルは a-data> の下にあり、local-transformation.lisp, total-transformation.lisp, lexical-transformation.lisp, pos.lisp がある。（pos 処理は変換部で行なわれるが、データの形式上は解析知識に分類する。）

これらにはそれぞれの処理の際に使われる変換、変形のためのルールが記述されているが、初めに1度すべてロードして内部データに置き換えるようになっている。

現在のシステムにロードされるルールの数は以下のとおりである。

P-DATA 1266

モデル会話（約 500 語彙）と新評価文（約 1500 語彙）を翻訳した場合、実際に採用されるルールはこのうち半分以下である。（データ未整理のため）

	種別	グループ数	ルール数
A-DATA	lexical	240	321
	local	76	199
	total	33	131
	pos	18	46

3-2. 知識ルールの変更

知識ルールはエディターで変更することができ、その変更したルールだけをロードすればよい。ただし、解析部用のルールを新たに追加した場合はファイル全体をロードする必要がある。

それぞれの全ファイルをロードするには以下のように関数を実行すればよい。

解析知識ルール	Command: (load-all-analysis-data)
変換知識ルール	Command: (load-all-pattern-data)

4. 辞書について

4-1. 辞書の概要

辞書には日英変換辞書、英語属性辞書、不規則変化動詞辞書があり、これらを参照しながら処理が進んでいく。

- ・ 日英変換辞書

日本語形態素を英語形態素に変換する際に用いられる。

- ・ 英語属性辞書

英語形態素の属性を得るために用いられる。様相による訳し分けにもこの情報が使われる。

- ・ 不規則変化動詞辞書

生成処理で動詞を活用させるとき用いられる。

4-2. 辞書の変更

辞書はエディターで変更することができる。日英変換辞書、英語属性辞書はエディター内でロードすることができるが、不規則変化動詞辞書は必ずセーブした後にロードしなければ有効とはならない。

それぞれの辞書をロードするには以下のように関数を実行すればよい。

日英変換辞書	Command: (make-japanese-to-english-dic t)
英語属性辞書	Command: (make-english-attribution-dic t)
不規則変化動詞辞書	Command: (make-conjugation-dic t)

5. UNIX 版について

ここでは UNIX の Lisp 上で TDMT 実験システムを使用する場合について説明する。

5-1. システム起動

翻訳処理部は Common Lisp の仕様にもとづいているので、TDMT:CODE;*.LISP のすべてのソース・ファイルが UNIX 上でそのまま使用できる。また、形態素解析システム MORPH についてもそのまま使用できる。ただし、データ等のキャラクタ・コードは変換してやる必要がある。

現在（平成 5 年 3 月現在）Sun Sparc の as21 上に次のようなセットアップ・ファイルを用意して動作可能である。

MORPH 用のセットアップ・ファイル（/home/kuma/morph/setup.lisp）

```
;;;
```

```
;;; Setup file for MORPH
```

```
;;;
```

```
(load "/home/kuma/morph/load-morph.lisp")
```

```
(load-morph)
```

```
(in-package :morph)
```

```
(load-all-data)
```

TDMT 用のセットアップ・ファイル（/home/hitoshi/tdmt/setup.lisp）

```
;;;
```

```
;;; Setup file for TDMT
```

```
;;;
```

```
(load "/home/kuma/morph/setup.lisp")
```

```
(load "/home/hitoshi/tdmt/code/tools.lisp")
```

```
(load-tdmt)
```

```
(in-package :tdmt)
```

```
(load-dictionaries t)
```

5-2. 使用方法

以下 Sun Sparc の as21 でのシステム使用法について述べる。

- (1) TDMT のトップ・ディレクトリで Lisp 環境を起動する。

```
as21@user> cd /home/hitoshi/tdmt
```

```
as21@user> lisp
```

- (2) TDMT 用のセットアップ・ファイルをロードする。

```
> (pwd)
```

```
#P "/home/hitoshi/tdmt/"
```

```
> (load "setup.lisp")
```

- (3) パッケージをセットする。

```
> (in-package :tdmt)
```

以上の操作で Lisp 環境に TDMT システムがロードされ、システムの用意する関数が使用可能となる。関数の詳細についてはプログラム説明の章を参照のこと。

ここで例えば、次のようにして翻訳処理が実行される。

実行例：

```
> (translate "こちらは会議事務局です")
```

```
((("this is the conference office" . 0.7738213095238096))
```

付録 A. MORPH

日本語形態素解析システム

1. システム機能説明

1. システムの概要

本システムは文字列として与えられた日本語文を形態素解析するためのものである。

一般に形態素解析処理では、形態素の区切り位置と形態素の選択に対して複数の可能性が生じ、最終的にはそれらの中から最適解を選ぶ必要がある。また、それらの複数解のすべて（全解）を求めようとすれば、解の数が爆発的に増え処理速度が著しく低下する。しかもその過程では処理の途中で先へ進めず結果的には役に立たない「解になりえない解候補」も多数現われる。そこで、これらの問題を解消するために処理の途中で何らかの制限を加えて解候補の枝刈りをしなければならない。

本形態素解析システムでは統計的接続頻度（バイグラム）を用いてこれを行なう。これはある形態素と次に来るべき形態素の接続する頻度で、詳しくは単語の接続頻度を表した「単語バイグラム」と、品詞の接続頻度を表した「品詞バイグラム」の二種類を用いている。これらの接続頻度から形態素接続尤度を求め、その値を評価して処理途中の解候補の枝刈りと最終的に最適解の選択を行なっている。

また、この種の解を求める方法として、一般には縦型探索法と横型探索法があるが、本システムでは横型探索法を用いている。これはモデル会話用の形態素辞書を使用した実験の結果、縦型探索では「解になりえない解候補」によるバックトラックの負担が大きく、横型探索での複数解を同時に扱うことの負担をうわまわったことと、横型探索の方が処理速度、処理精度の調整がしやすいことによる。本システムではビームサーチといわれる横型探索法を用いており、処理途中の解候補数に一定の幅（以下ビーム幅という）を持たせて、それを越えようとする場合に解候補の枝刈りを行なう（ビーム幅による枝刈り）。また、候補数がビーム幅の範囲でも部分接続尤度が一定のしきい値（本システムではこれは0である。従って、バイグラム・データが存在するかどうかがこの基準となる。）以下になるとその解候補は枝刈りされる（しきい値による枝刈り）。そして最後に得られた複数解の中から最適解を選択する。従って、ビーム幅が狭すぎると真の最適解が途中で枝刈りされることがあり処理精度が低下し、反対に広すぎると処理速度が低下するというトレード・オフの関係がある。

2. システムの動作原理

(1) 形態素解析

本システムは日本語文字列を入力として受け取り、それを適切な形態素に区切り、正規形、品詞、活用、活用形、意味コードといった形態素情報を付加することを目的としている。ここでは、これら一連の操作を形態素解析という。

以下に本システムでの形態素解析の実際の手順を示す。

まず、入力された n 文字で構成された日本語文字列を S とし、 S の第 k 番目の文字を s_k とすると、 S 中の第 i 番目の文字から第 j 番目の文字までの部分文字列 $S_{i,j}$ を次のように定義できる。

$$S_{i,j} = s_i s_{i+1} s_{i+2} \dots s_{j-2} s_{j-1} s_j \quad \dots \quad (1.1)$$

すると S は次のように表せる。

$$S = S_{1,n} \quad \dots \quad (1.2)$$

また、処理を進めるにあたって解候補を成長させるための w 個のバッファ B_u ($u=1, 2, \dots, w$)、それらに対する第 k 文字目からの辞書引きキャッシュ C_k ($k=1, 2, \dots, n$) を定義する。これらは第 k 文字目の形態素 m_k 、または形態素候補 m_k' 、 m_k'' 、 $m_k^{(u)}$ ($m_k^{(u)}$ は第 u 番目の候補。()内は'の数を表す。ただし、これらは互いに対等であり番号は区別のためのものである。) 等を用いて次のように表される。尚、この w が前述のビーム幅である。

$$B_u = m_1 \dots m_k \dots \quad \dots \quad (1.3)$$

$$C_k = m_k' m_k'' m_k^{(1)} \dots m_k^{(w)} \dots \quad \dots \quad (1.4)$$

初め B_u 、 C_k 共に空の状態でただ S のみが存在しており、 $k=1$ として、この状態から形態素解析が始まる。形態素解析の処理は「辞書引きおよび形態素連接尤度計算」と「バッファ間評価」という2段階手続きサイクルを繰り返し、その途中で成長する解候補がすべて無くなってしまう（形態素解析不能終了）か、最終的に残ったすべての解候補が終端する（完全終了）か、あるいはまた終端した解候補が規定の解の数に達する（制限終了）と処理を完了する。

次にその手続きサイクルを示す。

手続き 1: 辞書引きおよび形態素連接尤度計算

このフェーズは、すべてのバッファに対してそれぞれ独立に行なわれる。その内容は以下のようにさらに細かく 2 段階に分かれる。

a.) 辞書引き

まず、第 k 文字目からの辞書引きキャッシュ C_k を参照する。もし、それが空なら形態素辞書を引き、 $S_{k,k+1} \sim S_{k,n}$ の可能な形態素すべてを求め、辞書引きキャッシュ C_k に納める。その結果も C_k が空であるなら、その解候補をそれ以上成長させることができないので（これが「解になり得ない解候補」である。）、カレント・バッファ（現在処理中のバッファ）を削除し次のバッファの処理に移る。

b.) 形態素連接尤度計算

辞書引きキャッシュ C_k から形態素候補を取りだし、カレント・バッファに対して部分形態素連接尤度（後述）を計算し、それがしきい値を越えていれば実際に接続させる。ただし、本システムではこの値は 0 になっている。すなわち、バイグラム・データが存在するかどうかがこの基準となる。尚、一般に C_k は複数の形態素候補を持っているのでそれらすべてに互いに独立にこの操作を行なう必要がある。従ってこの結果、もともと単一だったカレント・バッファは複数のバッファへ分裂することがある。また、どの形態素候補に対しても部分形態素連接尤度がしきい値以下であれば、そのカレント・バッファはそれ以上成長できないので削除される。つまり、しきい値による枝刈りが行なわれる。

また、 k は新たに接続された形態素の文字数分だけ増加させられる。

手続き 2: バッファ間評価

手続き 1 によって増えたバッファがバッファ数の規定値（ビーム幅）を越した場合、それぞれのバッファの持つ部分形態素連接尤度を比較して、尤度の高いものを残しそれ以外を削除する。すなわち、ここでビーム幅による枝刈りが行なわれる。

次に実際これらのサイクルを通して S がどのように解析されるかを示す。

まず、 $k=1$ として手続き 1 により $S_{1,n}$ による辞書引きが行なわれる。その結果 C_1 は、

$$C_1 = m_1' m_1'' m_1''' \dots m_1^{(u)} \dots \quad (1.5)$$

そして、これらとバッファ B との部分形態素連接尤度を求めなければならないが、初めなのでこれは空である。そこで次項で述べるとおり、文頭を表すダミーの形態素を想

定し、これに対して形態素連接尤度を求める。

尤度がすべてしきい値を越えたとすると、バッファ B は次のようになる。

$$\begin{aligned} B_1 &= m_1' \\ B_2 &= m_1'' \\ B_3 &= m_1''' \\ &\dots \\ B_u &= m_1^{(u)} \\ &\dots \end{aligned} \quad (1.6)$$

ここで、バッファ B の数が規定値を越えていれば、次に手続き 2 によりこれらのバッファの部分形態素連接尤度が比較され、生き残ったバッファが 2 回目のサイクルに入る。2 回目以後のサイクルでは初めに選ばれた形態素候補の文字数の違いによって、バッファごとに手続き 1 の k の値が異なってくる。従って、同じサイクルを経過したのちでもバッファによって（文字列上の）成長の度合い、つまり解析の進み具合は異なる。ただし、形態素数は同じである。

第 4 サイクルの後に (1. 6) は例えば次のようになっている。

$$\begin{aligned} B_{1,1,1} &= m_1' m_4'' m_5' m_{10}' & (12) \\ B_{1,1,2} &= m_1' m_4'' m_5' m_{10}'' & (15) \\ B_{1,3} &= m_1' m_4'' m_5''' m_7' & (9) \\ B_5 &= m_1^{(5)} m_4'' m_5' m_{12}' & (13) \end{aligned} \quad (1.7)$$

バッファ B にふられた番号はジェネレーションを表し、例えば $B_{1,1,2}$ は B_1 から分裂した $B_{1,1}$ から更に分裂したものであることを示す。バッファの数は処理の過程で増加したり減少したりする。この例では 4 個に減少している。また、後ろにある () 内の数字は文頭から形態素解析の進んだ部分までの文字数を表している。

ところで C_k は一見、辞書引きの際に一時的に使われるだけで意味がなさそうに見えるが、これは辞書引きの効率に貢献している。というのは、前述のとおりバッファ B はそれぞれ文字列成長の度合いが違い、S 上のある点 k から辞書引きするタイミングには一般にずれが生じる。しかも、その辞書引きは $S_{k,k+1} \sim S_{k,n}$ までの部分文字列すべてを検索する必要があるため時間がかかる。そこで、一度辞書引きして得られた形態素候補を C_k に保存しておき、2 回目からはそれを使うようにして省力化を図っている。例えば、(1. 7) において $B_{1,3}$ は次のサイクルで $k=10$ として辞書引きの必要があるが、すでに C_{10} (m_{10}' や m_{10}'' 等が入っている。) が存在しているので、これを使えばよい。

またこのほか、文頭については品詞、文末については活用形に制限を加え、辞書引きの際にこれに反するものは除かれる。

(2) 形態素連接尤度

本システムでは統計的連接頻度をもちいて形態素連接尤度を評価することにより、解の絞り込みを行なっている。

まず、形態素連接尤度を次のように定める。

$$\begin{array}{l} W(n) = \frac{C(w_{n-1}, w_n) \cdots \cdots \text{単語バイグラム (連接頻度)}}{\text{単語連接頻度} \quad C(w_{n-1}) \cdots \cdots \text{単語モノグラム (出現頻度)}} \cdots \cdots (1.8) \end{array}$$

$$\begin{array}{l} P(n) = \frac{C(p_{n-1}, p_n) \cdots \cdots \text{品詞バイグラム (連接頻度)}}{\text{品詞連接頻度} \quad C(p_{n-1}) \cdots \cdots \text{品詞モノグラム (出現頻度)}} \cdots \cdots (1.9) \end{array}$$

$$\begin{array}{l} M(n) = \frac{32 \cdot W(n) + P(n)}{33} \cdots \cdots (1.10) \\ \text{形態素連接尤度} \end{array} \quad (n = 1, 2, 3, \dots)$$

ここで w_n は n 番目の単語を表す。ただし w_0 は文頭を表すダミー単語。また、 p_n は n 番目の品詞を表す。ただし p_0 は文頭を表すダミー品詞。

こうして得られる形態素連接尤度を先頭の形態素から順々に掛け合わせていくと、各解候補の先頭から n 番目の形態素までの部分形態素連接尤度が求められる。

$$\begin{array}{l} S(n) = \prod_{k=1}^n M(k) \cdots \cdots (1.11) \\ \text{部分形態素連接尤度} \end{array}$$

部分形態素連接尤度で n をその解候補全体 (文頭から文末まで) の形態素数 ($n = n_{\text{total}}$) とすると、全形態素連接尤度が求められる。

$$\begin{array}{l} S_{\text{total}} = S(n_{\text{total}}) \cdots \cdots (1.12) \\ \text{全形態素連接尤度} \end{array}$$

処理の途中で部分形態素連接尤度が定められたしきい値を越えるとき、また、解候補数が定められたビーム幅を越えようとするとき、部分形態素連接尤度 $S(n)$ を用いて枝刈りを行なう。また、最後に複数解が得られた場合は全形態素連接尤度 S_{total} を用いて最適解の選択を行なう。

MORPH

3. システムの構成

本システムは関数ファイル群、Nグラムファイル群、辞書ファイル群からなる。

以下論理ホスト名を "MORPH" とし、ディレクトリ、ファイルのパス名は論理パス名を使用する。

3-1. 関数ファイル群

ディレクトリ： MORPH:CODE;

ファイル： PACKAGE.LISP
VARIABLES.LISP
DATA-STRUCTURES.LISP
UTILITIES.LISP
BIGRAM.LISP
DATABASE.LISP
MAIN.LISP
LOADER.LISP
MAKE-NGRAM-FILES.LISP

3-2. Nグラムファイル群

ディレクトリ： MORPH:NGRAM;

ファイル： POS-ALISTS.LISP
POS-MONOGRAM.LISP
POS-BIGRAM.LISP
POS-BIGRAM-ENCODED.LISP
WORD-MONOGRAM.LISP
WORD-MONOGRAM-R.LISP
WORD-BIGRAM.LISP
WORD-BIGRAM-ENCODED.LISP
WORD-BIGRAM-R-ENCODED.LISP

3-3. 辞書ファイル群

ディレクトリ： MORPH:DIC;

ファイル： ATR-SEM-CODE.TEXT
ATR-SEM-CODE-REVISION.TEXT
MORPH-DIC.TEXT

4. システムの機能

本システムの主な機能は文字列として与えられた日本語文を形態素解析し、形態素情報を持った単語構造体の線形リストを返すことである。また、補助的な機能として意味コード辞書の参照機能がある。

本システムではデータ、辞書類はすべてオンメモリー方式をとっているので、形態素解析を実行するためには、あらかじめ、モノグラム、バイグラム、意味コード辞書、形態素辞書がすべてロードされていなければならない。

4-1. 実行関数群

実行関数はすべて Common-Lisp の仕様で書かれている。実行関数群にはデータ・ロード関数、辞書参照関数、形態素解析関数、解析結果表示関数などがある。

4-2. Nグラムのデータ

モノグラムは出現頻度、バイグラムは接続頻度を表し、それぞれ単語と品詞に関するものがある。

本データはATRのモデル会話1～10を用いて作成されている。

4-3. 辞書

意味コード辞書は、それぞれの形態素に付ける意味コードの辞書で形態素解析処理そのものには関係しない。

形態素辞書は解の中の形態素の辞書で、もし入力文中に本辞書未登録の形態素が存在すれば正しい解が得られない。

尚、意味コード辞書はATR言語コーパスから作成されている。形態素辞書はATRモデル会話1～10および新評価文から作成されている。

II. プログラム説明

1. 主要な関数

(1) データ・ロード関数

`load-all-data`

形態素解析システムに必要なすべての辞書、データをロードする。
引き数は不要。

(2) 辞書参照関数

`get-sem-code word`

意味コード辞書を参照する。`word` は日本語文字列で正規形。

`get-jword word`

形態素辞書を参照する。`word` は日本語文字列で表記形。

(3) 形態素解析関数

`morph-internal sentence`

日本語文字列で与えられた文を形態素解析して、結果の形態素構造列を返す。
`sentence` は日本語文字列。尚、結果は次のフォーマットで返される。

$((m_{11} \ m_{12} \ \dots) (m_{21} \ m_{22} \ \dots) \dots\dots\dots)$

ここで、 m_{ij} とは第 i 番目の解の第 j 番目の形態素を表す。

(4) 解析結果表示関数

`morph sentence &optional stream number-of-print`

日本語文字列で与えられた文を形態素解析して、結果を表示する。

`sentence` は日本語文字列。`stream` は結果を出力するストリーム。`number-of-print` は出力する解候補の数の上限。

2. モジュール説明

モジュール package

package.lisp

システムのパッケージ "morph:" を作る。

モジュール globals

variables.lisp

主な大域変数を定義している。以下にそのいくつかを示す。

大域変数：

hinshi-alist	品詞のコード表。
katsuyou-alist	活用のコード表。
ktype-alist	活用形のコード表。
buntou-kinshi-hinshi-list	文頭にくることのない品詞コードのリスト。
bunmatsu-kinshi-katsuyou-type-list	文末にくることのない活用形コードのリスト。
number-of-results	出力解の数。、
number-of-buffers	解析中のビーム幅。
pos-monogram	品詞モノグラムのテーブル。
pos-bigram	品詞バイグラムのテーブル。
word-monogram	単語モノグラムのテーブル。
word-bigram	単語バイグラムのテーブル。

datta-structures.lisp

構造体、マクロ等を定義している。

構造体：

jword

word-id	単語ID
pos-id	品詞ID
word	標準形
reg-exp	正規形
hinshi	品詞
katsyou	活用
katsyou-type	活用形
sem-codes	意味コード

priority	形態素接続尤度
word-length	単語文字数
total-length	全文字数
previous	前単語構造体

モジュール utilities

utilities.lisp

基本的な関数を定義している。

関数：

buntou-kinshi-p jword

*jword*の語が文頭にくることが可能かどうか判定する。

bunmatsu-kinshi-p jword

*jword*の語が文末にくることが可能かどうか判定する。

get-jword-pos-bigram f-jword b-jword

f-jword, *b-jword*の2単語間の品詞接続尤度を求める。

get-jword-word-bigram f-jword b-jword

f-jword, *b-jword*の2単語間の単語接続尤度を求める。

モジュール bigram

bigram.lisp

接続尤度の計算を行なう。

関数：

set-priority f-jword b-jwords

*f-jword*に対して*b-jwords*のそれぞれに接続尤度を計算して、値を埋め込む。

get-yuudo f-priority word-monogram word-bigram pos-monogram pos-bigram

品詞、単語の接続尤度から形態素接続尤度を計算する。

モジュール database

database.lisp

辞書の参照のための大域変数と関数が定義されている。

大域変数：

atr-sem-code-dic	意味コード辞書のテーブル。
morph-dictionary	形態素辞書のテーブル。
candidates-cache	辞書引きの際の他候補のキャッシュ。

関数：

get-sem-code word

wordの語の意味コードを返す。

get-jword word

wordの表記の形態素候補のリストを返す。

get-candidates sentence &optional position

sentenceのpositionの位置からの形態素候補を求める。

モジュール main

main.lisp

形態素解析のトップ関数等が定義されている。

関数：

morph sentence &optional stream number-of-print

sentenceで与えられた日本語文字列を形態素解析して、結果をstreamに出力する。
number-of-printは出力解の数。

morph-internal sentence

sentenceで与えられた日本語文字列を形態素解析して、結果の形態素構造列のリストを返す。

モジュール loader

loader.lisp

辞書、データのファイル名大域変数、ロード関数が定義されている。

大域変数：

morph-top-directory	システムのトップ・ディレクトリ。
morph-directory	プログラムのあるディレクトリ。
ngram-directory	Nグラム等のあるディレクトリ。
dic-directory	辞書のあるディレクトリ。
atr-sem-code-dic-file	意味コード辞書のファイル名。
atr-sem-code-revision-dic-file	追加意味コード辞書のファイル名。
morph-dictionary-file	形態素辞書のファイル名。

関数：

load-all-data

形態素解析システムに必要なすべての辞書、データをロードする。これの実行は以下に挙げるすべてのロード関数を以下の順序で呼ぶことに等しい。

load-pos-aliases

品詞、活用形のコード表をロードする。

load-monogram

単語、品詞モノグラムのデータをロードする。

load-bigram

単語、品詞バイグラムのデータをロードする。ただし、ファイルは次のモジュール make-filesによってコード化されたものである。

load-atr-sem-code-dic &key files remake

すべての意味コード辞書をロードする。

load-all-morph-dictionaries &key files

すべての形態素辞書をロードする。

モジュール `make-files`

`make-ngram-files.lisp`

コード化したNグラムファイルを作成するためのツール関数が定義されている。

関数：

`make-encoded-pos-bigram-file input output &optional initial-count`

ファイル名 `input` の品詞バイグラム・テキスト・ファイルをファイル名 `output` のコード・ファイルに変換する。

`make-encoded-word-bigram-file input output &optional initial-count`

ファイル名 `input` の単語バイグラム・テキスト・ファイルをファイル名 `output` のコード・ファイルに変換する。

III. 辞書説明

ここでは辞書、およびデータについての説明をする。以下で辞書という場合は意味コード辞書、形態素辞書を指し、データという場合は品詞、活用、活用形コード表およびモノグラム・データ、バイグラム・データを指している。

1. 品詞、活用、活用形コード表

pos-alist.lisp

品詞、活用、活用形コードのそれぞれの変換を記述したalistをセットしている。品詞コード表の例を以下に示す。

記述例：

```
(setq *hinshi-alist*
      '((-2 . "自イデ")
        (-1 . "付イデ")
        (0 . "記 号")
        (1 . "形容詞")
        (2 . "形動詞")
        . . . . .
        . . . . .
        . . . . .
        (19 . "補動詞")
        (20 . "自動詞")
        (21 . "他動詞")
        (30 . "固名詞")
        (31 . "形名詞")
        (32 . "本動詞")
        (33 . "間投詞")
        (34 . "準動詞")
        (35 . "並助詞")
        (36 . "係助詞")
        (80 . "慣用句")
        (99 . "その他")))
```

このほか、活用の場合は*katsuyou-alist*、活用形の場合は*ktype-alist*に対して同様に記述されている。

2. モノグラム・データ

(1) 単語モノグラム・データ

単語モノグラム・データはその単語のコーパス中の出現回数をカウントしたものである。ここでいう単語とは表記形、読み、標準形、品詞、活用、活用形のすべてに対して同一である形態素群を指す。

以下にファイル・フォーマットと記述例を示す。

フォーマット：

((表記形 読み 標準形 品詞 活用 活用形) 出現回数)

.....

記述例：

.....

((「わかり」 "わかり" "分かる" "本動詞" "五段" "連用") 129)

.....

尚、活用、活用形のない語の場合は、nil を入れてある。

(2) 品詞モノグラム・データ

品詞モノグラム・データはその品詞のコーパス中の出現回数をカウントしたものである。ここでいう品詞とは品詞、活用、活用形のすべてに対して同一である形態素群を指す。

以下にファイル・フォーマットと記述例を示す。

フォーマット：

((品詞 活用 活用形) 出現回数)

.....

記述例：

.....

((「補動詞」 "五段" "未然") 73)

.....

尚、活用、活用形のない品詞の場合は、nil を入れてある。

MORPH

3. バイグラム・データ

バイグラム・データはモノグラム・データに比較して大きいのでロードに時間がかかる。そこで、実際のロードにはモノグラム・データを用いてコード化したファイルを使っている。また、このためにも（尤度計算のためにも）バイグラム・データに現われる要素はモノグラム・データに必ず存在しなければならない。

(1) 単語バイグラム・データ

単語バイグラム・データはその単語対のコーパス中の出現回数をカウントしたものである。ここでいう単語とは表記形、読み、標準形、品詞、活用、活用形のすべてに対して同一である形態素群を指す。

以下にファイル・フォーマットと記述例を示す。

フォーマット：

```
((表記形 1 読み 1 標準形 1 品詞 1 活用 1 活用形 1)
(表記形 2 読み 2 標準形 2 品詞 2 活用 2 活用形 2) 出現回数)
.....
```

記述例：

```
((“一” “ ” “一” “記 号” “変則” “未然”)
(“わかり” “わかり” “分かる” “本動詞” “五段” “連用”) 129)
.....
```

尚、活用、活用形のない語の場合は、nil を入れてある。

(2) 品詞バイグラム・データ

品詞バイグラム・データはその品詞対のコーパス中の出現回数をカウントしたものである。ここでいう品詞とは品詞、活用、活用形のすべてに対して同一である形態素群を指す。

以下にファイル・フォーマットと記述例を示す。

フォーマット：

```
((品詞 1 活用 1 活用形 1) (品詞 2 活用 2 活用形 2) 出現回数)
.....
```

記述例：

```
((“記 号” “変則” “未然”) (“補動詞” “五段” “連用”) 1)
.....
```

尚、活用、活用形のない品詞の場合は、nil を入れてある。

4. 意味コード辞書

意味コード辞書は、それぞれの形態素に付ける意味コードの辞書で形態素解析処理そのものには関係しない。ただ得られた解の形態素にシソーラスの意味コードを付与するために使われる。また、必要であれば外部のアプリケーションからも同じ意味コードを参照することができる。

尚、意味コード辞書はATR言語コーパスから作成されている。

以下にフォーマットと記述例を示す。

フォーマット：

(正規形 品詞 意味コード)

.....

記述例：

.....

("振り込む" "32" "745b")

("加える" "32" ("262a" "263a" "428c" "712"))

.....

5. 形態素辞書

形態素辞書は解析結果として出力される形態素の辞書で、ここに登録されていない形態素は本システムでは出力されない。特に、活用語の場合は活用形ごとに登録されている必要がある。

尚、形態素辞書はATRモデル会話1～10および新評価文から作成されている。

以下にフォーマットと記述例を示す。

フォーマット1:

(表記形 標準形 (品詞 活用形 活用))

.....

*活用形、活用は存在しない場合は省略可能。

記述例1:

.....

("会議" "会議" (4))

("許さ" "許す" (32 0 1))

.....

形態素辞書は「形態素解析をするまでもなく常に結果が固定しているような部分」に対して次に示すように熟語登録ができる。

フォーマット2:

(表記形 標準分離形 ((品詞 活用形 活用) ...) 表記分離形)

.....

*活用形、活用は存在しない場合は省略可能。

記述例2:

.....

("でございます" "だ | ございます" ((12 1) (19 2 6)) "で | ございます")

("されるのでしたら" "する | れる | の | です | たら" ((19 0 4) (12 3) (34) (12 1) (14))

"さ | れる | の | でし | たら")

.....

IV. 操作説明

1. インストール

本システムは Symbolics 上のシステムとして開発したが、プログラムのソース・コード、およびデータは Common Lisp の仕様に基づいて書かれているので、容易に他のマシンに移植できるようになっている。

ここでは、Symbolics 上のシステムとしてのインストールについて説明する。

本システムはディストリビューション・フォーマットによりテープに保存されている。従って、以下の操作によりテープからのインストールができる。

(1) SYS:SITE;MORPH.TRANSLATIONS.NEWEST の作成とロード

プログラムおよびデータ・ファイルの物理パスと論理パスを結び付ける変換ファイル SYS:SITE;MORPH.TRANSLATIONS.NEWEST を作成し（すでに存在すればこの必要はない。）、それをロードする。

その内容は例えば次のようにすればよい。

```
;;; -*- Mode: LISP; Syntax: Common-lisp; Package: USER -*-
```

```
(fs:set-logical-pathname-host "morph"
```

```
  :physical-host "lm02"
```

; 物理ホスト名

```
  :translations
```

```
  '(("top;" ">morph>")
```

```
    ("**" ">morph>**>")))
```

(2) Command: Restore Distribution

テープをセットし Lisp listener から Restore Distribution コマンドをタイプする。インストールに必要なディレクトリが存在しない場合は新たに作成するかどうか確認してくるので Yes として作成する。ファイル構成については 1. システム機能説明 3. システムの構成の項を参照。

上記の操作により、テープからディスク上に次のシステムがインストールされる。

MORPH

プログラムを管理するシステム

MORPH-DATA

辞書、データを管理するシステム

2. プログラムのコンパイルとロード

プログラムはインストールした後に初めて起動する場合には、プログラムをコンパイルしてバイナリ・ファイルを作成する必要がある。このためには、Lisp listener から次のようにタイプすればよい。この操作によりプログラムはコンパイルおよびロードされる。

Command: Compile System "morph"

2 回目以後の起動時には（バイナリ・ファイルがすでにあれば）、次のようにタイプしてロードするだけでよい。

Command: Load System "morph"

3. 辞書およびデータのロード

辞書およびデータは次の関数を呼べばすべてロードされる。

Command: (morph::load-all-data)

尚、個別にロードしたい場合はそれぞれ次の関数を呼べばよい。ただし、相互に関連する項目があるので、関数を呼ぶ順序が以下のとおりとなるようにする。

load-pos-alists

品詞、活用、活用形コード表のロード。

load-monogram

単語、および品詞のモノグラム・データのロード。

load-bigram

単語、および品詞のバイグラム・データのロード。

load-atr-sem-code-dic

意味コード辞書のロード。

load-all-morph-dictionaries

形態素辞書のロード。

4. 実行関数の使用方法

本システムでは外部からの呼び出しを想定した関数をいくつか用意し、これを読み出すことによりデータのロード、および形態素解析の実行ができるようになっている。

本システムの属するパッケージは MORPH である。従って、外部パッケージから使用する場合には関数名に MORPH:: を冠して呼び出す必要がある。

尚、関数の種類、内容については II. プログラム説明でふれているのでここでは省略する。

5. 辞書の変更

辞書を変更、追加、削除した場合はファイルをセーブした後、再びロードしなければならない。この場合、変更した辞書だけをロードすることもできるが、ロードの際に関連する項目を参照するので、その順序に注意する必要がある。

また、モノグラム、バイグラムについてはあくまでも統計的なデータであるので部分的な変更を加えるのは望ましくない。しかし、統計的手続きによって得た新たなデータに置き換える場合にはモジュール make-files 中のコード化のツールを使うとよい。III. 辞書説明参照。

付録 B. TDMT-COLOR

カラー・デモンストレーション・システム

I. システム機能仕様

1. システム概要

本システムは TDMT（変換主導機械翻訳）の翻訳処理過程を表示し、その原理と特長を示すためのものである。

本システムは TDMT システムがロードされている環境でインクリメンタルにロードするようになっている。従って、本システムをロードする前に MORPH（形態素解析）システムと TDMT システムをロードしなければならない。また、本システムでは Symbolics COLOR システムを利用しているので、これもロードされていなければならない。

本システムではまず、各処理モジュールがどのような場合にどのような順序で起動されるかをカラー・ディスプレイ上に表示する。次にすべての処理が終わった後で、各処理モジュールの表示部分をマウスでクリックすると、ウィンドウが切り替わり、そのモジュールでどのような処理が行なわれたかが表示される。そこでは、変換、解析で用いられる用例距離計算の結果が色彩の差異によって表示される。

2. ユーザー・インターフェース

ユーザ・インターフェースはモノクロ・ウィンドウとカラー・ウィンドウとからなり、キー入力、マウスクリックによりコマンドを起動することができる。

2-1. ウィンドウ

本システムにはモノクロ・ウィンドウとカラー・ウィンドウとがある。モノクロ・ウィンドウはTDMTシステムのものとはほぼ同じで、メニュー・ウィンドウにいくつかのコマンドが追加されているだけである。カラー・ウィンドウはカラー主画面、Lexical processing（形態素処理）画面、Analysis（解析処理）画面、Transfer（変換処理）画面がある。

カラー・ウィンドウは初めに一度だけ作成し、以後モードを切り替えることによって操作可能になる。このウィンドウ内では、コマンド・メニューのボタンをマウス・クリックすることによって、様々なコマンドを実行できる。

以下にカラー・モードの構成を示す。

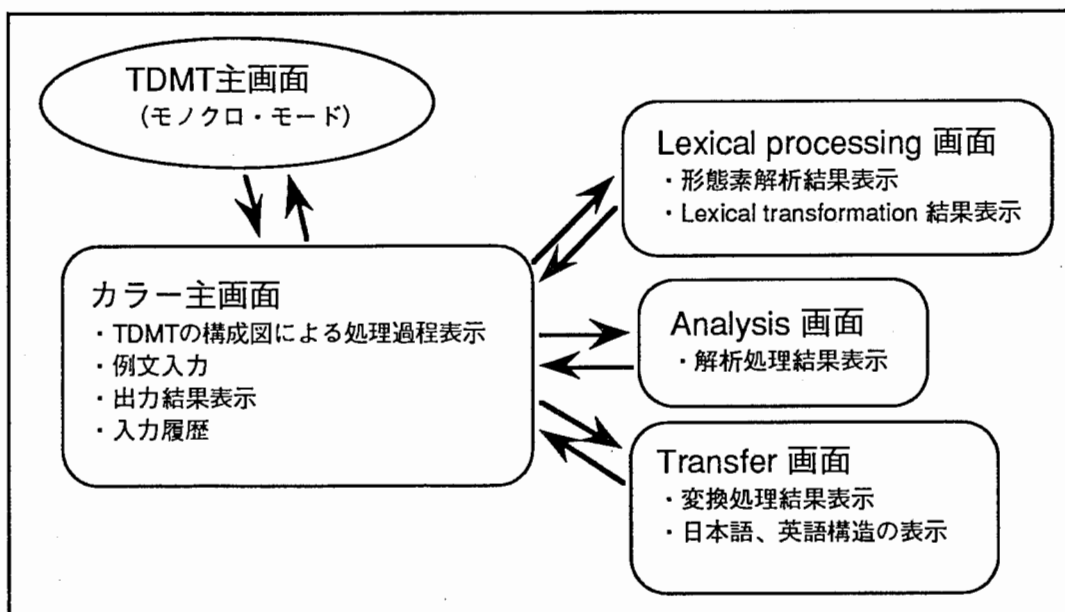


図1 カラー・ウィンドウ構成図

2-2. コマンド

本システムでは TDMT システムのコマンドはすべて使用でき、それらに加えて以下のようなカラー・ウィンドウ用のコマンドが追加されている。

Color Window

カラー・モードに切り替える。マウス・カーソルは自動的にカラー・ウィンドウに移る。もどる場合はマウスで Quit Color を選択する。

Set Up Color Window

カラー・ウィンドウの画面を作成する。初めに一度実行すればよい。

その他

各カラー・ウィンドウに付いているボタン等。(マウスクリックで起動)

3. ファイル構成

3-1. 関数ファイル群

ディレクトリ： TDMT:COLOR-WINDOW;

ファイル： GLOBALS.LISP

DRAWING-TOOLS.LISP

COLOR-FORMAT.LISP

SETUP-STUFF.LISP

DISPLAY.LISP

MAIN.LISP (パッチファイル)

WINDOW.LISP (パッチファイル)

3-2. イメージデータファイル群

ディレクトリ： TDMT:IMAGE;

ファイル： TDMT-LOGO.IMAGE

TDMT-BACKGROUND

LEXICAL-NEW

LEXICAL-GRAY

ANALYSIS-NEW

ANALYSIS-GRAY

TRANSFER-NEW

TRANSFER-GRAY

GENERATION-NEW

GENERATION-GRAY

II. プログラム説明

1. 主要な大域変数

ct-mode

カラー・ウィンドウかモノクロ・ウィンドウかのスイッチ。

ct-top-window

カラー・ウィンドウの最も外側のウィンドウ。

ct-main

カラー主画面。

ct-input-window

入力用ウィンドウ。

ct-output-window

出力用ウィンドウ。

ct-lexical-processing

形態素解析処理画面。

ct-morph-window

形態素解析結果出力用ウィンドウ。

ct-lexical-pattern-window

形態素処理パターン用ウィンドウ。

ct-analysis

解析処理画面。

ct-analysis-pattern-window

解析パターン用ウィンドウ。

ct-process-window

処理過程表示用ウィンドウ。

ct-transfer

変換処理画面。

ct-transfer-pattern-window

変換パターン用ウィンドウ。

ct-source-structure-window

原言語構造木表示ウィンドウ。

ct-target-structure-window

目的言語構造木表示ウィンドウ。

2. 主要な関数

ct-set-up-color-window

カラー・ウィンドウを作成する。

ct-display-output

結果をカラー・ウィンドウ上に表示する。

start-blink-process-box *box*

処理モジュール表示を点滅させる。 *box* は処理モジュール。

kill-blinking-process-box

処理モジュール表示の点滅をストップさせる。

III. 操作説明

1. インストールおよびシステム起動方法

1-1. インストール

TDMT システムと同様にディストリビューション・フォーマットのテープからシステムをファイルにおとす。

1-2. システム起動

本システムは TDMT システムに対しインクリメンタルにロードするようになっている。従って、MORPH, TDMT システムがロードされており、更に Symbolics COLOR システムがロードされている環境でロードしなければ起動できない。

システムをロードした後にモノクロ・ウィンドウ上で初めに一度だけ :Set Up Color Window コマンドを呼び出す。これで、カラー・ウィンドウが作成され、以後、:Color Window コマンドと :Quit Color コマンドでカラーモード、モノクロモードの切り替えができる。

2. ウィンドウおよびコマンド説明

1) カラー主画面

ウィンドウ：

- ・ Input

入力された日本文を表示。枠の部分をクリックすると、Input History コマンドが起動される。

- ・ Output

結果の英文を表示。枠の部分をクリックすると、All Candidates コマンドが起動される。また、枠内をクリックすると、Next Candidate コマンドが起動される。

ボタン：

- ・ Model Sentences

モデル会話文から入力文を選択。

- ・ Test Sentences

新評価文から入力文を選択。

- ・ Free Input

自由入力。

- ・ Quit Color

カラー・モードを抜ける。

2) Lexical Processing 画面

ウィンドウ：

- ・ Morph Analysis
形態素解析結果の表示。
- ・ Lexical Transformation
形態素合成処理の結果。

ボタン：

- ・ Quit
カラー主画面に戻る。

3) Analysis 画面

ウィンドウ：

- ・ Analysis
解析処理の結果を表示。
- ・ Process
処理の途中経過を表示。

ボタン：

- ・ Quit
カラー主画面に戻る。

4) Transfer 画面

ウィンドウ：

- ・ Transfer
変換処理の結果を表示。さらに、表示されたルールをクリックすると、そのルールに関する距離計算の詳細情報が表示される。
- ・ Source
日本語構造がグラフ表示される。
- ・ Target
英語構造がグラフ表示される。

ボタン：

- ・ All Candidates
コマンドAll Candidates が起動される。
- ・ Next Candidate
コマンドNext Candidate が起動される。
- ・ Distance Calc.
全適合ルールの距離計算の詳細を表示。
- ・ Quit
カラー主画面に戻る。

付録 C . 翻訳実行結果

日英変換主導型機械翻訳システムによる、モデル会話225文および機能試験文600文の翻訳実行結果を示す。

表示例：

("登録用紙は既にお持ちでしょうか"
("do you already have the registration form" . 1.000005))

一行目が入力文、二行目が翻訳結果、および変換における距離の総和である。この例では「登録用紙は既にお持ちでしょうか」を"do you already have the registration form"に翻訳し、距離の総和が 1.000005 であったことを示している。なお、翻訳結果は第一候補のみを表示している。これは翻訳結果候補の中で距離の総和が最小なものである。

MODEL-CONVERSATIONS

MODEL-CONVERSATION-1

{"もしもし"
{"hello" . 0)}

{"そちらは会議事務局ですか"
{"is this the conference office" . 1.5e-5)}

{"はい"
{"Yes" . 0)}

{"そうです"
{"that's right" . 0)}

{"どのようなご用件でしょうか"
{"what can I do for you" . 2.2777777)}

{"会議に申し込みたいのですが"
{"I would like to apply for the conference" . 0.0)}

{"どのような手続をすればよろしいのでしょうか"
{"how should I proceed" . 1.0)}

{"登録用紙で手続きをして下さい"
{"please proceed with the registration form" . 1.5e-5)}

{"登録用紙は既にお持ちでしょうか"
{"do you already have the registration form" . 1.000005)}

{"いいえ"
{"No" . 0)}

{"まだです"
{"not yet" . 0.0)}

{"分かりました"
{"I see" . 0)}

{"それでは登録用紙をお送り致します"
{"then we will send you the registration form" . 1.0)}

{"ご住所とお名前をお願いします"
{"please give me the address and name" . 2.0)}

{"住所は大阪市北区茶屋町二十三です"
{"the address is 23 Chaya-machi Kita-ku Osaka-shi" . 0.3333517)}

{"名前は鈴木真弓です"
{"the name is Suzuki Mayumi" . 1.0e-5)}

{"分かりました"
{"I see" . 0)}

{"登録用紙を至急送らせて頂きます"
{"we will send you the registration form as soon as possible" . 5.0e-6)}

{"よろしくお願いします"
{"thank you very much for your cooperation" . 0)}

{"それでは失礼します"
{"good-bye" . 0)}

MODEL-CONVERSATION-2

{"はい"
{"Yes" . 0)}

{"こちらは会議事務局です"
{"this is the conference office" . 0.023828816)}

{"会議の参加料について教えて頂きたいのですが"
{"I would like you to tell me about the attendance fee of the conference" . 1.5e-5)}

("今会議に申し込みれば参加料はいくらですか"
("how much is the attendance fee , if I apply for the conference now" . 0.16667168))

("はい"
("sure" . 0))

("参加料は現在お一人三万五千円です"
("the attendance fee is thirty five thousand yen per person now" . 1.00002))

("来月お申込みになりますと四万円です"
("if you apply next month , it is forty thousand yen" . 2.00002))

("参加料には予稿集代と歓迎会費が含まれています"
("the proceedings fee and reception fee are included in the attendance fee" . 0.30001202))

("わたしは情報処理学会の会員なのですが"
("I am the member of Information Processing Society" . 5.0e-6))

("参加料の割引はないのですか"
("is there not the discount on the attendance fee" . 0.33334002))

("今回は割引を行なっておりません"
("we don't discount this time" . 0.25))

("そうですか"
("I see" . 0))

("参加料はどのようにお支払いしたらよいのですか"
("how should I pay the attendance fee" . 1.0))

("参加料は銀行振り込みです"
("please pay the attendance fee by bank-transfer" . 0.0))

("案内書に記載されている口座番号に振り込んで下さい"
("please transfer into the account number that is described in the announcement" . 1.0e-5))

("また期限は今年いっぱいです"
("and the deadline is until the end of this year" . 0.33333334))

("分かりました"
("I see" . 0))

("どうもありがとうございました"
("thank you very much" . 0))

("どういたしまして"
("not at all" . 0))

("分からない点がございましたらいつでもお聞き下さい"
("if you have the question , please ask us at any time" . 1.000005))

("失礼致します"
("good-bye" . 0))

MODEL-CONVERSATION-3

("はい"
("Yes" . 0))

("こちらは会議事務局です"
("this is the conference office" . 0.023828816))

("会議に論文を発表したいと思っているのですが"
("I would like to present the paper at the conference" . 0.6666667))

("会議の内容について教えて下さい"
("please tell me about the contents of the conference" . 0.0))

("今回の会議は通訳電話に関連する広範な研究分野を含んでいます"
("this conference includes the various research fields that are related to the Interpreting Telephony"
. 0.500015))

("言語学や心理学を専攻する方にも参加して頂く予定です"
("those who specialize in linguistics and psychology would attend it" . 0.8333333))

("分かりました"
("I see" . 0))

("ところで会議での公式言語は何ですか"
("by the way, what is the official language at the conference" . 0.0))

("英語と日本語です"
("English and Japanese" . 1.0e-5))

("わたしは日本語が全然分からないのですが"
("I don't know Japanese at all" . 5.0e-6))

("発表が日本語で行なわれる場合英語への同時通訳はあるのですか"
("is there simultaneous interpretation into English , if the presentation is held in Japanese"
 . 0.8333584))

("はい"
("Yes" . 0))

("英語への同時通訳を用意しております"
("we are preparing simultaneous interpretation into English" . 0.66667))

("分かりました"
("I see" . 0))

("どうもありがとうございました"
("thank you very much" . 0))

("さようなら"
("good-bye" . 0))

MODEL-CONVERSATION-4

("こちらは会議事務局です"
("this is the conference office" . 0.023828816))

("会議について詳しいことを教えて下さい"
("please tell me the details about the conference" . 5.0e-6))

("会議の案内書をお持ちですか"
("do you have the announcement about the conference" . 1.3333333))

("いいえ"
("No" . 0))

("持っていません"
("I don't have it" . 0.0))

("そうですか"
("I see" . 0))

("会議は8月22日から25日まで京都国際会議場で開催されます"
("the conference will be held at Kyoto International Conference Hall from August 22nd to 25th"
 . 0.5555764))

("参加料は四万円です"
("the attendance fee is forty thousand yen" . 0.16668667))

("発表を希望されるのでしたら3月20日までに要約を提出して下さい"
("if you would like to present , please submit the summary by March 20th" . 0.2500025))

("会議の案内書をお送り致しますのでそれをご覧下さい"
("we will send you the announcement about the conference , so please take a look through it"
 . 2.3333333))

("失礼ですがお名前とご住所をお願い致します"
("excuse me, but could you tell me the name and address" . 2.5))

("アダムスミスです"
("I am Adam Smith" . 1.0e-5))

("住所は大阪市東区玉造2丁目27の7です"
("the address is 2 chome 2 7 7 Tamatsukuri Higashi-ku Osaka-shi" . 0.3333717))

("分かりました"

("I see" . 0))
("電話番号もお聞きしたいのですが"
("I would like to ask the telephone number" . 1.1666666))
("はい"
("Yes" . 0))
("372の8018です"
("3 7 2 8 0 1 8" . 0.1269989))
("372の8018でございますね"
("3 7 2 8 0 1 8 , right" . 0.1269989))
("はい"
("Yes" . 0))
("そうです"
("that's right" . 0))
("それではよろしくお願いします"
("thank you for your cooperation" . 0))
("失礼します"
("good-bye" . 0))
MODEL-CONVERSATION-5
("はい"
("Yes" . 0))
("こちらは会議事務局でございます"
("this is the conference office" . 0.023818815))
("ちょっとお願いがあるのですが"
("I ask the favor of you" . 1.0))
("私は会議に申込みをした者です"
("I have applied for the conference" . 5.0e-6))
("参加を取り消したいのですが"
("I would like to cancel the participation" . 0.16666667))
("お名前をお伺いできますでしょうか"
("can I ask you the name" . 2.16667))
("はい"
("Yes" . 0))
("ベル研のジムワイベルです"
("I am Jim Waibel of Bell Labs" . 1.5e-5))
("既に登録料の八万五千円を振り込まれておられますね"
("you have already transferred eighty five thousand yen for the registration fee , right"
 . 3.0e-5))
("はい"
("Yes" . 0))
("そうです"
("that's right" . 0))
("登録料を払い戻して頂けますか"
("would you refund the registration fee" . 0.6666717))
("お気の毒ですができません"
("I'm sorry, but it isn't possible" . 0.0))
("案内書にも書いていますが"
("it is written in the announcement" . 0.33333668))
("9月27日以後の取り消しに対する払い戻しはできません"
("we can't refund for the cancellation after September 27th" . 1.0889133))
("後日プログラムと予稿集をお送り致します"

("we will send you the program and proceedings later" . 1.6666666))

("では誰かが私の代わりに参加することはできますか"

("then can someone attend it instead of me" . 0.16666667))

("それは別に問題ありません"

("it doesn't matter in particular" . 0.6666667))

("代理人が参加する場合はあらかじめこちらまでお知らせ下さい"

("please inform us in advance , if substitute attend it" . 2.023814))

("分かりました"

("I see" . 0))

("代理人が決まりましたらお知らせ致します"

("when substitute is decided , we will inform you" . 1.7916759))

("では失礼します"

("good-bye" . 0))

MODEL-CONVERSATION-6

("はい"

("Yes" . 0))

("こちらは会議事務局ですが"

("this is the conference office" . 0.023823814))

("会議の間に市内観光があるそうですが"

("I have heard there is the city tour during the conference" . 0.0))

("まだ参加できますか"

("can I attend it yet" . 0.6666667))

("はい"

("Yes" . 0))

("まだ参加可能です"

("you can still attend it" . 0.5555556))

(" 8月5日の午後に清水寺金閣寺龍安寺などを見学します"

("we will see Kiyomizu-dera Kinkaku-ji and Ryoan-ji in August 5th afternoon" . 0.40001702))

("参加なさいますか"

("will you attend it" . 0.0))

("参加料はいくらですか"

("how much is the attendance fee" . 0.16667168))

("八千円です"

("it is eight thousand yen" . 2.0e-5))

("参加料には夕食代も含まれています"

("the cost for dinner is also included in the attendance fee" . 5.0e-6))

("講演者も参加されるのですか"

("will the speaker also attend it" . 5.0e-6))

("講演者の何人かは参加する予定になっています"

("some of speakers will attend it" . 0.055558894))

("そうですか"

("I see" . 0))

("それでは参加したいと思います"

("then I would like to attend it" . 0.0))

("ではお名前と人数をお願い致します"

("then could you tell me the name and number of people" . 1.5))

("ケンブラウンと申します"

("I am Ken Brown" . 5.0e-6))

("家内と参加します"

("I will attend it with my wife" . 0.0))

("集合場所は会議場の受付の前になっております"
("the meeting place is in front of the registration desk at the conference hall" . 0.071432866))

("参加料は当日集合場所でお支払い下さい"
("please pay the attendance fee at the meeting place on the day" . 1.0))

("分かりました"
("I see" . 0))

("ありがとうございました"
("thank you very much" . 0))

("ではお待ちしております"
("then we are waiting" . 1.0))

MODEL-CONVERSATION-7

("はい"
("Yes" . 0))

("こちらは会議事務局です"
("this is the conference office" . 0.023828816))

("会議で扱う話題に関して質問したいんですが"
("I would like to ask you about the topic treated at the conference" . 0.33333835))

("はい"
("Yes" . 0))

("何でしょうか"
("what is it" . 0.0))

("機械翻訳という話題が案内書に載っていますが"
("the topic about machine translation is written in the announcement" . 0.0))

("具体的にこれはどういう内容のものなんですか"
("what kind of contents is this specifically" . 0.0))

("申し訳ありませんがこちらでは専門的な質問にお答えできません"
("excuse me, but we can't answer the technical question" . 1.3666667))

("第二版の案内書に会議で発表される論文の題目が載っております"
("the title of the paper which will be presented at the conference is written in the second version of the announcement"
 . 0.6083354))

("そちらを見て頂けないでしょうか"
("wouldn't you see it" . 0.0))

("いいですよ"
("that's good" . 0.0))

("それでは早急にその案内書を送って下さい"
("then please send me the announcement immediately" . 0.9444478))

("送り先は大阪市東区城見2の1の61渡辺明です"
("the address is 2 1 6 1 Shiromi Higashi-ku Osaka-shi Akira Watanabe" . 0.57781506))

("大阪市東区城見2の1の61渡辺明様ですね"
("2 1 6 1 Shiromi Higashi-ku Osaka-shi mr. Akira Watanabe , right" . 0.27780503))

("はい"
("Yes" . 0))

("早速送らせて頂きます"
("we will send you it immediately" . 0.0))

("他に何かございますか"
("is there anything else" . 0.0))

("いいえ"
("No" . 0))

("ありません"
("nothing" . 0))

("ありがとうございました"
("thank you very much" . 0))

("失礼します"
("good-bye" . 0))

MODEL-CONVERSATION-8

("はい"
("Yes" . 0))

("会議事務局です"
("the conference office" . 1.0e-5))

("ちょっとお聞きしたいことがあるんですが"
("I have something to ask" . 1.000015))

("私は今度の会議に発表したいと思っているんですが"
("I would like to present at the next conference" . 1.0))

("どのような手続きをすればよろしいでしょうか"
("how should I proceed" . 1.0))

("先ず二百字の要約を3月20日までにこちらまでお送り下さい"
("first, please send us two hundred word summary by March 20th" . 1.1666766))

("こちらで審査を行ない5月20日までに結果をお送りします"
("after we judge , I will send you the result by May 20th" . 1.4166741))

("投稿が受理された場合原稿用紙を同封致します"
("if the submission is accepted , we will enclose the manuscript paper" . 0.0))

("6月30日までに原稿の送付をお願いします"
("please send me the manuscript by June 30th" . 0.40001202))

("分かりました"
("I see" . 0))

("要約はどのような書式で書けばいいんですか"
("what kind of form should I write the summary in" . 1.3333399))

("所定の申込み用紙がありますのでそれに記入して下さい"
("there is the prescribed application form , so please write in it" . 5.0e-6))

("それでは申込み用紙を送りますので送り先をお願いします"
("then I will send you the application form , so please give me the address" . 0.5))

("分かりました"
("I see" . 0))

("人工知能研究所のジョージオハラです"
("I am George Ohara of AI Labs" . 2.0e-5))

("住所は東京都豊島区東池袋3丁目2番5号です"
("the address is 3 chome 2 ban 5 gou Higashi Ikebukuro Toshima-ku Tokyo-to" . 0.46669236))

("人工知能研究所のジョージオハラ様ですね"
("mr. George Ohara of AI Labs , right" . 0.33333334))

("ご住所は東京都豊島区東池袋3丁目2番5号でよろしいですね"
("the address is 3 chome 2 ban 5 gou Higashi Ikebukuro Toshima-ku Tokyo-to , right" . 1.0000217
)

("はい"
("Yes" . 0))

("そうです"
("that's right" . 0))

("それでは申込み用紙の送付をよろしくお願いします"
("then please send me the application form" . 1.000005))

("はい"
("I see" . 0))

("分かりました"
("I see" . 0))

("では早速お送り致します"
("then we will send you it immediately" . 1.0))

("失礼致します"
("good-bye" . 0))

MODEL-CONVERSATION-9

("そちら会議事務局ですか"
("is this the conference office" . 0.023828812))

("はい"
("Yes" . 0))

("会議事務局です"
("the conference office" . 1.0e-5))

("何のご用件でしょうか"
("what can I do" . 1.0))

("会議場へどうやって行ったらいいか教えて欲しいんですが"
("I would like you to tell me how I should go to the conference hall" . 1.600002))

("今京都駅にいます"
("I am at Kyoto station now" . 5.0e-6))

("地下鉄で北大路駅まで行って下さい"
("please go to Kitaoji station by the subway" . 1.0e-5))

("そこから国際会議場へ行くバスが利用できます"
("you can use the bus that goes to the international conference hall from there" . 5.0e-6))

("北大路駅ではタクシーも利用できます"
("you can also use the taxi at Kitaoji station" . 5.0e-6))

("京都駅からタクシーで会議場まで行くにはいくらぐらいかかりますか"
("how much does it cost to go to the conference hall by the taxi from Kyoto station"
 . 0.7619062))

("京都駅からですとおよそ六千円かかります"
("it will cost about six thousand yen from Kyoto station" . 2.5e-5))

("では北大路駅からですといくらぐらいかかりますか"
("then how much does it cost from Kitaoji station" . 0.2666717))

("北大路駅からですとおよそ九百円です"
("it is about nine hundred yen from Kitaoji station" . 2.5e-5))

("分かりました"
("I see" . 0))

("どうもありがとうございました"
("thank you very much" . 0))

("いいえ"
("you are welcome" . 0))

("どういたしまして"
("not at all" . 0))

MODEL-CONVERSATION-10

("もしもし"
("hello" . 0))

("はい"
("Hello" . 0))

("会議事務局でございます"
("the conference office" . 0.0))

("会議の宿泊施設についてお尋ねしたいのですが"
("I would like to ask you about the hotel accommodations for the conference" . 1.4444478))

("そちらでどこか紹介して頂けますか"
("would you introduce me somewhere" . 0.16666918))

("はい"
("Yes" . 0))

("私どもでご紹介できるホテルは京都ホテルと京都プリンスホテルです"
("the hotel that we can introduce you is Kyoto Hotel and Kyoto Prince Hotel" . 1.00003))

("一人部屋の値段は一晚七千円から一万円です"
("the price of the single room is from seven thousand yen per night to ten thousand yen"
 . 0.055593893))

("二人部屋の値段は九千五百円から六万円です"
("the price of the twin room is from nine thousand five hundred yen to sixty thousand yen"
 . 4.0e-5))

("そうですか"
("I see" . 0))

("どちらのホテルが会議場に近いですか"
("which hotel is closer to the conference hall" . 2.0285761))

("京都プリンスホテルが会議場には近いんですが"
("Kyoto Prince Hotel is closer to the conference hall" . 0.7777828))

("それでは京都プリンスホテルを予約したいのですが"
("then I would like to reserve Kyoto Prince Hotel" . 0.33333835))

("ホテルの手配もして頂けるのですか"
("would you also make the arrangement for hotel rooms" . 5.0e-6))

("はい"
("Yes" . 0))

("京都ホテルと京都プリンスホテルは予約できます"
("Kyoto Hotel and Kyoto Prince Hotel will be available" . 0.6666817))

("そうですか"
("I see" . 0))

("では京都プリンスホテルの七千円の一人部屋をお願いします"
("then could you reserve seven thousand yen the single room at Kyoto Prince Hotel" . 1.000025))

("はい"
("Yes" . 0))

("京都プリンスホテルの七千円の一人部屋ですね"
("seven thousand yen the single room at Kyoto Prince Hotel , right" . 2.5e-5))

("はい"
("Yes" . 0))

("そうです"
("that's right" . 0))

("いつからお泊まりになりますか"
("when will you stay" . 1.0))

("8月4日の夜からです"
("from the evening on August 4th" . 1.0666738))

("8日の朝までお願いします"
("until 8th morning , please" . 0.23333533))

("分かりました"
("I see" . 0))

("少々お待ち下さい"
("please wait for a moment" . 1.0))

("お部屋が取れるかどうか調べます"
("I will check whether I can take the room or not" . 2.0259259))

("お部屋をお取りできます"
("we can reserve the room" . 2.0))

("ではお名前とご住所をお願いします"
("then please give me the name and address" . 3.5))

("中村一雄です"
("I am Kazuo Nakamura" . 1.0e-5))

("住所は東京都港区新橋1丁目1番3号です"
("the address is 1 chome 1 ban 3 gou Shimbashi Minato-ku Tokyo-to" . 0.5407642))

("電話番号をお願いします"
("please give me the telephone number" . 0.0))

("電話番号は331の2521です"
("the telephone number is 3 3 1 2 5 2 1" . 0.08731112))

("分かりました"
("I see" . 0))

("京都プリンスホテルに8月4日から8日まで一人部屋をお取りしました"
("I reserved the single room at Kyoto Prince Hotel from August 4th to 8th" . 2.31848))

("どうもありがとうございました"
("thank you very much" . 0))

("失礼します"
("good-bye" . 0))

NEW-TEST-SENTENCES

NEW-TEST-SENTENCE-A

("アカンパニーパーソンの欄に妻の名前を記入したはずです"
("I have written my wife's name in the category of the accompanied person" . 0.16668002))

("明日いっぱい申し込みを受け付けています"
("we are accepting an application by the end tomorrow" . 0.66667))

("明日相談しましょうか"
("shall we consult it tomorrow" . 0.7777778))

("明日そちらに伺います"
("I will visit you tomorrow" . 1.0357194))

("明日そちらに参ります"
("I will visit you tomorrow" . 0.16666918))

("明日はそちらに伺えると思います"
("I think I can visit you tomorrow" . 1.5e-5))

("宛先は大学の言語学研究室内でお願いします"
("please send it to linguistics laboratory at the university" . 0.500005))

("あとは先生のご返事を待つのみです"
("all we have to do is waiting for the professor's reply" . 1.2777778))

("あなたも会議に出席されますか"
("do you also attend the conference" . 0.75))

("あのちょっとお聞きしたいことがあるのですが"
("well , I have something to ask" . 1.0))

("アブストラクトとはつまり論文の要約のことです"
("the abstract is the summary of the paper" . 0.5555556))

("あまり突然なお話なのでこの場ではお返事できません"
("it is a very unexpected matter , so I can't reply right now" . 2.0))

("雨がふった"
("it rained" . 0.0))

("あらかじめビザを取る必要があります"
("you need to get the visa in advance" . 0.6060606))

("ありがとうございます"
("thank you very much" . 0))

("案内書にはこう書いてあるのですが"
("it is written like this in the announcement" . 0.33333668))

("案内書をご覧になりましたでしょうか"
("did you see the announcement" . 0.33333668))

("案内書を二百通送った"
("we sent you two hundred copies of announcements" . 1.0e-5))

NEW-TEST-SENTENCE-I

("いいえ協賛の条件はございません"
("no , there isn't the condition for the co-sponsor" . 0.6666667))

("いいえ使用に対する制限はありません"
("no , there isn't the limit on usage" . 0.5))

("いいお天気ですね"
("it is the good weather , isn't it" . 0.0))

("いいですよ論文の原稿は速速で送ります"
("that's good . I will send you the manuscript of the paper by express mail" . 0.33334336))

("委員会には十一人のメンバーがいる"
("there are eleven members in the committee" . 0.33333835))

("委員会には女性のメンバーがいない"
("there isn't the female member in the committee" . 0.5555556))

("委員会の日取りが決まりましたお知らせします"
("when the day of the committee is decided , I will inform you" . 1.6666667))

("いずれの場合もわたしでもお世話できます"
("in any case, we can take care you" . 1.16667))

("以前田中先生の講演を聞きました"
("I heard the Dr. Tanaka's speech before" . 0.6666717))

("急ぎますので手書きのコピーでも結構です"
("I am in a hurry , so the handwritten copy is all right" . 0.16666868))

("一時間後にもう一度コンタクトしてみます"
("I will contact one hour later again" . 1.8583359))

("一番大きな会議場は一千名まで収容できます"
("the largest conference hall can seat one thousand people" . 2.055584))

("一番速い交通機関は新幹線です"
("the fastest transportation is the shinkansen" . 0.44444445))

("いつおいでになりますか"
("when will you come" . 0.0))

("いつ登録なさいましたでしょうか"
("when did you register" . 0.6666667))

("いつまでにアレンジしないといけないのでしょうか"
("until when do I have to arrange" . 1.0))

("いつまでに準備しなくてはいけないのでしょうか"
("until when do I have to prepare" . 1.0))

("いつまでに登録しなくてはならないのでしょうか"
("until when do I have to register" . 0.0))

("伊藤さんはもう登録なさいましたか"
("did mr. Itoh already register" . 1.00001))

("今案内書を見ながら電話をかけています"
("I am calling the telephone looking the announcement now" . 0.16666667))

("今から会議に登録しておくことはできませんか"
("can't I register for the conference now" . 0.33333334))

("今から講演者として申し込みをしても大丈夫ですか"
("can I apply as the speaker now" . 1.000005))

("今北大路駅でバスを待っています"
("I am waiting for the bus at Kitaoji station now" . 0.50000405))

("今北大路駅に着きました"
("I arrived at Kitaoji station now" . 0.5555622))

("今出張から戻ったところです"
("I have just returned from the business trip" . 0.0))

("今宣伝のための記事を書いている"
("I am writing the article for the advertisement now" . 0.16668168))

("今のところ手続きを急ぐ必要はありません"
("you don't need to rush the procedure now" . 0.16666667))

("今は観光シーズンなのでホテルの予約はなかなか取れないと思います"
("it is the sightseeing season now , so I think we can't make the reservation for hotel rooms easily"
 . 1.8333333))

("いろんな分野の研究者が集まります"
("the researcher from the various fields will gather" . 0.8333384))

("インタビューの内容を全部載せることはできません"
("we can't print the all of contents of the interview" . 0.33333334))

NEW-TEST-SENTENCE-U

("うーんそうおっしゃられましても困ります"
("well , I am in trouble about that" . 0.0))

("受付で名札を渡します"
("I will give you the name card at the registration desk" . 0.5523914))

("受付では登録証を提出して下さい"
("please submit the registration card at the registration desk" . 0.44444448))

("受付と会場の間に講師の部屋があります"
("there is the room for the speaker between the registration desk and conference hall" . 1.0e-5
))

("受付に申込用紙があります"
("there is the application form at the registration desk" . 5.0e-6))

NEW-TEST-SENTENCE-E

("えーと登録費のことなんですが"
("well , it is concerning the registration fee" . 0.0))

("英語では簡単な会話さえ満足にできません"
("I don't speak even the basic conversation in English properly" . 0.0))

("ええそれで結構です"
("yes , that's good" . 0.0))

("ええ特典がいろいろあります"
("yes , there are many privileges" . 0.0))

NEW-TEST-SENTENCE-O

("オープニングセッションは朝九時に始まります"
("the opening session will begin at nine o'clock in the morning" . 0.9083521))

("お忙しい中申し訳ありません"
("I'm sorry for taking your time" . 1.0))

("OHPを使うことができますか"
("can I use the overhead projector" . 0.6666667))

("多くの講演がありますが機械翻訳に関するものは注目されています"
("there are many speeches , but ones concerning the machine translation are paid the attention"
 . 0.38889724))

("多くの講演者が質問された"
("many speakers were asked" . 1.0277778))

("多くの聴衆から質問された"
("I was asked by many audiences" . 0.6666667))

("大勢の研究者がアメリカから参加します"
("many researchers will attend it from America" . 2.0e-5))

("お気の毒ですが締め切りが過ぎています"
("I'm sorry, but the deadline has passed" . 5.0e-6))

("お支払いは銀行振込もしくは現金をお願いします"
("please pay it by the bank transfer or cash" . 1.4444544))

("遅くとも五月二十日までには予算の申請をお願いします"
("please make an application for the budget by May 20th at the latest" . 1.400007))

("おそらく参加者は二千人程度でしょう"
("probably, the participant will be about two thousand people" . 0.22224723))

("おそれいりますが高木先生をお願い致します"
("excuse me, but could I speak to Dr. Takagi" . 1.0e-5))

("お手元に登録用紙はお持ちですか"
("do you have the registration form at the hand" . 1.6666667))

("お名前のつづりを教えて頂けますか"
("would you tell me spelling of the name" . 1.5))

("お名前はローマ字でも書いて下さい"
("please write the name in Roman letters" . 1.5))

("お一人でお泊まりになる方のみ特別料金が必要になります"
("the special cost is necessary for only those who stay alone" . 1.2857193))

NEW-TEST-SENTENCE-KA

("カードが使えると非常に助かります"
("if I can use the card , it would be very helpful" . 0.0))

("海外からの参加者もかなりいます"
("there are many participants from overseas" . 0.16667502))

("会期中に観光ツアーがあります"
("there is sightseeing tour during the conference" . 0.28334433))

("会議が終わった後歓迎会があります"
("there is the welcome party after the conference is over" . 0.16668002))

("会議が催されている間雨がふっていた"
("it was raining while the conference was held" . 0.0))

("会議場にカメラを持ち込むことはできません"
("you can't bring the camera into the conference hall" . 0.0))

("会議場にスライド用のスクリーンはありますか"
("is there the slide for the screen at the conference hall" . 0.33333334))

("会議場は五百人ほどの参加者でいっぱいになった"
("the conference hall was filled by about five hundred participants" . 0.111137785))

("会議場は町の中心にあります"
("there is the conference hall in the center of the city" . 0.50001127))

("会議場までバスでいくらかかりますか"
("how much does it cost by the bus to the conference hall" . 0.26667368))

("会議中に京都を回りたいけれど時間が足りません"
("I would like to go to Kyoto during the conference , but the time isn't enough" . 0.41667166))

("会議で京都へお越しならばお寄り下さい"
("please drop in , when you come to Kyoto at the conference" . 2.4940538))

("会議でさまざまな発表を聞くのを楽しみにしています"
("I am looking forward to hearing the various presentations at the conference" . 0.22222388))

("会議で使われる言葉は日本語だけです"
("is the language which will be used at the conference only Japanese" . 1.3888988))

("会議でのOA機器は国内のメーカーが提供しております"
("the office automation equipment at the conference is provided by the domestic manufacturer"
 . 1.1666718))

("会議での講演内容を録音することは禁じられております"
("it is prohibited to record content of the speech at the conference" . 0.0))

("会議で発表する論文をまとめています"
("I am summarizeing the paper that I will present at the conference" . 0.33333334))

("会議に関する資料を送らなければなりません"
("we have to send you the materials concerning the conference" . 0.73333335))

("会議に参加する国を挙げて下さい"
("please mention the country which will attend the conference" . 0.41667002))

("会議に出席しない"
("I won't attend the conference" . 0.0))

("会議に出席する際にどのような手続きが必要ですか"
("what kind of procedure is necessary when I attend the conference" . 0.6666667))

("会議に出席できるかどうか至急お知らせ下さい"
("please inform us whether you can attend the conference or not as soon as possible"
. 1.6666667))

("会議にはいろいろな国から講演者がいらっしゃいます"
("the speaker will come to the conference from many countries" . 1.3333418))

("会議の後に論文集の出版も予定されています"
("the publication of the proceedings is also planned after the conference" . 1.0e-5))

("会議の一週間前までには論文集を作る予定です"
("we will make the proceedings by one week before the conference" . 1.2047721))

("会議の開会式には参加しないつもりです"
("I am not going to attend opening ceremony of the conference" . 0.7083334))

("会議の詳細いことが分かりましたら必ず通知を差し上げます"
("when we know the details about the conference , I will make sure to give you a notification"
. 0.5166832))

("会議の具体的な日程はまだ未定です"
("the definite schedule for the conference is not decided yet" . 0.0))

("会議のことで個人的な相談をしたいのですが"
("I would like to consult privately concerning the conference" . 0.08333583))

("会議の参加者があまりに多いので驚きました"
("I was surprised that there were too many participants in the conference" . 1.0e-5))

("会議の参加料は高いが参加する人は多いらしい"
("the attendance fee of the conference is expensive , but it seems that the people who will attend it are
many"
. 1.0e-5))

("会議のテーマはどのような分野ですか"
("what kind of field is the theme about the conference" . 1.9888889))

("会議の登録について知りたいのですが"
("I would like to know about the registration for the conference" . 0.41666666))

("会議の内容についてはプログラムを見て下さい"
("please see the program about the contents of the conference" . 0.0))

("会議の冒頭で首相のメッセージが読まれます"
("the message of the prime minister will be read at the beginning of the conference"
. 1.3055556))

("会議のポスターができました"
("we have printed the poster of the conference" . 0.0))

("会議の申し込みの方法を教えてください"
("please tell me the way of an application to the conference" . 5.0e-6))

("会議は今年開催されることに決まっている"
("the conference has been decided to be held this year" . 1.00001))

("会議は充実していてしかも観光ツアーまで用意されています"
("the conference is satisfactory , and even sightseeing tour is prepared" . 0.25))

("会議は十時から始まります"
("the conference will begin at ten o'clock" . 0.11906049))

("会議は十時に始まります"
("the conference will begin at ten o'clock" . 0.04168042))

("会議への参加を申し込みたいのですが"
("I would like to apply the participation for the conference" . 0.6666667))

("会議を開催するたびに参加者が増えています"
("each time we hold the conference , the number of participant is increasening" . 5.0e-6))

("開催の決定に至るまで大変でした"
("we have a hard time until the decision of the opening" . 0.0))

("会社あての領収書をもらえますか"
("can I get the receipt for the company" . 0.16667502))

("会場にいらした時名札を差し上げます"
("when you come to the conference hall , I will give you the name card" . 0.13334034))

("会場にはOHPの設備がある"
("there is the equipment of the overhead projector at the conference hall" . 5.0e-6))

("会場にはOHPの設備がない"
("there is not the equipment of the overhead projector at the conference hall" . 0.16667001))

("会場の近くに郵便局はありますか"
("is there post office near the conference hall" . 0.16667001))

("会場の近くに郵便局はありませんか"
("isn't there post office near the conference hall" . 0.46296632))

("会場の近くのホテルに泊まっています"
("I am staying at the hotel near conference hall" . 0.25))

("会場は広いでしょうか"
("is the conference hall large" . 0.0))

("会場は広いでしょうね"
("the conference hall is large , isn't it" . 0.0))

("会場は広いはずです"
("the conference hall must be large" . 0.0))

("会場までバスが来ています"
("the bus comes to the conference hall" . 0.16666667))

("会場までバスで行けますか"
("can I go by the bus to the conference hall" . 0.33333835))

("各分野から五名ずつ教授を選んでおります"
("we have chosen five professors from the each field" . 1.3055656))

("かしこまりました"
("certainly" . 0))

("肩書なんか別にありません"
("there isn't the title in particular" . 1.0e-5))

("カラー印刷ではさらにコストがかかります"
("it will be more expensive at the color printing" . 1.1888989))

("彼は来るようです"
("it seems that he will come" . 0.21428715))

("彼は日本を二度訪れている"
("he has visited Japan twice" . 0.8976255))

("彼は四ヶ国語を話します"
("he speaks four languages" . 0.38890925))

("観光ツアーがあることを知りませんでした"
("I didn't know there is sightseeing tour" . 1.5e-5))

("観光の件については旅行社に相談して下さい"
("please consult the travel agency about a sightseeing matter" . 0.5833408))

("簡単な履歴書を同封して頂けるとありがたいのですが"
("if you would enclose brief personal history , I appreciate it" . 1.3333334))

("関連するプロシーディングス類が展示される予定です"
("the related proceedings will be displayed" . 0.6666717))

("学生および大学関係者には登録料が割引になります"
("the registration fee is discounted for the student and person related to university"
 . 0.35715288))

("学会に出席される際お立ち寄り下さい"
("please drop in when you attend the society" . 1.3333334))

("学会に出席するために三日ほど留守にします"
("I am out for about 3rd to attend the society" . 0.9761972))

("学会には出席しません"
("I won't attend the society" . 0.16666667))

("学会には大学から代表を出します"
("the university will send representative to the society" . 0.71428573))

NEW-TEST-SENTENCE-KI

("北大路駅から国際会議場までバスが通じています"
("there is the bus from Kitaoji station to the international conference hall" . 0.48718286))

("北大路駅よりバスがあります"
("there is the bus from Kitaoji station" . 5.0e-6))

("昨日までずっとホテルは満室でした"
("hotel room had been booked until yesterday" . 1.0e-5))

("昨日読んだ本はおもしろかった"
("the book that I read yesterday was interesting" . 0.33333835))

("急なご依頼なのでこの場で即答できません"
("that is the unexpected request , so I can't give an immediate answer right now" . 1.33334))

("京都駅からの道筋がはっきりしないのですが"
("the route from Kyoto station isn't clear" . 1.2666667))

("京都駅から北大路までバスが出ています"
("there is the bus from Kyoto station to Kita-oji" . 0.52526134))

("京都駅を出るとすぐタクシー乗り場です"
("if you go out of Kyoto station , taxi stand is close" . 0.0))

("京都国際会議場の正確な場所を思い出せないのですが"
("I can't remember the exact place of Kyoto International Conference Hall" . 0.33333835))

("京都大学の長尾教授など早速出席のお返事を下さいました"
("Prof. Nagao of Kyoto university gave the reply of attendance immediately" . 2.277787))

("京都大学の長尾教授等が組織委員会のメンバーです"
("Prof. Nagao of Kyoto university and others are the members of the organizing committee" . 0.0
))

("京都でこの会議が開かれるのは初めてです"
("it is first time that this conference is held at Kyoto" . 1.05))

("京都に一度行ってみたいです"
("I would like to go to Kyoto once" . 0.6666667))

("京都のお寺を見たいのですが"
("I would like to see the temple in Kyoto" . 1.500005))

("京都の市内観光では寺をたくさん回ります"
("we will visit a lot of temples during the city tour of Kyoto" . 0.05555889))

("京都の市内観光は約四時間かかります"
("it will take about four hours for the city tour of Kyoto" . 0.05557889))

("京都の三つのホテルは既にほとんど予約されています"
("three hotels in Kyoto are already reserved almost completely" . 0.69445944))

("京都の夜のツアーには現在で百名ばかり集まっております"
("about hundred people have applied for the evening tour of Kyoto now" . 0.0))

("京都は暑いですが会場には冷房が入っています"
("it is hot in Kyoto , but there is an air-conditioner in the conference hall" . 0.33333835))

("京都へ行きたい"
("I would like to go to Kyoto" . 5.0e-6))

("京都へ一度行ってみたいです"
("I would like to go to Kyoto once" . 0.6666717))

("今日のセッションの内容には何ら新しいものがなかった"
("there was not anything new in the contents of today's session at all" . 0.3809724))

("去年の会議の参加者は約五百名でした"
("there were about five hundred people participants in the last year's conference" . 1.00003))

("銀行振込かまたは郵便為替でお支払い下さい"
("please pay by the bank transfer or mail money order" . 1.8777778))

("銀行振込か郵便為替でお支払い下さい"
("please pay by the bank transfer or mail money order" . 1.8777778))

("銀行振込でも構いません"
("the bank transfer is all right" . 0.44444445))

NEW-TEST-SENTENCE-KU

("九時五十分から基調講演が行なわれます"
("keynote speech will be held from nine fifty" . 0.25000668))

("クレジットカードもご利用になれます"
("you may also use the credit card" . 1.3333334))

NEW-TEST-SENTENCE-KE

("掲載されるかどうか分かりません"
("I don't know whether it will be printed or not" . 0.3333334))

("結果は書面で確認しましょう"
("I will check the result on the paper" . 0.0))

("原稿を京都駅でなくしたらしい"
("it seems that I have lost the manuscript at Kyoto station" . 0.5555631))

("言語のセッションを聞くかまたは機械処理のセッションを聞くか迷っています"
("I am wondering whether I should hear the language session or hear the session the machine processing"
 . 1.3333384))

NEW-TEST-SENTENCE-KO

("講演会は一時間ほどで終わります"
("lecture meeting will be over in about one hour" . 0.3333334))

("講演して頂ければ十分です"
("if you make a speech , that is enough" . 0.0))

("講演者から返事が届くまで時間がかかります"
("it will take the time to receive the reply from the speaker" . 0.2222222))

("講演の準備はちゃんとできています"
("the preparation for the speech has been completed properly" . 0.3888889))

("後援を依頼するためアメリカに担当者を派遣します"
("I will send the person in charge to America to ask support" . 1.250011))

("広報の担当者を通して情報を確認します"
("I will check information through the person in charge of the advertisement" . 0.16667502))

("公立の機関の方の登録料は割引になります"
("the registration fee for person from the public institute is discounted" . 5.0e-6))

("国際会議において技術協力について話し合われた"
("the technology cooperation was discussed at International Conference" . 0.6666667))

("ここからはその建物が見えません"
("I can't see the building from here" . 0.8333334))

("個人でホテルの予約をして下さい"
("please reserve for hotel rooms by individual" . 0.16667168))

("こちらから申し上げることは何もありません"
("there aren't we tell anything" . 2.0e-5))

("この会議には世界中から関心が寄せられています"
("this conference draws the attention from all over the world" . 0.0))

("この会議のことが新聞に出ました"
("this conference was in newspaper" . 0.0))

("この会議は科学と産業の両方にとって有意義です"
("this conference is meaningful to both the science and industry" . 0.500005))

("この会議は研究者の間で特に重視されています"
("this conference is taken seriously by the researcher in particular" . 1.0277778))

("この国際会議は毎年催されます"
("this International Conference will be held every year" . 0.083335005))

("この資料を読まれましたか"
("did you read this material" . 0.7777778))

("このセッションは五時で終わるようです"
("it seems that this session will be over at five o'clock" . 1.00003))

("このセッションは四時までです"
("this session is until four o'clock" . 0.50002))

("このセッションは少し難しい"
("this session is a little difficult" . 1.0e-5))

("細かい点についてよく検討しなければならないと思います"
("I think we have to investigate the detailed point carefully" . 1.604764))

("これから言語研究が盛んになっていくでしょう"
("the language research will become more popular from now" . 0.33333334))

("これでシンポジウムを終わります"
("we will conclude the symposium now" . 0.0))

("これは大きな会議です"
("this is the large conference" . 0.11111945))

("これは小さな会議です"
("this is the small conference" . 0.11111945))

("これはなかなかよく書けている論文です"
("this is the paper that is well written" . 0.0))

("今回の会議では多くの企業と情報処理学会がタイアップしています"
("many companies and Information Processing Society are being involved in this conference"
 . 1.00001))

("コンGRESキットはちゃんと一式用意されています"
("the set properly of the congress kit is prepared" . 1.0))

("今月中に登録料をお支払い下さい"
("please pay the registration fee within this month" . 1.1666716))

("今後の要望についての質疑を受けました"
("I received the question about the request for the future" . 0.7333454))

("今度の会議には私が代表として派遣されます"
("I will be sent to the next conference as representative" . 0.7083333))

("こんな所にホテルがあるのですか"
("is there the hotel at the such place" . 0.25000626))

("コンピューターに関する国際会議があると聞いているんですが"
("I have heard there is International Conference on the computer" . 1.0e-5))

("ご希望のホテルを具体的におっしゃっていただければ予約をお取りします"
("if you tell specifically the hotel you would like , we will get the reservation" . 2.8611212))

("午後のセッションの通訳はあまり良くありませんでした"
("there wasn't the interpretation for the session in the afternoon very good" . 0.65001))

("午後のセッションは一時ちょうどから始まります"
("the session in the afternoon will begin at just one o'clock" . 0.11112279))

("午後のセッションは一時半から始まります"
("the session in the afternoon will begin at one thirty" . 0.33334503))

("ご質問がございましたらお尋ね下さい"
("if you have the question , please ask us" . 2.0))

("ご住所とお名前をお願いします"
("please give me the address and name" . 2.0))

("ご住所は東京都ですか"
("is the address Tokyo-to" . 1.00001))

("五十件もの論文を書いている"
("I am writing as many as fifty papers" . 1.5e-5))

("午前九時までに受付に来て下さい"
("please come to the registration desk by nine o'clock a.m." . 0.5606194))

("午前十時から開会式を行います"
("we will hold opening ceremony from ten o'clock a.m." . 0.30003005))

("ご提案の主旨を拝見した上で決めたいと思います"
("after seeing the purpose of the proposal , I would like to decide" . 1.500005))

("ご発表になる内容を簡単に書いて提出して下さい"
("please briefly write the contents that you will present and submit it" . 2.0))

("五百人ほどの参加者が出席するでしょう"
("about five hundred participants will be attending" . 0.33335668))

NEW-TEST-SENTENCE-SA

("最近参加者の数が増えてきました"
("recently the number of the participants has been increaseing" . 5.0e-6))

("最終的には私の代理の者を出席させることになるでしょう"
("I will have my substitute attend after all" . 1.1333374))

("最初の部分の字が読めません"
("I can't read the letter in the first part" . 0.33334836))

("最低十人はいないとツアーは中止になります"
("unless there are at least ten people , the tour is cancelled" . 0.22223888))

("再度書類をお送り願えますか"
("can you send me document again" . 2.5833333))

("先ほどお尋ねした件ですが"
("it is concerning a matter that I have asked you earlier" . 2.3055556))

("昨年は参加者が定員を越えました"
("the participant was over capacity last year" . 1.4435358))

("さようなら"
("good-bye" . 0))

("参加資格は前回と同じです"
("the qualification for participation is the same as last time" . 0.0))

("参加者が多かったためホテルの手配ができませんでした"
("because there were many participants , we couldn't arrange for hotel rooms" . 0.33334836))

("参加者が二百名集まった"
("there were two hundred participants" . 0.7222389))

("参加者としては二百人ないし三百人を見込んでいます"
("we assume two hundred people to three hundred people as the participant" . 0.33336332))

("参加者は受付の前に集まって下さい"
("could the participant please gather in front of the registration desk" . 0.6111162))

("参加者は既に二百人を越えました"
("the participant already exceeded two hundred people" . 1.4583434))

("参加者は二千人以上だという事実をご存知ですか"
("do you know the fact that the participant is two thousand people or more" . 2.2222524))

("参加者は二百人ぐらいです"

("the participant is about two hundred people" . 0.50001))
("参加者を二つのグループに分けるつもりです"
("we are going to separate the participant into two groups" . 1.1051688))
("参加しようと思います"
("I am going to attend it" . 0.0))
("参加人員は五百人以上になると思います"
("I think the participant will be five hundred people or more" . 1.0e-5))
("参加するつもりです"
("I am going to attend it" . 0.0))
("参加人数は六十人までです"
("number of participants is no more than sixty people" . 0.33334336))
("参加費はいくらでしょうか"
("how much is the attendance fee" . 5.0e-6))
("参加費は会社が出すことになっています"
("the company will pay the attendance fee" . 0.8055606))
("参考のため参加者の人数を知りたいのですが"
("I would like to know the number of the participant for the reference" . 1.1388988))
("三種類の内どれを送れば良いですか"
("which one of three kinds should I send you" . 0.8))
("残念ですがプレスの取材を制限することに決めました"
("I'm sorry, but we have decided to restrict the coverage of press" . 1.4583359))
NEW-TEST-SENTENCE-SHI
("しかし団体の方々には若干の割引がございます"
("but there is some discount for the group" . 0.66667503))
("四月十日頃にまたお電話します"
("I will call you around April 10th again" . 2.0666687))
("仕事の目的を一応理解しました"
("I understood the purpose of the business roughly" . 0.500015))
("失礼ですがお名前をおっしゃって下さい"
("excuse me, but could you please tell the name" . 1.0))
("失礼ですがどなたでしたっけ"
("excuse me, but who is calling" . 0.0))
("指定の金額を送ります"
("I will send you the specified amount of money" . 0.16667502))
("市内観光には参加なさらないのですか"
("don't you attend the city tour" . 0.0))
("支払いは現金でもいいみたいです"
("it seems that the payment can be made in the cash" . 5.0e-6))
("締め切りが迫っているので早急に対処しなければなりません"
("the deadline is approaching , so we have to deal immediately" . 0.33333835))
("締め切りっていつですか"
("when is the deadline" . 0.0))
("写真を送って下さるのでしょうか"
("would you send me the photograph" . 0.0))
("シャンク先生が参加するかどうかわかりません"
("I don't know whether Dr. Schank will attend it or not" . 0.06667067))
("集合時間を間違えてツアーに参加できませんでした"
("I mistook the meeting time , and I couldn't attend the tour" . 0.26666668))
("集合場所は受付の後ろです"
("the meeting place is behind the registration desk" . 0.071432866))

("集合場所は受付の前です"
("the meeting place is in front of the registration desk" . 0.071432866))

("集合場所を聞き忘れておりました"
("I forgot to ask the meeting place" . 0.33333334))

("宿泊費は五万円で間に合いますか"
("is fifty thousand yen enough for the hotel expenses" . 0.6666867))

("出張から戻ったばかりです"
("I have just returned from the business trip" . 0.0))

("詳細は当日決めることにしましょう"
("shall we decide the detail on the day" . 0.6111145))

("少々お待ち下さい"
("please wait for a moment" . 1.0))

("招待者への連絡は明日でも構いません"
("it is all right to contact the guest even tomorrow" . 0.0))

("初日に首相が特別講演をなさいます"
("the prime minister will make the special speech on the first day" . 0.44447285))

("初日の夜に親睦会がございます"
("there is social meeting on the evening on the first day" . 0.700002))

("資料を直接送って欲しいのですが"
("I would like you to send me the material directly" . 0.0))

("新幹線で京都まで来るのはそれほど大変ではありません"
("it isn't so hard to come to Kyoto by the shinkansen" . 5.0e-6))

("新幹線は約十分おきに東京駅を出ています"
("the shinkansen leaves Tokyo station about every ten minutes" . 0.3030491))

("シンポジウムは午前十時より開かれます"
("the symposium will be held from ten o'clock a.m." . 1.30001))

("次回の会議は千九百九十三年に開かれる予定です"
("the next conference will be held in nineteen ninety three" . 0.80002505))

("実行委員長の下に八つの委員会が置かれています"
("eight committees are put under the direction of executive chairperson" . 1.2777828))

("実際にご自分でご覧下さい"
("please actually take a look through it yourself" . 2.16667))

("実際の入金はいつですか"
("when does the actual payment come in" . 0.0))

("実はまだ研究が終わっていないのです"
("to tell the truth, research has not been completed yet" . 0.5))

("じゃあまた明日ご連絡します"
("then I will contact you tomorrow again" . 1.8888956))

("じゃ次の回で結構です"
("then the next time is all right" . 0.9722222))

("十月になればかなり涼しくなります"
("when October comes , it will become very cool" . 0.5873033))

("住所氏名年齢職業の欄は必ず記入して下さい"
("please make sure to write the category of the address , name , age , and profession"
 . 1.3333434))

("十時以後は会場に入らないで下さい"
("please do not enter the conference hall after ten o'clock" . 1.912714))

("準備のための余裕は少しもないと思います"
("I think there is not the space for the preparation at all" . 0.0))

("人工知能研究所の渡辺ですけれどまだ登録用紙は届いていませんか"
("this is Watanabe of AI Labs . don't you receive the registration form yet" . 0.7833459))

NEW-TEST-SENTENCE-SU

("すいませんもう一度おっしゃって下さい"
("excuse me , could you please say it again" . 0.8333333))

("すぐ車を手配します"
("I will arrange the car right away" . 0.16667333))

("既に多くの論文を書いていた"
("I was already writing many papers" . 0.33333835))

("既に多くの論文を書いている"
("I am already writing many papers" . 0.33333835))

("既に原稿はできております"
("the manuscript is already completed" . 0.7500059))

("既にこちらに登録用紙をお送りでしょうね"
("do you already send me the registration form" . 1.0))

("既にサマリーを書いてあった"
("I had already written the summary" . 5.0e-6))

("既にサマリーを書いてある"
("I have already written the summary" . 5.0e-6))

("既に申し上げた通りお約束はできません"
("as we told you before , we can't promise it" . 0.33334002))

("すなわちアブストラクトです"
("that is to say, it is the abstract" . 0.0))

("スピーチの時間があまりありません"
("there isn't enough time for the speech" . 0.16666667))

("すべての論文を収録した論文集を出す予定です"
("we will send the proceedings in which we collected the all paper" . 1.0))

("すみませんこちらは事務局ですか"
("excuse me , is this the office" . 0.7500075))

("スライドかOHPなんかがあればとても助かるんですが"
("if there is the slide or overhead projector , it is very helpful" . 1.6777878))

("スライドは五かける六インチのサイズです"
("the slide is five by six inches sizes" . 0.33334336))

NEW-TEST-SENTENCE-SE

("正式な登録はまだ行なわれていませんね"
("the formal registration isn't made yet" . 0.0))

("せっかく京都へ来たので見物しようと思います"
("I came all the way to Kyoto , so I am going to see" . 0.33333835))

("先日招待講演の依頼を受けたのですが"
("I received the request of lecture the other day" . 1.0))

("先日は大変お手数をかけました"
("I caused you so much trouble the other day" . 1.2619047))

("先生が会議の詳細を知りたいとおっしゃいました"
("the professor said I would like to know for the detail about the conference" . 1.7083333))

("先生に敬意を表してパーティーを開く予定です"
("we will hold the party in honor of the professor" . 5.0e-6))

("先生には厚く御礼申し上げます"
("I thank the professor sincerely" . 0.0))

("先生には専用のリムジンをご用意致します"
("we will prepare the special limousine for the professor" . 1.4444529))

("先生の講演の順番を一番目から三番目に直したいのですが"

("I would like to change the professor's speech order to third from the first" . 1.3958384))

("先生の同僚の松田様に連絡を取りたいのですが"
("I would like to get contact with mr. Matsuda of the colleague of the professor" . 1.0416805))

("先生は何時に会場に来られますか"
("what time will the professor come to the conference hall" . 0.77778447))

("是非会議には出席したいって伝えて下さい"
("please tell that I would like to attend the conference without fail" . 0.6666667))

("是非編集のための部会に出席して下さい"
("please attend the committee on editing without fail" . 0.6666667))

("全員に会議の案内状を出しました"
("I sent information letter for the conference to all members" . 1.6666718))

NEW-TEST-SENTENCE-SO

("そういう次第で歓迎会を欠席させていただきます"
("I will be absent from the welcome party because of the reason" . 1.5e-5))

("そういう事情なのではございません"
("that isn't the case" . 0.0))

("そういうわけではございません"
("that isn't the reason" . 0.0))

("そう言ってるのではない"
("I am not saying so" . 0.33334002))

("そう言ってるんじゃないのです"
("I am not saying so" . 0.33334002))

("早急にファックスで原稿をお送り下さい"
("please send me the manuscript by the fax immediately" . 1.2777811))

("そうじゃない"
("that's not true" . 0))

("そうだとは思いません"
("I don't think so" . 0.0))

("そうですね京都駅からは少し不便ですね"
("well , it is a little inconvenient from Kyoto station" . 1.0e-5))

("そうではございません"
("that isn't true" . 0.0))

("そうではないと思います"
("I think it is not true" . 0.0))

("そこまで気を配って頂いておそれいます"
("thank you very much for considering so much" . 0.5))

("そしたらもう一度お電話致します"
("then we will call again" . 2.5555556))

("そしてファックスの番号は一三〇八ですね"
("and the fax number is one three o eight" . 0.6944586))

("そちらだけに特別な便宜を図ることはできません"
("we can't promote the special convenience for only you" . 1.2611151))

("そちらに連絡が行きましたか"
("did you receive the information" . 0.0))

("そちらの会議に参加したいけど場所はどこでしょうか"
("I would like to attend the conference . where is the place" . 0.0))

("そのあたりにホテルはありますか"
("is there the hotel around there" . 0.22222835))

("その件に関しては判断が分かれると思います"
("I think the decision is broken about the matter" . 0.6507999))

("その件についてははっきり申し上げたはずですが"
("I have told about the matter clearly" . 0.08334584))

("そのご質問には何ともお答えできません"
("I can't answer to the question somehow" . 2.0))

("その時間帯には四つのセッションが重複します"
("four sessions will overlap on that period" . 1.0833484))

("その時期に会議に出席するのは難しいです"
("it is difficult to attend the conference at that time" . 0.2619062))

("その事実は初めて知りました"
("I knew the fact for the first time" . 0.0))

("その段階で事務局に問い合わせるつもりです"
("I am going to contact the office at the stage" . 0.3888889))

("そのテーマについて話すつもりです"
("I am going to speak about the theme" . 0.0))

("その用紙を送って下さい"
("please send me the paper" . 0.0))

("その論文は信頼できますか"
("can I rely on the paper" . 0.33333334))

("それから終了時間までは中に入れません"
("and you can't enter until the time of ending" . 0.100012004))

("それこそ問題じゃありませんか"
("isn't it the problem" . 0.7963113))

("それじゃまた後で連絡します"
("then I will contact you later" . 0.33334002))

("それで説明する機会がありませんでした"
("that's why there wasn't the opportunity to explain" . 0.0))

("それでは口座番号を申し上げますか"
("then shall I tell the account number" . 0.16667168))

("それでは早速飛行機の予約をしましょう"
("then I will reserve for airplane immediately" . 0.6666717))

("それではちょっと考えてみます"
("then I will think for a moment" . 0.7777778))

("それでは登録用紙を送ります"
("then I will send you the registration form" . 0.0))

("それでも大体的な内容は了解しました"
("even so, I understood most of the contents" . 0.16668418))

("それともタクシーでおいでになりますか"
("or will you come by the taxi" . 0.16667502))

("それなら別の方法を考えます"
("in that way, I think about another way" . 1.0238153))

("それはありがたく頂きます"
("I will be happy to receive it" . 0.071432866))

("それはおもしろい企画ですね"
("it is the interesting plan , isn't it" . 0.16666667))

("それはおもしろい企画ですねえ"
("it is the interesting plan , isn't it" . 0.16666667))

("それはちょっと困ります"
("it is a little trouble" . 0.11111445))

("それはまったくの誤解です"
("it is the absolute misunderstanding" . 0.0))

("それほど大規模な会議だとは知りませんでした")

("I didn't know so the large conference" . 0.9333334))

("そんな事情は知りませんでした"

("I didn't know that circumstance" . 5.0e-6))

NEW-TEST-SENTENCE-TA

("滞在する場所と期間をそれぞれ記入して下さい"

("please write the place and period where you stay respectively" . 0.3333334))

("滞在費用が作れるかどうか心配です"

("I worry whether I can make the hotel expenses or not" . 0.375))

("タクシーよりも地下鉄を使うことをお勧めします"

("I will recommend to use the subway rather than the taxi" . 1.6666667))

("ただ会議場は少し不便な所にあるのです"

("but there is the conference hall at a little inconvenient place" . 0.5000075))

("ただ現在の状況は分かりません"

("but I don't understand the present situation" . 0.0))

("ただしそのホテルが予約できないこともあります"

("but we also have possibility that the hotel isn't available" . 1.5e-5))

("たぶん会議の最後の日には出席できないでしょう"

("probably, I can't attend on the last day of the conference" . 0.35238582))

("担当者がすぐ来ますからお待ち下さい"

("the person in charge will come soon , so please wait" . 1.7857172))

("担当者とは相談します"

("I will consult with the person in charge" . 0.41667417))

("担当者に今の状態を確認させます"

("I will have the person in charge check the current situation" . 0.3333334))

("代理に助手を行かせましょう"

("I will have the assistant go on my behalf" . 0.0))

NEW-TEST-SENTENCE-TI

("近いうちにご連絡致します"

("we will contact you soon" . 2.0))

("ちなみに参加者は二百人くらいです"

("by the way, the participant is about two hundred people" . 0.50001))

("昼食を特別にご用意することはできません"

("we can't prepare lunch specially" . 1.45))

("ちょうど三名分だけパンフレットが残っていました"

("the pamphlet was left just only for three people" . 1.0e-5))

("ちょうどホテルの前から会議場行きのバスが出ます"

("there is the bus for conference hall from the right in front of hotel" . 0.33333585))

("ちょっと手間ですけど何とかかなと思います"

("that is a little troublesome , but I think it will be settled somehow" . 0.0))

NEW-TEST-SENTENCE-TSU

("ツアーの希望者は十人しかいません"

("there is only ten people applicant for the tour" . 1.5e-5))

("ツアーの参加者が多いとバスは二台になります"

("if there are many participants at the tour , there will be two buses" . 0.33334336))

("ツアーの参加者は二百人もいます"

("there is the participant at the tour two hundred people" . 0.45240384))

("ツインビルは大阪市内のどこからでも見えます"

("I can see Twin building from anywhere in Osaka-shi" . 0.6666717))

("通常の会議ですと参加者は二百人くらいです"
("the participant is about two hundred people at the regular conference" . 0.50001))

("通訳がいますのでご安心下さい"
("there is the interpreter , so please feel easy" . 0.8333433))

("通訳料が高いと思われるかもしれません"
("you might think the interpretation fee is expensive" . 0.3888889))

("次の委員会で来年の会議の日程を決める予定です"
("we will decide the schedule for the next year's conference at the next committee" . 1.2500159
))

("次の会議がいつかご存知ですか"
("do you know when the next conference is" . 2.0555556))

("次の会議は京都ないし大阪で開催されます"
("the next conference will be held at Kyoto or Osaka" . 0.16667917))

NEW-TEST-SENTENCE-TE

("手紙を送って頂けるのでしょうか"
("would you send me the letter" . 0.16666667))

("手紙を書いておいた"
("I have written the letter" . 0.5000067))

("手紙を書いておく"
("I write the letter" . 0.5000067))

("手紙を書きます"
("I will write the letter" . 0.90000004))

("寺とか神社などを見物します"
("we will see the temple and shrine" . 2.5e-5))

("寺も博物館も見物します"
("we will also see the temple and museum" . 0.50001))

("天候によってはバスの運行が中止になることがあります"
("we have possiblity that the bus operation is cancelled by the weather" . 1.4523809))

("展示場に出席したり会議参加者と商談したりして頂いて結構です"
("it is good to present in exhibition hall and have business discussions with the participant in conferen
ce"
 . 0.08334083))

("では会議でお目にかかるのを楽しみにしています"
("then I am looking forward to seeing at the conference" . 0.33333835))

("では会議は二月十八日つまり来週の月曜日から始まるんですね"
("then the conference begins from February 18th which is next Monday" . 0.51111645))

("では京都プリンスホテルの七千円の一人部屋をお願いします"
("then could you reserve seven thousand yen the single room at Kyoto Prince Hotel" . 1.000025))

("では再度案内書を送ります"
("then I will send you again the announcement" . 0.7777811))

("では詳細なパンフレットができましたらお送り致します"
("then when the detailed pamphlet is possible , we will send you it" . 1.3333334))

("では先生は奥様とご一緒なんですね"
("then the professor will come with the wife" . 1.1388888))

("ではちょっとスポンサーに問い合わせさせてみましょう"
("then I will contact the sponsor for a moment" . 1.1666666))

("では登録方法から説明しましょう"
("then I will explain registration procedure first" . 1.0))

("では登録用紙をお送り致します"
("then we will send you the registration form" . 1.0))

("でもこちらの通訳は優秀です"

("but our interpretation is excellent" . 0.0))

("電車で来るつもりです"
("I am going to come by the train" . 5.0e-6))

("電話が通じなくて連絡ができませんでした"
("we couldn't contact you because the telephone wasn't connected" . 0.6666667))

("電話番号をもう一度お願いします"
("please give me the telephone number again" . 0.8))

("電話番号を言いますか"
("shall I say the telephone number" . 0.5000067))

NEW-TEST-SENTENCE-TO

("当初の計画を変更しました"
("I changed the original plan" . 0.8333334))

("当日はブースの前にも椅子を置きます"
("I will put the chair in front of the booth on the day" . 1.3381026))

("登録期限はもう過ぎました"
("the deadline of registration already passed" . 0.53333837))

("登録証は後日送付して下さる訳ですね"
("you would send me the registration card later , right" . 0.9889029))

("登録証を受け取ったら手続きは終わりです"
("when you receive the registration card , the procedure is over" . 1.0916677))

("登録証を紛失した場合は再発行はできません"
("if you lose the registration card , we can't reissue" . 1.4722222))

("登録証を紛失した場合再発行はできません"
("if you lose the registration card , we can't reissue" . 1.4722222))

("登録する前に出席者の人数をもう一度確かめて下さい"
("please check the number of the participant again before registering" . 1.750005))

("登録の際は論文集の代金も含めてお支払い下さい"
("please pay the amount including the cost of the proceedings at the time of the registration"
 . 1.6666716))

("登録の手続きは前回と同じ要領です"
("the procedure of the registration is the same as last time" . 0.33333835))

("登録は先月で締め切らせて頂きました"
("we have closed the registration last month" . 0.73333836))

("登録はもう済ませましたでしょうか"
("did you already complete the registration" . 0.0))

("登録用紙が欲しい"
("I want the registration form" . 0.33333334))

("登録用紙がまだ届いていないのですが"
("I don't receive the registration form yet" . 0.8666667))

("登録用紙の書き方を間違えてしまいました"
("I made a mistake about how to write the registration form" . 0.2777778))

("登録用紙は机の上にあります"
("there is the registration form on the desk" . 0.16667001))

("登録用紙をお持ちでしょうか"
("do you have the registration form" . 1.0))

("登録用紙をお持ちですか"
("do you have the registration form" . 1.0))

("登録用紙をお送りしましょうか"
("shall we send you the registration form" . 1.8666717))

("登録用紙を送ってあげる"

("I send you the registration form" . 0.0))
("登録用紙を送って頂けないでしょうか"
("wouldn't you send me the registration form" . 0.0))
("登録用紙を送って頂けるのでしょうか"
("would you send me the registration form" . 0.0))
("登録用紙を送って下さい"
("please send me the registration form" . 0.0))
("登録用紙を送ってくれませんか"
("wouldn't you send me the registration form" . 0.0))
("登録用紙を送ってくれるように頼みました"
("we have asked you to send me the registration form" . 0.5))
("登録用紙を送って欲しい"
("I would like you to send me the registration form" . 0.0))
("登録用紙を送って欲しいのですが"
("I would like you to send me the registration form" . 0.0))
("登録用紙を送ってもらいたいのですが"
("I would like you to send me the registration form" . 0.8))
("登録用紙を送ってもらえますか"
("would you send me the registration form" . 5.0e-6))
("登録用紙を送ってもらえませんか"
("wouldn't you send me the registration form" . 0.0))
("登録用紙を送りました"
("I sent you the registration form" . 0.611111))
("登録料と論文集代のほかに費用は要りません"
("the expense isn't necessary except the registration fee and the cost for proceedings"
 . 0.6889159))
("登録料は銀行口座に振り込んで下さい"
("please transfer the registration fee into bank account" . 1.0e-5))
("登録料は日本円またはアメリカドルでお支払い下さい"
("please pay the registration fee by Japanese yen or U.S. dollar" . 1.0))
("登録料をまだお支払い頂いておりません"
("you haven't paid the registration fee yet" . 1.6666718))
("登録料を払ったのですから当然出席できる訳ですね"
("because I paid the registration fee , I can attend as a matter of fact , right" . 0.16666667))
("トピックの一覧をお送りしますので短いコメントをお願いします"
("I will send you the list of the topic , so please give me the short comment" . 2.0555556))
("どうしても見たいお寺が一箇所あるのですが"
("there is one temple that I would really like to see" . 0.37778446))
("どうぞ事務局にご用件を申し付けて下さい"
("please feel free to ask anything to the office" . 1.7777778))
("どうもありがとうございます"
("thank you very much" . 0))
("どうもこの案内書はよく分かりません"
("I don't understand this announcement well" . 0.60002005))
("どこかビジネスホテルを知りませんか"
("don't you know any business hotel" . 1.0777812))
("どちらの用紙をお送りすればよいですか"
("which paper should I send you" . 2.9666717))
("どの用紙を送るのですか"
("which paper do I send you" . 1.000005))

("長尾教授なら既に出発なさいました"
("Prof. Nagao already left" . 0.7777778))

("長尾先生に日時をお伝えしました"
("I told Dr. Nagao the date and time" . 2.1523857))

("長尾先生はおられますか"
("is Dr. Nagao there" . 0.48333332))

("長尾先生はすぐ京都にいらっしゃいます"
("Dr. Nagao will come to Kyoto soon" . 1.2888997))

("長尾先生は京都大学の教授でいらっしゃいます"
("Dr. Nagao is the professor of Kyoto university" . 0.65000904))

("長尾先生は出席されると思いますか"
("do you think Dr. Nagao will attend" . 0.23333834))

("長尾先生は組織委員長として出席されます"
("Dr. Nagao will attend as executive chairperson" . 1.0666666))

("何か分からない点がございましたらお問い合わせ下さい"
("if you have some question , please contact us" . 1.000015))

("なにぶん交通が不便なので時間にご注意下さい"
("please take care about the time because the transportation is quite inconvenient" . 1.00001))

("名前は田中一郎と申します"
("the name is Ichiro Tanaka" . 5.0e-6))

("なるべく会議の前日に京都に着くようにします"
("we will try to arrive at Kyoto on the previous day of the conference if possible" . 0.0))

("なるべく早くご返事できるよう努力致します"
("we will try to reply as soon as possible" . 1.6555576))

("なるほどそうですか"
("well, I see" . 0.0))

("何とかキャンセルできないでしょうか"
("can't I cancel somehow" . 1.0))

("何とか払い戻しをして頂けないでしょうか"
("wouldn't you refund somehow" . 1.3333334))

("何度も論文を書き直している"
("I am rewriting the paper again and again" . 0.66668504))

NEW-TEST-SENTENCE-NI

("日本へ行くのは初めてです"
("it is first time to go to Japan" . 5.0e-6))

("二百名の参加者が集まった"
("two hundred participants gathered" . 1.000025))

("日本語の場合は原稿用紙三枚をお願いします"
("please write it within three sheets of the manuscript paper in the case of Japanese"
 . 0.05556889))

("二万五千円です"
("it is twenty five thousand yen" . 2.0e-5))

NEW-TEST-SENTENCE-NU

NEW-TEST-SENTENCE-NE

("値段はいくらでも構いません"
("I don't care how much the price is" . 1.0e-5))

("念のためご住所を書いて下さい"
("please write the address just in case" . 1.2777778))

NEW-TEST-SENTENCE-NO

("残りのスペースを広告に当てたいと思います"
("I would like to spend the remaining space for the advertisement" . 0.7222272))

("後ほど詳しいスケジュールをお伝えします"
("I will tell you the detailed schedule later" . 2.468257))

("後ほど主催者による正式なアナウンスメントがあるそうです"
("it is said that we will have the formal announcements by the sponsor later" . 1.5666867))

("後ほど事務局からご連絡させていただきます"
("we will contact you from the office later" . 2.0555556))

("後ほど電話を致します"
("I will call you later" . 0.33333334))

NEW-TEST-SENTENCE-HA

("はあ新聞記者の取材は自由だと思います"
("yeah , I think the coverage by press is free" . 0.41111946))

("はいまだ若干席があります"
("yes , there are still some seats" . 0.11112945))

("はいまだ若干席がございます"
("yes , there are still some seats" . 0.0))

("配布した資料の数は多くないです"
("the number of the material that I delivered isn't many" . 1.5833333))

("八万円の登録料は全然高くない"
("eighty thousand yen of the registration fee isn't expensive at all" . 0.16669))

("発表者に対する質問は一回です"
("the question is once to presenter" . 0.5))

("発表なさるテーマについて要点をまとめて頂きます"
("I would like you to summarize the point about the theme that you will present" . 0.8333334))

("発表にはスライドを使いたいのですが"
("I would like to use the slide for the presentation" . 0.3888889))

("発表の日時の変更をご了承頂きたいのです"
("I would like you to agree to change of the presentation date and time" . 1.3333384))

("早く登録したほうがいいです"
("it is good to register soon" . 0.6805556))

("早く論文を提出して頂かないと印刷に間に合いません"
("unless you submit the paper soon , it isn't in time for the printing" . 1.7916667))

("バスが来た"
("the bus came" . 0.0))

("バスはすぐ来ます"
("the bus will come soon" . 0.6190505))

("パーティーへの参加は早めにご返事下さい"
("please reply about the participation for the party as soon as possible" . 1.000005))

("パーティーや観光ツアーなどの機会は設けるつもりです"
("we are going to have the opportunity for the party and sightseeing tour" . 0.6111228))

("パネル討論は毎日決まった時間に設定されています"
("the panel discussion is set up at the same time every day" . 0.5))

("パンフレットと一緒に登録用紙を送ります"
("I will send you the registration form with the pamphlet" . 0.33333334))

("パンフレットの三ページに料金のことが書かれていますね"
("the cost is written in page three of the pamphlet , isn't it" . 1.4861205))

("パンフレットは会議当日に受付で配布致します"
("we will deliver the pamphlet at the registration desk on the day of the conference"
 . 1.1583383))

NEW-TEST-SENTENCE-HI

("ひかり二百三十一号は京都に十二時に到着します"
("Hikari superexpress No. 231 will arrive at Kyoto twelve o'clock" . 0.8389087))
("飛行機の予約が取れなくて困っているんですが"
("I am in trouble since I can't take the reservation for airplane" . 0.600004))
("広い会場が必要でしょうか"
("is the large conference hall necessary" . 1.0))
("広い会場が必要でしょうね"
("the large conference hall will be necessary , isn't it" . 1.2592592))
("広い会場が必要です"
("the large conference hall is necessary" . 0.33333334))
("ビデオをそちらに運ぶとかいうことはできますか"
("can I bring the video there" . 1.1587349))

NEW-TEST-SENTENCE-HU

("複雑な手続きは必要ではありません"
("the complicated procedure isn't necessary" . 0.61111444))
("複雑な手続きは必要ではないと思います"
("I think the complicated procedure is not necessary" . 0.61111444))
("フラッシュをたいたり録音機材を持ち込んだりということはできません"
("you can't use the flash and bring the recording equipment" . 1.83333334))
("部会には助手を代理に出します"
("I will send the assistant to the committee on my behalf" . 0.5151565))
("プリンスホテルのほうがロイヤルホテルより安いです"
("Prince Hotel is cheaper than Royal Hotel" . 0.33333334))
("プリンスホテルのほうがロイヤルホテルより会場に近いです"
("Prince Hotel is closer to the conference hall than Royal Hotel" . 0.33333334))
("プログラム委員会が詳しい日程を決めることになっています"
("the planning committee will decide the detailed schedule" . 1.7222307))
("プログラム委員会のメンバーを座長に選びたいと思います"
("I would like to choose the member of the planning committee as the chairman" . 0.6468385))
("プログラムに対するご意見を聞かせて下さい"
("please let me ask opinion about the program" . 1.56667))
("プログラムはすぐにお送り致します"
("we will send you the program soon" . 1.000005))

NEW-TEST-SENTENCE-HE

("部屋が空いているかどうかホテルに問い合わせます"
("I will contact the hotel whether the room is available or not" . 0.5))

NEW-TEST-SENTENCE-HO

("ほかの研究論文から教えられることが多い"
("I often learn from the other research paper" . 0.33333334))
("ほかの研究論文によって教えられることが多い"
("I often learn by the other research paper" . 0.33333334))
("ほかの研究論文に教えられることが多い"
("I often learn by the other research paper" . 0.33333334))
("ホテルの部屋は確保されているようです"
("it seems that the room at the hotel is reserved" . 0.6666667))
("ホテルの予約が取れなくて困っています")

("I am in trouble since I can't take the reservation for hotel rooms" . 0.600004))
("ホテルの予約が取れなくて困っているところです"
("I am in trouble since I can't take the reservation for hotel rooms" . 0.600004))
("本日登録料を受け取りました"
("I received the registration fee today" . 0.33334002))
("本当に参加されるのですか"
("will you attend really" . 0.5))
NEW-TEST-SENTENCE-MA
("まあ多少の変更は許されると思います"
("well, I think a few change is allowed" . 0.0))
("毎年参加者が増えています"
("the number of participant is increaseing every year" . 1.000005))
("また掲載の許可については学会にお問い合わせ下さい"
("and please contact the society about the printing" . 1.7500075))
("まだお送りしてないのですが"
("I haven't sent you it yet" . 1.0))
("まだ会議の四ヵ月前ですから登録は間に合います"
("because it is four months before the conference now , the registration will be in time"
 . 0.57144284))
("まだすべての原稿が集まっていません"
("we haven't received the all manuscript yet" . 1.3333334))
("まだ登録を受け付けています"
("we are still accepting the registration" . 0.16666667))
("まだ申し込みの締め切りに間に合いますか"
("is it still in time at the deadline for an application" . 1.0277803))
("全く理解できません"
("I can't understand at all" . 0.0))
("まもなく予定ははっきりすると思います"
("I think the schedule will be clear soon" . 0.0))
NEW-TEST-SENTENCE-MI
("皆様には決して迷惑はおかけしません"
("I will never cause you the trouble" . 1.0))
("見本を作るために小切手のコピーを取らせて下さい"
("please let me make the copy of the check to make sample" . 1.530303))
NEW-TEST-SENTENCE-MU
("むしろ当日伝えるほうが良いと思います"
("I think it is rather good to tell on the day" . 0.7037126))
NEW-TEST-SENTENCE-ME
NEW-TEST-SENTENCE-MO
("もう会場に入ってもいいです"
("you can enter the conference hall now" . 1.0))
("申し上げた通り全体を独占することはできません"
("as we told , you can't monopolize the whole" . 0.66667837))
("申し込み書をお送り致します"
("we will send you the application form" . 1.9166666))
("申し込みの期限はいつですか"
("when is the deadline for an application" . 1.0e-5))

("申し込みの締め切りは五月の第一週です"
("the deadline for an application is the first week of May" . 0.36666667))

("申し訳ありませんが登録用紙を送って下さい"
("excuse me, but please send me the registration form" . 0.0))

("申し訳ないのですがホテルには午後三時までチェックインできません"
("I'm sorry, but you can't check-in the hotel by three o'clock p.m." . 0.16668834))

("もう長い間バスを待っています"
("I am waiting for the bus for a long time" . 0.0))

("もし京都観光が中止になったらパーティーを開く予定です"
("if sightseeing in Kyoto is cancelled , we will hold the party" . 0.16667168))

("もしそれが無理ならば日時を変更しましょう"
("if it is impossible , I will change the date and time" . 1.6388973))

("もちろん出席します"
("of course, I will attend" . 0.0))

("最寄りの駅は京都駅あるいは奈良駅です"
("the nearest station is Kyoto station or Nara station" . 5.0e-6))

NEW-TEST-SENTENCE-YA

("やっぱり国際会議の講演の内容は違いました"
("as I expected , the contents of the speech of International Conference were different"
 . 0.16666667))

("やっぱり変更があるみたいです"
("as I expected , it seems that there is change" . 0.33333835))

("やはり京都は見ておくべきですよ"
("you should see Kyoto after all" . 0.5555606))

("やはり二万円は高いと思います"
("I think twenty thousand yen is expensive after all" . 2.0e-5))

("山田先生に共同でインタビューをしたい"
("I would like to interview Dr. Yamada by collaboration" . 0.50001))

NEW-TEST-SENTENCE-YU

("有意義な会議になるだろう"
("it will become the meaningful conference" . 0.0))

("夕方レセプションが開かれる予定です"
("the reception will be held in the evening" . 0.85555893))

("夕食を取った後に京都の町に出かけます"
("I will go to downtown Kyoto after I get dinner" . 0.750005))

("友人とグループで参加したいと思います"
("I would like to attend it in the group with the friend" . 0.22222222))

("友人の誰かがわたしを紹介したようです"
("it seems that someone of the friend introduced me" . 1.0))

("郵送で申し込んだ旨担当者にお伝え下さい"
("please tell the person in charge that I applied by the mail" . 2.2962964))

NEW-TEST-SENTENCE-YO

("用紙を送ってもらいたいのですが"
("I would like you to send me the paper" . 0.8))

("予稿集は会場で直接渡します"
("I will give you the proceedings at the conference hall directly" . 0.26667067))

("予定が変わるようでしたら前もってお知らせ下さい"
("if it seems that the schedule will change , please inform us in advance" . 3.133335))

("予約金を払わなければなりませんか"
("do I have to pay deposit" . 1.0000025))

NEW-TEST-SENTENCE-RA

("来月の会議に登録したいのですが"
("I would like to register for next month conference" . 0.16666667))

("来月の会議に登録したいのですけれど"
("I would like to register for next month conference" . 0.16666667))

NEW-TEST-SENTENCE-RI

("リストの中から必要な事項を選んで下さい"
("please choose necessary items from the list" . 0.3888889))

("立派な会場に驚きました"
("I was surprised by the excellent conference hall" . 0.0))

NEW-TEST-SENTENCE-RU

NEW-TEST-SENTENCE-RE

("レセプションに参加されるかどうかは各自の自由です"
("whether you will attend the reception or not is under individual choice" . 0.0))

NEW-TEST-SENTENCE-RO

("ロイヤルホテルとプリンスホテルでは値段が違いますか"
("is the price different between Royal Hotel and Prince Hotel" . 0.43333334))

("ロイヤルホテルと比べてプリンスホテルのほうが安いです"
("compared to Royal Hotel , Prince Hotel is cheaper" . 0.33333334))

("ロイヤルホテルと比べればプリンスホテルは安いです"
("compared to Royal Hotel , Prince Hotel is cheap" . 0.5833384))

("ロイヤルホテルに比べてプリンスホテルのほうが安いです"
("compared with Royal Hotel , Prince Hotel is cheaper" . 0.33333835))

("ロイヤルホテルに比べればプリンスホテルは安いです"
("compared with Royal Hotel , Prince Hotel is cheap" . 0.5833434))

("ロイヤルホテルよりプリンスホテルのほうが安いです"
("Prince Hotel is cheaper than Royal Hotel" . 0.33333334))

("論文の選考が済み次第お知らせ致します"
("we will inform you as soon as we complete the decision of the paper" . 1.833345))

("論文のタイトルを直したいのですが間に合いますか"
("I would like to change the title of the paper , is it in time" . 1.12501))

("論文のテーマを考える"
("I think about the theme of the paper" . 0.58333504))

("論文の中に読めない所が三箇所あるのですが"
("there are three sections I can't read in the paper" . 0.0))

("論文の要約は二百語まででお願いします"
("please write the summary of the paper within two hundred words" . 1.0e-5))

("論文を事務局に提出したのですが気に入りません"
("I have submitted the paper to the office , but I don't like it" . 0.16666667))

NEW-TEST-SENTENCE-WA

("分からないことがありましたら事務局にお尋ね下さい"
("if you have the question , please ask the office" . 1.2500125))

("私達は新幹線の予約が取れませんでした"
("we couldn't take the reservation for the shinkansen" . 1.4333425))

("私どもでお世話致しますのでご心配は要りません"
("we will take care you , so you don't need to worry" . 2.0333366))

("私の言ったことが通じなかったようです"
("it seems that what I said wasn't understood" . 0.0))

("私の代わりに助手を遣ります"
("I will send the assistant instead of me" . 1.2375748))

("私の渡航費と滞在費はそちらで負担して下さいますか"
("would you pay my transportation fee and hotel expenses" . 0.3904855))

("私の名前を彼に使わせました"
("I had him use my name" . 0.0))

("私の発表は午前のセッションの一番です"
("my presentation is the first session in the morning" . 0.8889006))

("私は今京都駅にいるんですが"
("I am at Kyoto station now" . 5.0e-6))

("私は教授に代わって登録するよう依頼されています"
("I am asked to register on the professor's behalf" . 1.0694469))

("私は個人的にロボットに興味を持っています"
("I have the interest in the robot personally" . 0.5))

("私はそう思いません"
("I don't think so" . 0.57143146))

("私はその時点では全く知りませんでした"
("I didn't know at the point at all" . 0.16666667))

("私は妻と共に京都へ行く予定です"
("I will go to Kyoto with my wife" . 0.16667168))

("私はノーマン教授を知っています"
("I know Prof. Norman" . 0.16667001))

("私はノーマン教授を知っております"
("I know Prof. Norman" . 0.16667001))