

情報フィルタリング機能をもつ発散的思考支援環境の試作

森 康真 國藤 進

北陸先端科学技術大学院大学 情報科学研究科

本論文は、検索要求生成過程におけるユーザの問題意識に対して、発想の初期段階である発散的思考の過程に注目し、人間の創造的知的活動を支援する環境の提供を検討するものである。ここでは、情報フィルタリングに対し、問題意識を明確化するためのインタフェースとして、ブレインストーミングを支援する環境を提供することを検討する。ユーザがその時点で提示できる知識の断片に対して、システムがユーザの自由連想にきっかけを与え、様々な方向に発想を拡げる過程を支援する。その結果形成される空間配置に関連するユーザの関心/観点に基づく情報をフィードバックすることにより、ユーザの知的活動を支援するシステムを提案・構築する。

An Experimental Divergent Thinking Support Environment with Information Filtering

Yasuma MORI Susumu KUNIFUJI

Japan Advanced Institute of Science and Technology

E-mail:{yasuma,kuni}@jaist.ac.jp

15,Asahidai,Tatsunokuchi,Nomigun,Ishikawa 923-12,JAPAN

About an awareness of the issues when the user makes an information requirement, we take notice of idea generation process at the beginning of creative thinking process and propose a computer to assist human creative thinking activity. The information filtering system has a human-interface with brainstorming support environment. The user gives a fragment of ideas to the system and then the system stimulates the creativity of the user by showing words generated by the system. As a result of generating a two-dimensional space from the words the user selects, the system presents information characterizing the space. The system assists human intelligent activity.

1 はじめに

近年、われわれの知的活動を支援するシステムに対して様々な分野からのアプローチがなされている。なかでも、情報の整理、検索、通信等で人間の持つ能力をはるかに凌ぐコンピュータの存在が、人間との共存において新しいレベルの知的システムを生み出す結果となっている。今や技術革新によるダウンサイジングのおかげで、個人環境でも大容量メディアによる情報の受給や、ネットワークへのアクセスが、可能となっている。その結果、コンピュータの役割も従来の人間の持つ能力の「代行」といった立場から「支援」へと移り変わってきた。[1, 2]

情報に対するサービスが提供される一方、情報の氾濫による新たな問題が引き起っていることも忘れてはならない。大量の情報の中から、必要な情報を見つけ出す情報検索は、計算機を有効に利用する一つの分野として確立されてきた。しかし、アクセスできるデータベースが増えるにつれて情報過多になる傾向があり、データベースを検索しても、検索した結果を人間が処理し切れないという、事態に陥っている。情報検索の場合、データベースに用いられている検索の付け方を理解したり、データベース言語を用いて検索の問い合わせを作成するためには、通常かなりの専門的な知識が必要とされる。ほとんどのエンドユーザはこのような専門知識をもっていないので、検索を実行するためには熟練したサーチャによる助けが必要となる。この場合、ユーザはデータベースとその構造に関する適切な認知モデルを持っていないため、サーチャがこの食い違いを克服しなければならない。以上の問題はサーチャにとっても過負荷なものであり、情報の洪水に対するフィルタリング技術が望まれている。[3]

現状ではコンピュータが情報の持つ意味を理解するのは不可能に等しく、情報検索法として、キーワードによる検索や、フルテキストサーチが用いられている。しかし、時々刻々と変わるユーザの問題意識や興味に対しての配慮が十分にはなされていない。インタフェースに要求される環境と能力は、情報検索において、そこに介在する人間の探索行動に反映されなければならない。

コンピュータは与えられた情報を処理するだけで、ユーザの意図を理解してくれるわけではない。必要とする情報に対する検索要求を生成するのは人間であり、システムはユーザの要求を明確化する過程を支援する環境を提供する必要があると考える。なかでも計算機上に構築された発想支援システムの典型は「KJ法」[4]

にヒントを得た空間配置を用いることにより支援するタイプ[5]が多く、問題意識の把握/解決に対してその有効性が報告されている。

2 システムの構成

本章では、今回提案する情報フィルタリングシステムの全体構成についてその概略を述べる。

2.1 情報フィルタリングシステムの全体構成

情報フィルタリングシステムの全体構成を図1に示す。

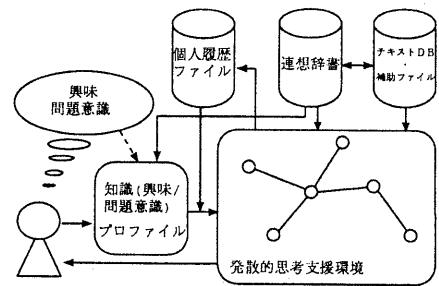


図1: 情報フィルタリングシステム

ユーザは頭の中に思い描く問題意識/興味をテキストの形式で入力する。システムはテキストのキーワードを抽出し、連想辞書の情報を参照することによってプロフィール(ユーザの興味を表す)を作成する。さらに、個人情報を加味するため履歴ファイルを加える。次に、プロフィール情報に基づいてシステムから提示されるキーワードに対して、ユーザが取捨選択を繰り返し、空間配置によって情報に対する要求を整理する。部分あるいは全体として空間配置がまとまった段階でシステムへフィードバックし、操作に応じた情報を得る。ユーザはシステムから得る情報により思考を上げ、空間配置を再変更していく。以上の操作を繰り返し、ユーザは情報要求を明確にしていく。その結果を用い、問題意識にそった情報を得る。

2.1.1 キーワード

本稿で扱うキーワードとは、文章を構成する最小の意味要素(形態素)である。単純に単語と置き換えても構わない。また、未定義語の影響が無視できないため、

既存の辞書を用いるのではなく、実際にテキストから抽出したキーワードを用いる。

2.1.2 テキストデータベース

電子化されたテキストであれば対象とする分野を問わない。ユーザの興味の対象をデータベース化したものであればよい。ブレンテキストを対象とする。

2.1.3 連想辞書

テキストから、キーワードの出現頻度・位置関係等の統計量を使い、連想辞書を構築する。連想辞書を用いてユーザの刺激を促すキーワードを提示する。

2.1.4 補助ファイル

連想辞書の情報を用いテキストファイルにフィードバックし、検索用の補助ファイルを作成する。

2.1.5 発散的思考支援環境

ヒューマンインタフェースとして個人で行うブレンストーミングを支援し、ユーザの情報要求を明確にしていく環境を提供する。発想過程そのものはユーザにまかせ、システムは発想の過程で有用となり得るさまざまな情報を提供する環境を与える。ユーザはキーワードの空間配置を操作することによってシステムから関連キーワードの提示を受け、刺激を促される。空間配置に対するユーザの基本操作として、キーワードの追加/削除/移動、グルーピングを用いる。キーワードのグルーピング操作によって情報要求を集約し、その結果を用い必要な情報をフィルタリングする。

2.1.6 個人履歴ファイル

言葉は人それぞれに違ったニュアンスで使用される。また、興味の対象領域が異なれば使用される語の意味も異なってくる。連想辞書は統計量に基づいて作成されるため平均化がなされる。そのため問題意識の焦点がボケてしまう可能性がある。そこで、ユーザの興味や問題意識をより明確化するために、個人情報やカスタマイズする必要がある。以上の理由から個人情報をより反映させるために、履歴ファイルを作成する。

3 連想辞書

本章では、連想辞書の構築方法について述べる。

3.1 検索要求生成過程

ユーザの要求を満たす情報を得るためには、その情報要求の背景にある問題意識/興味をシステムに理解させる必要がある。すなわち、無機質なコンピュータとの対話において、ユーザ自身どのような情報を求めているかを明確にし、システムに伝えなければならない。しかし、情報に対する要求は常に最初から明確ではなく様々な段階がある。^[6]

- 何か情報が不足していると感じているが、まだそれを言語では表現できない段階
- どのような情報が必要であるか頭の中には描いているが、まだ曖昧で漠然としている段階
- 情報要求が明確に声や文字で表現された段階
- 情報システムに適した表現に変換された段階

通常の検索において、ユーザの意図とシステムの動作との間を仲介する知的な情報伝達(翻訳)の形式を必要とする。ほとんどのユーザはデータベース特有の構造や実行に関する知識、手続き的な専門知識、領域知識、戦略知識などを持っていない。このことから一般には、情報検索に対してサーチャのような人間による介在者が、ユーザとコンピュータの間のインタフェースとして作用することが多い。

システムが提供するインタフェースとしてユーザ/システムの相互作用における操作の中で、ユーザの情報要求を明確化するための付加的な関連情報を与える機能が提供されるべきであると考えられる。ユーザの不十分な知識や戦略のないアドホックな検索では、意図している要求を明確にシステムに伝えることは不可能である。

ユーザは様々なキーワードの組合せにおいて検索を行なう。ところが認知的負荷の制限による制約により、意識的/無意識的にどの情報に注意を向けるかを決めなければならない。また、記憶に対する負荷により、発想はすぐには生まれない。新しい発想を妨げる要因として、知識の不足、連想が拡がらない、意識の固着により新しい枠組みを見いだせない、といったことが挙げられる。これは普段慣れ親しんでいる思考のパターンに流されてしまうことや、問題対象領域のどこからどのように手をつければ良いかを考える訓練の不足によるものが大きい。

発想はすでに存在する情報の新しい結合の中に新しい価値をみいだすことによることが多い。我々は外界からの情報により知識を得、視点を変えることにより発想のヒントを得る。発想の過程で様々な情報にふれるこ

との有効性は多くの人が経験しているであろう。発想支援システムを考える場合、情報を自由に得られる環境を提供する必要がある。システムはユーザに対して、情報を提供する環境を用意しなければならない。これまでも工学的な応用例として、ニューラルネットを用いた連想記憶をはじめ様々な研究が行なわれてきた。なかでも連想語を提示するシステムとして情報理論による統計量を用いた Keyword Associator^[7]が注目されている。

本稿では、キーワード間の距離による重みづけによる統計量を用いた連想辞書の構築により、ユーザの刺激を促す情報を提供することを検討する。

3.2 連想辞書の構築

各キーワード間に互いの関連度を表す重みづけのリンクを張り、連想辞書を構築する。各テキストのキーワードの出現頻度・位置関係等の統計量を使用する。以下その方法について述べる。

テキストデータベース全体を対象としてキーワードの抽出を行なう。抽出した N 個のキーワードから N 次元キーワード空間を定義する。各キーワード w_i の単位ベクトルとして以下のように表すことができる。

$$\begin{aligned} w_1 &= (1, 0, \dots, 0) \\ w_2 &= (0, 1, \dots, 0) \\ &\vdots \\ w_N &= (0, \dots, 0, 1) \end{aligned}$$

各テキスト中に現れるキーワードのそれぞれの関係は、文脈の中でより近くに出現するもの程、互いの関連度が大きいと考える。キーワード間の距離を定義することによって、そのテキスト中の注目するキーワードに対するキーワードベクトルを定義する。

cc is the C compiler .

例として "cc is the C compiler." を考える。"cc" に対して "C" は、距離 $d_{cc,C} = 3$ 、同様に "compiler" は $d_{cc,compiler} = 4$ と考える。また、注目するキーワードに対して前後関係を考慮することから、"compiler" に対する "cc" は $d_{compiler,cc} = -4$ とマイナス符号を用いて定義する。よって任意のテキスト T_i (l 個のキーワードからなる) の先頭から j 番目のキーワードにおけるキーワードベクトルを定義でき、

$$k_j^{T_i} = \sum_{k=j-r}^{j+r} \alpha \exp^{-\beta(k-j)^2} w_k$$

ここで、テキストの両端に注意して、

$$1 \leq k \leq l$$

が成り立つ。但し、 r は j 番目のキーワードから前後どれだけの距離に含まれるキーワードを関連キーワードとして用いるかのパラメータとし、 α, β は任意の定数とする。 $(r, \beta$ がキーワードベクトルの特徴づけに大きく影響するパラメータとなる。) また、 $w_k^{T_i}$ はテキスト T_i 中の先頭から k 番目のキーワードが表す単位ベクトルを意味する。

以上に定義したキーワードベクトルを用いて、各キーワード w_i に関する連想キーワードベクトルを定義する。

$$a_i = \sum_{k_1=1}^M \sum_{k_2=1}^{l_{k_1}} \delta_{k_2;i}^{T_{k_1}} k_{k_2}^{T_{k_1}}$$

但し、

$$\delta_{k_2;i}^{T_{k_1}} = \begin{cases} 1 & a_i \equiv k_{k_2}^{T_{k_1}} \\ 0 & \text{otherwise} \end{cases}$$

ここで、 M はテキストの総数、 l_i はテキスト T_i のキーワード総数を表す。(記号 \equiv は両辺同じキーワードに注目していることを表す。)

生成される連想キーワードベクトル中には、連想を刺激する語と一般に良く用いられる語が混在していると考えられる。一般語は各連想キーワードベクトル中での出現頻度も高く、全テキストに対してかたよりなく均等に出現することから、平均をとることにより抽出が可能である。よって連想キーワードベクトルから平均の差分をとることによって、一般に良く用いられる語を抑制することができる。

さらに連想キーワードベクトルの平均を計算する。

$$a_{ave} = \sum_{k=1}^N a_k$$

各ベクトルをそれぞれ正規化し、

$$\hat{a}_i = \frac{a_i}{|a_i|} \quad \hat{a}_{ave} = \frac{a_{ave}}{|a_{ave}|}$$

(記号 $|a_i|$ は a_i の各成分の自乗和の平方根を表す。) 連想キーワードベクトルは以下のように再定義される。(γ は定数)

$$\hat{a}'_i = \hat{a}_i - \gamma \hat{a}_{ave} \quad (i = 1, \dots, N)$$

3.3 キーワードベクトルへの変換

3.3.1 テキストデータベース

テキストに出現するキーワード毎に出現頻度に応じた値を乗じ、連想キーワードベクトルを加算していく。

これを正規化したものをテキストに関する補助ファイルとして用いる。(P(w_i)は全テキスト中でのキーワードの出現確率)

$$t_i = \sum_{i=1}^M \frac{1}{P(w_i)} \hat{a}_i \quad \hat{t}_i = \frac{t_i}{|t_i|} \quad (i = 1 \dots M)$$

3.3.2 個人履歴ファイル

個人履歴ファイルはユーザが実際にシステムを使用することによって得られる空間配置に基づくキーワードベクトルを記録として用いることによって構築される。システムを幾度も使用することによってユーザの興味の対象領域を反映させるものとなる。空間配置はキーワードベクトルとして表現することができ(cf.4章)、個人履歴ファイルにベクトル*Rec*として記録する。

3.3.3 問題意識

ユーザは問題意識や興味をテキストの形式で入力する。このテキストに基づいてテキストデータベースの補助ファイル*t_i*と同様に、問題意識を表すキーワードベクトル*Con*を作成する。さらに個人履歴ファイルを用い、ユーザの個人情報を反映させる。

$$Con' = (1 - \delta)Con + \delta Rec \quad (0 \leq \delta < 1)$$

但し、 δ はユーザの履歴を反映させるための定数。

*Con'*を各要素の重み順にソーティングしユーザにそのキーワードを提示する。ユーザはテキスト情報から一旦キーワードに置き換え、システムによって提示されるキーワードに刺激を受けながら、空間配置を形成するなかで情報要求を明確化していく。

4 発散的思考支援環境

本章ではシステムが提供する思考支援環境について述べる。

4.1 思考支援環境

人間の思考は発散的思考と収束的思考の2種類に大別される。^[9]発想の過程では発散的思考を行っている間、発想を展開させるために一つのタスクから他のタスクへの移動を頻繁に生じる。さらにレベルの異なるタスク間でも移動を行う。このことからシステムはユーザに対して、情報の管理/整理を容易に取り扱うことができる環境を用意しなければならない。

以上の問題に対し、計算機上に構築された発想支援システムの典型は「KJ法」にヒントを得た空間配置を

用いることにより支援するタイプが多く、問題意識の把握/解決に対してその有効性が報告されている。問題の基本構造の認識に関して、空間配置を用いるメリットとして、以下のような点が挙げられる。

- 全体の構造の把握
- 全体の中での個々の位置付け
- 任意の部分空間を思考対象化可能

本稿で提案するシステムは、情報に対する要求として問題意識/興味を文章の形式で記述し、それに基づくプロフィールを作成する。ユーザは文章をキーワードに置き換え、キーワードオブジェクトを空間配置の中で操作することによって、連想辞書からの関連情報の提示により刺激を受け、情報を集め/整理する。KJ法的な空間配置を用いたブレインストーミングは知識を階層構造化し整理するのに役に立つが、ユーザの問題意識/興味の対象領域をシステムに理解できる形で表現しなければ、必要な情報を得ることができない。

できるだけ容易な操作でユーザに負担をかけることなく要求に応じた情報を提供するシステムが構築されなければならない。本稿ではシステムに対するユーザの操作として、キーワードの空間配置変更(移動/削除/追加)とグルーピングを対象とする。ユーザはグルーピングによって、情報の要求を集約し明確化していく。

4.2 情報の集約

4.2.1 グルーピング

ユーザは空間配置を続けることによって情報の要求を整理する。ある程度思考がまとまった段階で情報を知識として集約することを試みる。部分/全体としてのグルーピング操作に対して関心と観点に基づいた集約を提案する。この操作によって1語へ情報の集約、あるいは関連情報を得ることによって新たに情報の追加を行なうことができる。

4.2.2 関心に基づく集約

関心を表す集約(図2)に関して、中心からの距離に応じた重みづけを行なうことによってグルーピングされたキーワードをまとめる。関心の中心をユーザが指定することによって情報の集約を行なう。

中心は必ずしもキーワードでなくてもよく、以下の式で定義される。

$$I = \sum \frac{1}{r_i + 1} \hat{a}_i$$

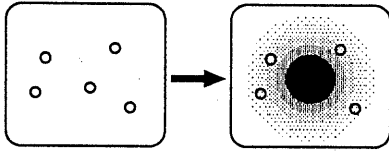


図 2: 関心

但し、 r_i は中心からの距離を表す。また、 \sum はユーザがグルーピングしたそれぞれの連想キーワードベクトルに距離による重みづけを行なったものの加算和をとることを意味する。(以下観点の場合も同様)

4.2.3 観点に基づく集約

観点を表す集約(図3)に関して、ユーザは空間配置の中で、しばしば軸を用いた配置を行なう。観点の軸をユーザが書くことによって情報の集約を行なう。

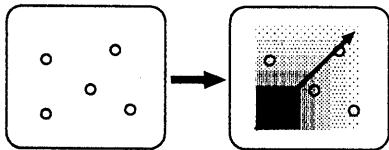


図 3: 観点

これは以下の式で定義される。

$$V = \sum \left(\frac{1}{r_i^h + 1} + \epsilon \frac{1}{r_i^v + 1} \right) a_i'$$

但し、 r_i^h は軸の始点から水平方向の距離、 r_i^v は軸に対する垂直方向の距離を表し、 ϵ は軸から離れたことによる重みづけを表すものとする。

4.3 関連情報の提示

4.3.1 連想度/上昇度

ユーザの発想をより刺激するキーワードを提示するための指標として、連想度と上昇度を導入する。

システムによって提示されるキーワードは、予想できるものや思いもつかなかったもの等さまざまである。一般にユーザが必要とする情報は個人によって違うし状況によっても異なるためその区別は困難である。

ユーザがまとめあげた空間配置によるキーワード空間のベクトル合成ではせいぜいユーザが思いつく連想

語が提示されるにすぎない。発想を妨げる要因として、多くの容易に連想される語が発想のステレオタイプ化を引き起こし、意識の中で大きく浮かび上がってくる重要となるかもしれない情報を抑制していると考えられる。空間の中で容易に連想される語を提示するだけでなく、大きく浮かび上がってきた語にも注目することで新たな展開を生じる可能性がある。

システムは連想空間を以下の2つの視点に基づいて提示することを考える。

- 情報の連想率(連想度)
- 情報の上昇率(上昇度)

4.3.2 キーワードの提示

キーワードの提示はユーザの興味の程度によってその内容を変化させる必要がある。ここでは、連想度と上昇度による視点の違う2つの指標を用いてその対応を行なう。

連想度は1語以上のキーワードのグルーピング操作に対して定義される。例えば1語の場合、連想キーワード中の重みの大きな成分から順にソーティングしてやることによって順位づけをすることができる。また、2語以上の場合にはベクトル I や V によって定義されたキーワードベクトル空間に対して重み順にソーティングすることによりユーザに提示される。連想度は空間を最もよく満足する解を導き出す指標である。この操作では通常良く用いられる語が提示されると予想される。

上昇度は1語に対して定義することはできず、2語以上のグルーピング操作が対象となる。複数のキーワードに注目したことにより浮かび上がってきた情報を取り出す指標である。ここでは連想キーワード内の重み順ソーティングによる順位に対して点数づけを行なう。点数は順位によって与えられるものとする。上昇度は集約前の各キーワードベクトル内での点数を合計したものから集約後の点数を引くことによって得られる。以下に式を定義する。

$$S_{I(or)V}^{w_i} = \left(\sum^n s_{a_j}^{w_i} \right) - n \dots I_{(or)V}^{w_i}$$

但し、 $s_{a_j}^{w_i}$ 、 $s_{I(or)V}^{w_i}$ はベクトル合成前後の w_i の点数、 $S_{I(or)V}^{w_i}$ は w_i に関して関心(or観点)の上昇度の点数となる。全ての w_i について演算を行ない $S_{I(or)V}^{w_i}$ を大きい順にソーティングしたものを提示することによって情報としての上昇度を知ることができる。

実際にはユーザの刺激に応じ、連想度と上昇度を併用して用いる。連想度/上昇度それぞれによる点数づけ

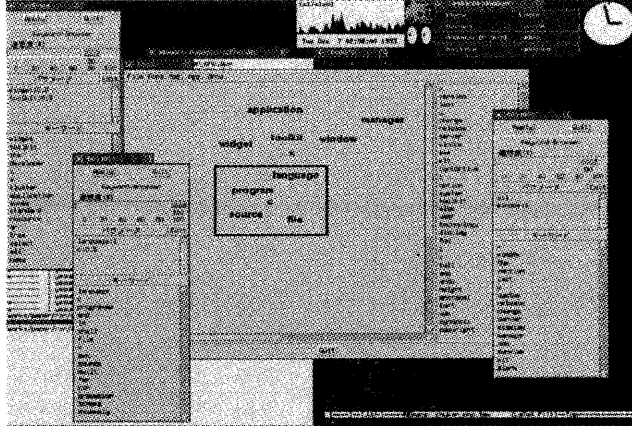


図 4: システム使用例

を行ない、それぞれに重みを乗じたものを加算したものに對してさらにソーティングを行なう。どちらにどれくらいの重みを置くかはユーザ次第である。以上の操作により、単なるベクトル空間の合成では乗り越えることのできなかつたキーワードがユーザに提示され、ユーザの発散的思考が促進される。

5 情報フィルタリング

5.1 ユーザの情報要求

ユーザの情報に対する要求を明確化するため、連想辞書による関連情報の提示と、空間配置により要求をまとめあげる環境を提供した。以上の環境の中でまとめあげたユーザの情報要求を用いて、ユーザは必要とする情報を得る。

ユーザの情報要求に対して検索を行なう対象としてテキストデータベースを考える。データベースそのものは受動的なシステムであり、与えられた要求に対して能動的な答を返してくれるわけではない。いかにして必要な情報にたどり着くか、かつ、限られた制約の中で解を出していくかが焦点となる。

5.2 テキストの提示

空間配置の内容をベクトル I や V によって表すことを提案した。これらの情報を用いたテキストの各補助ファイル i_j との比較により情報のフィルタリングを行なう。ここでは順位づけを用いた得点計算による戦略を用いた計算量の低減化を検討する。

I 、 V あるいは i_j 中の各要素の値は対象とする空間の中で w_i がどれ程の影響を与えるかに関するパラメータである。重みの大きいほどその空間内を反映することから、 I や V に関して重み順にキーワード w_i に注目する。各テキスト i_j の w_i に対して重みづけによるソーティングを行ない、順位づけによって各テキストに点数を与える。 w_i を数語から数十語用いることにより各テキストに点数を加算していく。以上の操作で各テキストに対して得点づけがなされ、得点の小さい順にユーザに提示してやることによってユーザの要求にこたえる。得点づけによるソーティングを行なった結果として、情報の洪水に対する制御ができ、ユーザに負担をかけることなくテキストの提示を行なうことができる。

6 実行例

現在インプリメント中のシステムの実行結果の一例を図4に示す。

システムの動作確認のため便宜上UNIXのオンラインマニュアルを使用した。(将来的には対象変更を予定) 現段階では各パラメータに関する十分な評価/調整は行なっていない。そのデータは以下のようなものである。

- 処理したファイル数:約 4000
- 抽出したキーワード数:約 25000(基本形)
- 連想辞書の大きさ:約 100MB

構築した連想辞書から、キーワード“x”に関するキーワードベクトルの上位/下位 16 語を表1に示す。実験

