

SD式意味モデルを用いた 英日会話文意味検索システムの試作

吉原 将太¹⁾・脇山 正博²⁾・河口 英二³⁾

A Prototype of Semantic Retrieval System for Conversational English/Japanese Texts Using SD-Form Semantics Model

Shota YOSHIHARA, Masahiro WAKIYAMA and Eiji KAWAGUCHI

長崎純心大学 人文学部 英語情報学科 講師¹⁾

北九州工業高等専門学校 制御情報工学科 教授²⁾

慶應義塾大学 デジタルメディア・コンテンツ統合研究機構 教授³⁾

Abstract

The SD-Form Semantics Model, developed by the authors, is a framework to deal with the meaning of natural language in a quantitative way. Using this model, we are constructing a conversational-text database of English & Japanese language combined with semantic data described by the SD-Form. The feature of this model is the most remarkable in the point that it can calculate the difference of the meaning in two expressions. This paper describes a prototype of the conversational English/Japanese texts semantic retrieval system using SD-Form semantics model. We have found the effectiveness of our method.

1. はじめに

日本人が英語で話そうとするとき、先ず英語の文法の知識を借りて日本語を英語に翻訳しようと試みることが多い。しかし、文法のみを頼るやり方では適切な会話表現とはならない。即ち、会話で使う口語的英語表現は、書き言葉的な英語表現とは相当違っていて、この口語的英語表現の知識がなければ、会話の際に話したり聴きとったりすることが困難である。

英会話の学習には、従来から「様々な会話場面を想定した模範会話例文集」が教材として利用されている。学習者は、できるだけそのような例文を暗記しようと試みるが、現実に暗記した例文がそのまま使える場面に遭遇することはあまりない。また、「英会話辞書」を持っていたとしても、言いたい意味と一致する英語表現を探し出すことは容易ではない。なぜなら、辞書には表

現したい意味の見出し語が欠落していたり、思い浮かべた見出し語に別の意味の英語表現が載せてあったりするからである。このような不都合が生じるのは、従来の辞書が「意味データ」を検索キーとして利用できないからである。もし、伝えたい意味そのものの表現データが容易に作れ、さらに、そのような意味表現を「意味見出し語」とする辞書があり、各見出し語には会話状況に応じた「英語」や「日本語」が対応させてあれば、状況は異なる。

多くの会話文検索システムは、会話文中の語をキーワードとして検索するものや^[1]、あらかじめ用途別に分類された会話文の例文をユーザーが選択するといったものである^[2]。しかし、このような検索システムでは、上述したような問題点は解消されない。そこで、本研究では、表現したい事柄（文）を検索キーとして、その意味により近い表現を複数個出力する検索システムを試作する。本システムでは、検索結果として出力された会話文がどのような場面で用いられたものが分かるように、その会話が行われている場面全体の会話文集合データ（シーンデータ）も出力できるようにしている。このようなシステムの実現のためには、「検索キー」と「意味見出し語」間の相対的な類似度（意味差）が計算できる仕組みが必要となる。

近年、著者等の研究グループは、自然言語概念の意味表現形式の一つとしてSD式（Semantic structure Description Form）を提案している^[3, 4, 5]。SD式は、自然言語における個々の概念、陳述表現、感情表現、あるいはシステムに与える知識データ等を記述するための一種の中間言語であり、その構文は、曖昧さの無い文脈自由文法で規定されている^[3]。自然言語概念をSD式として捉え、その記述データを基にして意味処理を行おうとするモデルをSD式意味モデルと言う。SD式意味モデルの最大の特徴は、従来からの意味記述モデルでは扱い難かった2つの概念間の意味的な近さを定量的に扱える点である。

著者らは、15～20年にわたって「NHKラジオ英会話テキスト」や「NHKテレビ英会話テキスト」などを基にした英会話例文の「英日会話文データベース」の構築を行っている^[6]。これは、主に会話文（英語と日本語）とそれに対応するSD式で記述した意味データ、および会話が行われている背景を述べたデータ（背景データ）からなる。

本研究の目的は、この英日会話文データベースを用いて、「会話文を意味的に検索」するシステムをWeb上に実現することである^[7]。このシステムを「英日会話文意味検索システム（以下、本システム）」と呼ぶ。本システムは、検索キーとしてSD式を用い、完全に一致するものがなくても、意味的にできるだけ近いものを検索結果とするものである。この検索の過程において、SD式意味モデルにおける意味差の尺度を利用する。

意味差の計算には、多くの計算時間を要する。ゆえに、データベース中の全ての会話文との意味差を計算するのは非効率的である。また、SD式には多様な概念ラベルを用いることが可能であるため、異なる概念ラベルが用いてあったとしても、意味的には同じような内容であることがある。この場合、検索キーのSD式に用いられているラベルと一致するものだけを検索対象としたのでは、意味的に近くても検索対象から漏れてしまうことになる。そこで、本研究では、概念ラベル間の類似性に関しては、国立国語研究所の分類語彙表^[8, 9]を用い、意味的に近そうなデー

タが漏れないようにして検索対象のデータを絞り込むこととした。そして、絞り込んだデータと検索キーとの意味差を計算し、より近い意味の会話文を検索できるようにした。

以下、2章ではSD式の記述例とSD式意味モデルにおける意味差の尺度について述べる。3章では、英日会話文データベースについて、4章では、英日会話文意味検索システムについて述べる。5章では、検索例を用いて本システムの処理手順を示す。最後に6章で、まとめと今後の課題を述べる。

2. SD式意味モデルの概要^[3, 4, 5]

SD式意味モデル (Semantic-structure Description Form Semantics Model) は、自然言語の意味を定量的に分析するための枠組みである。このモデルに従った意味記述がSD式である。

2.1 会話文のSD式記述例

会話文は、①発話意図を含む陳述形式のSD式 (発話意図SD式)、または②感情SD式で記述する。

発話意図とは会話において話し手が聞き手に伝達したい意図のことである。発話意図のラベルとしては、質問・告げる・依頼・命令・許可・提案・陳謝・感謝・確認などを設定している。発話内容は陳述SD式で記述する。発話意図SD式の構文は次の通りである。

[s(発話者), v(発話意図), o(対話相手), c(発話内容)]

英語で5W1H型と呼ばれている疑問文は、「発話意図」の部分で「質問」とし、「発話内容」の部分で5W1H型のSD式ラベル (何時、何処、誰、何、何故) を用いた陳述SD式で記述する。HOW疑問詞については、How many、How farなどいろいろな種類の疑問文があり、一意に決めることは難しい。発話意図SD式の例を以下に示す。

〈例1〉

・ [s(自分), v(命令), o(相手), c([s(相手), v(開ける), o(窓)])]

：窓を開けなさい。

・ [s(自分), v(質問), o(相手), c([s(相手), v(行く / (過去) para(時/昨日) para(場所/何処)]))]

：あなたは昨日何処に行ったのですか？

感情SD式は、自然言語による感情的な発声や発話の状況の記述に用いる。感情SD式では、呼びかけ・応答・感嘆の3種類を定義している。感情SD式の例を以下に示す。

〈例2〉

・ [a(ジョン)] : ジョン (呼びかけ)

・ [r(肯定)] : はい (応答)

・ [e(賞賛)] : すばらしい! (感嘆)

2.2 SD式の意味的情報量

SD式では、記号列の構造で何か固有の概念を表現しようとするだけでなく、その概念の意味量の大小も表すこととしている。そのため、SD式意味モデルでは、各SD式記号に意味素量を設定している。任意のSD式をDとすると、その意味量を

$$si(D) = n \text{ [semit]}$$

と表す。また、その単位をsemitとしている。意味素量を以下のように設定している。

(1) 変数ラベル	“X, Y, Z, ...”	: 1 [semit]
(2) 単純ラベル	“馬, CAT, ...”	: 10 [semit]
(3) 修飾子	“/”	: 1 [semit]
(4) 規定子	“nega, only, ...”	: 2 [semit]
(5) 結合子	“incl, kdof, ...”	: 1 [semit]
(6) 機能項目記号	“s, v, c, o, ...”	: 1 [semit]
(7) 区切り記号	“[]”	: 1 [semit]
(8) 区切り記号	“()”, “,”	: 0 [semit]

SD式全体の意味量は、それらの総和となる。

SD式意味モデルでは、これらを定義するときの一般的指針は示しているが、意味量の値については、モデルの利用者が独自に定めて良いとしている。

〈例3〉

$si(\text{帽子/赤い}) = 21$: 赤い帽子
$si(\text{(男)plus(女)}) = 21$: 男と女
$si([s(\text{美紀}),v(\text{である}),c(\text{先生/英語})]) = 45$: 美紀は英語の先生である。

2.3 SD式意味モデルにおける意味差の尺度

2つのSD式概念D₁、D₂に関して、D₁の意味をより具体化したものがD₂であり、逆にD₂の意味をより抽象化したものがD₁であるとき、D₁とD₂には「詳述関係 (elaboration relation)」があるという。この関係を

$$elab(D_1, D_2) = n$$

と表す。このときのnを「詳述量 (単位: semit)」と呼ぶ。このようなD₁をD₂の先祖、D₂をD₁の子孫と呼ぶ。ただし、詳述関係が成り立たない場合は、

$$elab(D_1, D_2) = \infty$$

と定義している。

SD式意味モデルでは、概念間の「意味差の尺度」を次のように定義している。2つの概念D₁、D₂の全ての先祖D₀₁、D₀₂、…、D_{0i}、…の中でD₁、D₂に最も近い先祖を「D₁、D₂の最近共通先祖」と呼び、D₀で表す。

$$elab(D_0, D_1) + elab(D_0, D_2)$$

$$= \min\{elab(D_{0i}, D_1) + elab(D_{0i}, D_2)\}$$

$$= n_1 + n_2 = n_0$$

この関係を

$$ncoa(D_1, D_0, D_2, n_1, n_0, n_2)$$

または

$$ncoa(D_1, D_0, D_2) = n_0$$

と表す。このときの n_0 を「 D_1 と D_2 の意味差」といい、

$$diff(D_1, D_2) = n_0$$

と表す。すなわち、意味差は与えられた D_1 、 D_2 の最近共通先祖を探索することにより求められる。

3. 英日会話文データベース^[6,7]

本研究における英日会話文データベースは、原資料を「NHKラジオ英会話テキスト」(1989年4月号～1997年12号)として、「背景データベース」と「会話文データベース」の2つで構成している。

3.1 背景データベース

背景データとは、会話が行われている状況(背景)を記述しているデータであり、「背景データ番号」、「タイトル(英語)」、「タイトル(日本語)」、「タイトル(SD式)」、「背景文(英語)」、「背景文(日本語)」、「背景文(SD式)」、「背景知識」の8種類で構成したものである。背景データベースとは、これらの背景データをデータベース化したものである。

背景データの例を表1に示す。

表1 背景データの例

種類	データ
背景データ番号	r971023
タイトル(英語)	A Visitor
タイトル(日本語)	訪問客
タイトル(SD式)	訪問客
背景文(英語)	Sam puts a large box on the table.
背景文(日本語)	サムはテーブルの上に大きな箱を置く。
背景文(SD式)	[s(サム),v(置く/場所/上/テーブル),o(箱/大きい)]
背景知識	(人間)incl(サム) (人間)incl(ナオコ)

(1) 背景データ番号

背景データ番号は、次の7桁で表記する。

SYMMWN₁

- S** : 原資料の種類 (ただし、本研究で用いた原資料は「NHKラジオ英会話テキスト」であり、これを“r”と表記している)
- YY** : 年
- MM** : 週
- N₁** : 各週における背景文の連番

〈例4〉

r971023 : NHKラジオ英会話テキスト1997年10月号、第2週 第3番目の背景データ

(2) 背景データ番号

原資料では、各週の単元にタイトル (英語文/日本語文) が付けられている。これを、英語、日本語、それに対応するSD式の3種類の形式で背景データベースに登録している。

(3) 背景文

原資料には、会話が行われている背景が日本語および英語で記述されている。この日本語および英語、それに対応するSD式に登録している。

(4) 背景知識

会話が行われている場面に関連した情報を、知識データとして登録している。

3.2 会話文データベース

会話文データとは、個々の会話文を記述したデータであり、「背景データ番号」、「会話文データ番号」、「話者 (英語)」、「話者 (日本語)」、「会話文 (英語)」、「会話文 (日本語)」、「会話文 (SD式)」、「重要文指定」の8種類で構成したものである。会話文データベースとは、これらの会話文データをデータベース化したものである。

会話文データの例を表2に示す。

(1) 背景データ番号

その会話文が使用されている場面に対応する背景データの番号を、3.1節 (1) の形式で登録している。

(2) 会話文データ番号

会話文データ番号は、背景データ番号に3桁の会話文の連番を結合したものであり、次の10桁で表す。

SYMMWN₁TTN₂

- TT** : 話者の連番 (ただし、同一人物が数度登場し発話しても、それぞれ1回の話者とする)
- N₂** : 1回の話者による会話文の連番

〈例5〉

r971023070: NHKラジオ英会話テキスト1997年10月号、第2週 第3番目の場面において、第7番目の話者による0番目（最初）の会話文データ。

(3) 話者

その会話文の話者を英語と日本語で登録している。

(4) 会話文

英語文1文を基準として、英語文、日本語文、およびそれに対応するSD式を登録している。

(5) 重要文指定

原資料において、「会話に役立つ重要表現」としてマークされている文を重要文としている。重要文の場合1を、重要文でない場合は0を登録している。

表2 会話文データの例

種類	データ
背景データ番号	r971023
会話文データ番号	r971023070
話者（英語）	Sam
話者（日本語）	サム
会話文（英語）	Have dinner with me.
会話文（日本語）	僕と食事をしてくれないか。
会話文（SD式）	[s(サム),v(依頼),o(ナオコ),c([s(相手),v(食事/随伴/自分)])]
重要文指定	0

4. 英日会話文意味検索システム

英日会話文データベース（3節参照）を用いて、「会話文を意味的に検索」するシステムをWeb上に試作した。ただし、自然言語文からSD式データへの自動変換システムは現在開発中であるため、本システムの現段階では、検索キーはSD式で記述することとした。本システムの特徴は、完全に一致するものがなくても、意味的にできるだけ近いものを検索結果とするものである。この検索の過程において、次の意味処理を行う。

(1) 分類語彙表^{18,9)}による検索対象の絞り込み処理

(2) 意味差の計算処理

検索結果は、検索キーに意味的に近いと判断された会話文データおよびシーンデータ（その会話が行われている場面全体の会話文集合データ）を出力する。

4.1 意味検索処理

分類語彙表は、単語を「体の類（名詞の仲間）」、「用の類（動詞の仲間）」、「相の類（形容詞の仲間）」、「その他」の4つに大分類したシソーラスである。

SD式の意味ラベルは、既成の単語（英語や日本語）を使用する。そのため、多様な概念ラベルを用いることが可能であるが、異なる概念ラベルが用いてあったとしても、意味的には同じような内容であることがある。そこで、SD式に用いられているラベル間の類似性に関しては分類語彙表を用い、意味的に近そうなデータが漏れないように検索対象のデータを絞り込むようにしている。

使用した「分類語彙表[フロッピー版]^[8]」には、語彙表と索引が収録されている。語彙表には、語を意味的に分類し、分類番号と見出し語が与えられている。本システムでは、語彙表のみを用いた。語彙表の一部を図1に示す。

1.3332 食	
1	食事 食 齋 飲み食い 飲食 酒食 酒色 腹ごしらえ 舌鼓 早飯
2	試食 試飲 愛飲 立食 立ち食い 買い食い つまみ食い 盗み食い 食べ歩き バイキング
3	大食 少食 過食 多食 飽食 暴飲 暴食 鯨飲馬食 痛飲 深酒 食べ過ぎ 飲み過ぎ 食い上げ
4	欠食 絶食 断食

図2 語彙表のデータ例

分類語彙表を用いた絞込み処理手順を以下に述べる。ただし、入力する検索キーはSD式としているが、そのSD式が「発話意図SD式」の場合と「感情SD式」の場合がある（2.1節参照）。

(1) 発話意図SD式の場合

①入力文の意味を表すSD式をD_{IN}とする。D_{IN}の「発話内容」の部分からラベルを抽出し、リスト1とする。

リスト1: U₁, U₂, …, U_n

ただし、“自分”、“相手”、“当該”などのようなSD式記述ルールにおいて、一意に決められたラベルは除外する。

②リスト1のラベルU_nの同義語を分類語彙表から抽出し、リスト2に加える。このときU_nも加える。

リスト2: U₁, C_{U11}, C_{U12}, …, U_n, C_{Un1}, C_{Un2}, C_{Unm}

③次の条件を満たすSD式を英日会話文データベースから抽出する。得られたSD式をD₁, D₂, …, D_kとする。

・「発話意図」のラベルにおいて、入力SD式と一致している。または、背景知識に

(*assu(D_i)caus(D_j)*)の形式の知識データが登録されており、入力SD式とD_iまたはD_jと詳述関係がある(D_i、D_jはSD式)。

・「発話内容」にリスト2のラベルを1つでも持つ。

- ④D_{IN}と③の絞込み処理で得られたD_kとの意味差を求め(2.3項参照)、意味差が小さい順に会話文データを出力する。

(2) 感情SD式の場合

- ①入力文の意味を表すSD式をD_{IN}とする。D_{IN}の全てのラベルを抽出し、リスト3とする。

リスト3: E₁、E₂、・・・、E_n

- ②リスト3のラベルE_nの同義語を分類語彙表から抽出し、リスト4に加える。このときE_nも加える。

リスト4: E₁、CE₁₁、CE₁₂、・・・、E_n、CE_{n1}、CE_{n2}、CE_{nm}

- ③リスト4のラベルを1つでも持つ感情SD式を、英日会話文データベースから抽出する。得られたSD式をD₁、D₂、・・・、D_kとする。

- ④D_{IN}と③の絞込み処理で得られたD_kとの意味差を求め、意味差が小さい順に会話文データを出力する。

ラベルの同義語とは、分類語彙表(語彙表)において、同一の分類番号と段落番号をもつものである。例を以下に示す。

〈例6〉

ラベル: 食事

同義語: 食 齋 飲み食い 飲食 酒食 酒色 腹ごしらえ 舌鼓 早飯

(分類番号: 1.3332、段落番号: 1)

SD式では、区切り記号“++”を用いて複数のSD式を連結して1つの意味を表すことができる。英日会話文データベースには、このようなSD式で記述したデータも存在する。この場合には、“++”の前後のSD式についてリスト2またはリスト4のラベルを用いているかを調べ、次のパターンで対象データを絞り込む。ここで、D₁、D₂は任意のSD式としたとき、対象データの形式がD₁++D₂であるとする。

- ①D₁のみにリスト2またはリスト4のラベルが用いてある場合は、D₁のみを意味差の計算対象とする。
 ②D₂のみにリスト2またはリスト4のラベルが用いてある場合は、D₂のみを意味差の計算対象とする。
 ③D₁およびD₂ともにリスト2またはリスト4のラベルが用いてある場合は、D₁++D₂を意味差の計算対象とする。

4.2 システムの概要

英日会話文意味検索システムは、Perl Ver. 5.8.0 for MSWin32でプログラミングし、Web上で動作するようにしている。翻訳処理の過程で意味差を計算しているが、その処理は、SD式意味モデル実験システムSDENV-4^[10]の機能を利用している。SDENV-4も、Perlでプログラミングしており、SD式意味モデルの応用システムでも利用できるように、それぞれの処理をサブルーチンとして記述している。

英日会話文意味検索システムの入力ページを図2に示す。

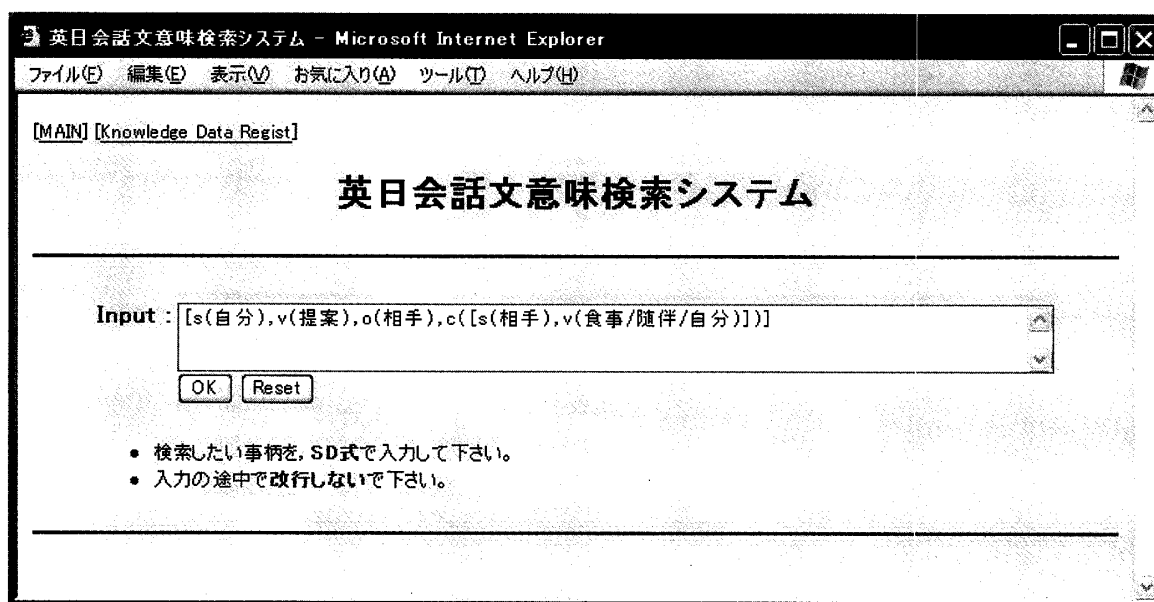


図2 本システムの入力ページ

本システムで検索する際には、図2において、“Input”の欄に検索したい事柄をSD式で記述する。ただし、現在の本システムでは、1文のみの入力に限定する。“OK”ボタンを押すと意味検索処理が実行され、結果が表示される。結果は、「会話文 (SD式)」、「会話文 (英語)」、「会話文 (日本語)」、「意味差」の情報を表示する。このとき、意味差が小さいものから順にリスト表示される。

検索結果の表示例を図3に示す。

英日会話文意味検索システム - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

[MAIN] [Knowledge Data Regist]

英日会話文意味検索システム

INPUT (SD-Form) : [s(自分),v(提案),o(相手),c([s(相手),v(食事/随伴/自分)])]

Result :

背景データ番号	シーン番号	英語 / 日本語
r971023	2	[s(サム),v(依頼),o(ナオコ),c([s(相手),v(食事/随伴/自分)])]
		Have dinner with me.
		僕と食事をしてくれないか？
r970321	4	[s(ウェイン),v(質問),o(レーシー),c([s(相手),v(nesa(構う/随伴/自分)),o(食事)])]
		Do you mind having dinner alone with me?
		僕と2人だけで食事するのかまわない？
r930622	7	[s(クミ),v(依頼),o(アレックス),c([s(自分),v(ご馳走する),i(相手),o(昼食)])]
		Let me buy you lunch.
		お昼は私にご馳走させて。
r960624	26	[s(ウェイン),v(依頼),o(白ーコ),c([s(相手),v(食べる/(随伴/自分)para(時/今夜)),o(夕飯)])]
		How about having dinner together tonight?
		今夜、一緒に夕飯を食べないか？
r950442	29	[s(リサ),v(告げる),o(七ロキ),c([s(自分),v(連れて行く/(mood/欲求)para(昼食)),o(相手)])]
		They're expecting to take you to lunch.
		あなたを昼食に連れて行きたいと思っているのよ。
r970231	29	[s(ウェイン),v(依頼),o(レーシー),c([s(相手),v(来る/(夕食)para(時/木曜日)])]
		Will you come over for dinner on Thursday?
		木曜日に夕食に来ない？
r890421	52	[s(スコット),v(提案),o(エミ),c([s(我々),v(話す/間/[s(我々),v(食べる/場所/どこか),o(ハンバーガー)])]
		Let's grab a hamburger somewhere and talk.
		どこかでハンバーガーでも食べながら話そう。

図3 検索結果の表示例

検索結果表示ページ中の「背景データ番号」をクリックすると、その会話文が使われている場面全体の会話文集合データ（シーンデータ）が表示される。

シーンデータを表示したページを図4に示す。

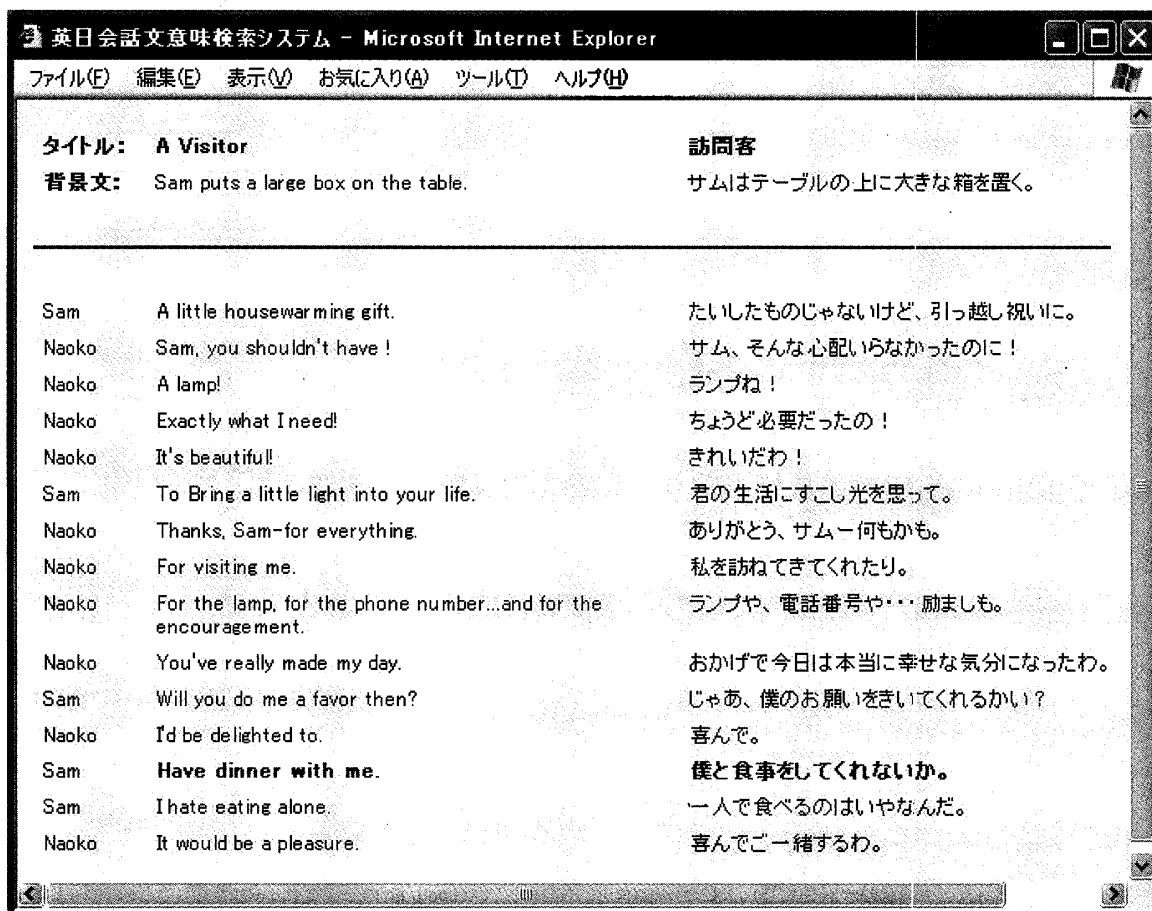


図4 シーンデータの表示例

5. 実験

英日会話文意味検索システムの実験を行い、その有効性を評価した。実験の一つを用いて、意味検索処理を例示する。

5.1 意味検索処理の例示

本研究における英日会話文データベースは、「背景データベース」と「会話文データベース」の2つで構成しており、現在のデータ件数は、背景データベースが約1,000、会話文データベースが約15,000件である(3節参照)。しかし、SD式データは、手作業で作成しており、その登録作業は完了していない。そこで、今回はシステムの動作検証を主とした基礎実験として、会話文データの件数を100件に限定して実験を行った。

以下に、英日会話文意味検索システムを用いた実験例を、処理手順を追いながら示す。ただし、自然言語文からSD式への自動変換プログラムは、現在開発中であるため、入力文はSD式とした。

入力SD式を次のD_{IN}とする。

D_{IN} : [s(自分),v(提案),o(相手),c([s(相手),v(食事/随伴/自分)])]

(一緒に食事をしませんか?)

- (1) D_{IN} は、発話意図SD式であるので (2.1節参照)、発話内容の部分からラベルを取り出す。

リスト1: 食事、随伴

- (2) リスト1のラベルの同義語を分類語彙表から抽出し、リスト2とする。

リスト2: 食事、食、飲み食い、飲食、・・・、随伴、同伴、同道、同行、・・・

- (3) 4.1節 (1) ③の条件を満たすSD式を、英日会話文データベースから抽出する。そのSD式を

D_1 、 D_2 、・・・、 D_i とする。この例の場合、次のSD式がその結果となる。

D_1 : [s(サム),v(依頼),o(ナオコ),c([s(相手),v(食事/随伴/自分)])]

(僕と食事をしてくれないか?)

(Have dinner with me.)

D_2 : [s(クミ),v(依頼),o(アレックス),c([s(自分),v(ご馳走する),i(相手),o(昼食)])]

(お昼は私にご馳走させて。)

(Let me buy you lunch.)

D_3 : [s(リサ),v(告げる),o(ヒロキ),

c([s(自分),v(連れて行く/(mood/欲求)para(昼食)),o(相手)])]

(あなたを昼食に連れて行きたいと思っているのよ。)

(They're expecting to take you to lunch.)

D_4 : [s(ウェイン),v(依頼),o(ヨーコ),

c([s(相手),v(食べる/(随伴/自分)para(時/今夜)),o(夕飯)])]

(今夜、一緒に夕飯を食べないか?)

(How about having dinner together tonight?)

D_5 : [s(ウェイン),v(依頼),o(ルーシー),c([s(相手),v(来る/(夕食)para(時/木曜日)]))]

(木曜日に夕食に来ない?)

(Will you come over for dinner on Thursday?)

D_6 : [s(ウェイン),v(質問),o(ルーシー),c([s(相手),v(nega(構う/随伴/自分)),o(食事)])]

(僕と2人だけで食事するのかまわない?)

(Do you mind having dinner alone with me?)

D_7 : [s(スコット),v(提案),o(エミ),

c([s(我々),v(話す/問/[s(我々),v(食べる/場所/どこか),o(ハンバーガー)]))]]

(どこかでハンバーガーでも食べながら話そう。)

(Let's grab a hamburger somewhere and talk.)

- (4) 入力SD式 D_{IN} と絞り込み処理で得られたSD式 D_i ($i=1, 2, \dots, 7$) との意味差 n_{oi} を求める。

① D_{IN} と D_1 との意味差計算

ここで利用する知識データは、システムに登録されている次の2件である。

- (自分)*equa*(サム) : 自分とサムとは同じである
 (相手)*equa*(ナオコ) : 相手とナオコとは同じであるs

このとき、 D_{IN} と D_1 の最近共通先祖は、

$[s(\text{自分}),v(\text{提案}),o(\text{相手}),c([s(\text{相手}),v(\text{食事/随伴/自分})])]$

となり、意味差 n_{01} は、

$$n_{01} = \text{diff}(D_{IN}, D_1) = 2$$

となる。

② D_{IN} と D_2 との意味差計算

知識データは次の通りである。

- (自分)*equa*(クミ) : 自分とクミとは同じである
 (相手)*equa*(アレックス) : 相手とアレックスとは同じである
 (食事)*incl*(昼食) : 食事は昼食を含む
 $(\text{asuu}([s(X),v(\text{ご馳走する}),i(Y),c(\text{食事})])])\text{caus}([s(X),v(\text{提案}),o(Y),c([s(Y),v(\text{食事/随伴/X})])])]$
 : もしXがYに食事をご馳走するならば、YはXと食事する。

このとき、 D_{IN} と D_2 の最近共通先祖は、

$[s(\text{自分}),v(\text{提案}),o(\text{相手}),c([s(\text{相手}),v(\text{食事/随伴/自分})])]$

となり、意味差 n_{02} は

$$n_{02} = \text{diff}(D_{IN}, D_2) = 7$$

となる。

以下、同様にして D_{IN} と各 D_i との意味差を求めると次のようになる。

$$n_{03} = \text{diff}(D_{IN}, D_3) = 29$$

$$n_{04} = \text{diff}(D_{IN}, D_4) = 26$$

$$n_{05} = \text{diff}(D_{IN}, D_5) = 29$$

$$n_{06} = \text{diff}(D_{IN}, D_6) = 4$$

$$n_{07} = \text{diff}(D_{IN}, D_7) = 52$$

(5) (4) より、意味差が小さい順に会話文データを結果として表示する。

このとき意味差が最小となるSD式は、

$$n_{01} = 2$$

$D_2 : [s(\text{サム}),v(\text{依頼}),o(\text{ナオコ}),c([s(\text{相手}),v(\text{食事/随伴/自分})])]$

である。このSD式に対応する会話文は次の通りである。

英語：Have dinner with me.

日本語：僕と食事をしてくれないか？

4.2節の図3、図4は、この実験例の入力画面と結果画面を示したものである。

5.2 主観評価

本システムの有効性を評価するために、54人の被験者に、10種類の実験結果について、アンケートに答えてもらった。質問内容は、「検索結果に、検索したい事柄の意味を表す会話文が含まれているか」である。その問いに対して「はい」または「いいえ」のどちらかで答えてもらった。ただし、会話文データの件数を100件に限定した実験であるため（5.1節参照）、その中に含まれているような話題（食事、ショッピング、スポーツなど）に関する会話文10個を実験者が考え、それらを入力文（検索キー）として実験した。アンケートの結果を表3に示す。

評価の全体の平均は80.4%となった。どのようなものを検索キーにするかに当然依存するが、調べたい事柄の意味に近い会話文が得られているという評価を得ることができ、本システムの有効性を確認できた。

表3 評価実験結果

入力文	検索結果に意味的に近い会話文が含まれていたと回答した人の割合
それはいくらですか？	100.0%
一緒に食事をしませんか？	100.0%
おなかが空いたな。	81.5%
テレビをつけて。	25.9%
久しぶりだね。	88.9%
今、忙しい？	100.0%
あなたへのプレゼントです。	55.6%
きれいだなあ。	100.0%
お願いがあります。	88.9%
一緒にテニスの練習をしよう。	63.0%
平均	80.4%

6. おわりに

英日会話文データベース（3節参照）を用いて、「会話文を意味的に検索」するシステム（英日会話文意味検索システム）をWeb上に試作した。意味検索処理の過程で意味差を計算しているが、その処理は、SD式意味モデル実験システムSDENV-4^[10]の機能を利用している。ただし、意味差の計算には多くの計算時間を要するため、国立国語研究所の分類語彙表^[8, 9]を用いて、意味的に近そうなデータを絞り込んだ後で、意味差の計算処理を行うようにしている。

54人の被験者による主観評価実験を行い、入力文の意味に近いものが検索結果として得られているという評価を得ることができ、システムの有効性を確認することができた（5節参照）。

今後は、より多くの会話文データに対しての意味検索実験を行っていくことと、明らかに意味的に類似していない結果は表示しないしくみを研究する予定である。

参考文献

- [1] スペースアルク 英語表現辞典
http://home.alc.co.jp/db/owa/ehj_idx
- [2] スペースアルク 英会話トレーニングジム
<http://www.alc.co.jp/eng/speaking/training/index.html>
- [3] Wakiyama,M., Noda,H., Nozaki,K. and Kawaguchi,E.: Computation Algorithm of Semantic Difference Measure in the SD-Form Semantics Model, Trans. IPS Japan, Vol.40, No.3, pp.1065-1079 (1999)
- [4] Wakiyama,M., Yoshihara,S., Noda,H., Nozaki,K. and Kawaguchi,E.: A Discussion on the Property of the Semantic Space in the SD-Form Semantics Model, Transactions of Information Processing Society of Japan, Vol.41 No.2, pp.509-512 (2000)
- [5] Wakiyama,M., Yoshihara,S., Noda,H., Nozaki,K. and Kawaguchi,E.: An Extended SD-Form Semantics Model, Transactions of Information Processing Society of Japan, Vol.41 No.2, pp.513-516 (2000)
- [6] Shao,G., Nozaki,K., Kamata,S. and Kawaguchi,E.: SD-Forms as Interlingua and a Prototype of a Conversational-Text Retrieving System, 人工知能学会 Vol.9 No.5, pp.684-693 (1994)
- [7] 峯脇さやか、吉原将太、脇山正博、河口英二：SD式と分類語彙表を用いた英日会話文意味検索システム、電子情報通信学会技術研究報告 信学技報 Vol.101 No.484, pp.45-50 (2001)
- [8] 国立国語研究所：分類語彙表[フロッピー版]、秀英出版 (1994)
- [9] 国立国語研究所、中野洋：「分類語彙表」形式による語彙分類表、増補版、国立国語研究所 (1996)
- [10] 吉原将太、峯脇さやか、脇山正博、河口英二：CGIを用いたSD式意味モデル実験システムの試作、電子情報通信学会技術研究報告 信学技報 Vol.101 No.484, pp.29-36 (2001)
- [11] 峯脇さやか、吉原将太、脇山正博、河口英二：SD式と分類語彙表を用いた英日会話文意味検索システム、電子情報通信学会技術研究報告 信学技報 Vol.101 No.484, pp.37-44 (2001)
- [12] Masahiro Wakiyama, Shota Yoshihara, Koichi Nozaki and Eiji Kawaguchi : A Prototype of Semantic Retrieval for Video Data Using SD-Form, The 13th European-Japanese Conference on Information Modelling and Knowledge Bases, pp.189-196 (2003)
- [13] Masahiro Wakiyama, Shota Yoshihara, Koichi Nozaki and Eiji Kawaguchi : Prototype of Video Data Semantic Retrieval System Based on SD-Form Semantics Model, The 8th World Multiconference on Systemics, Cybernetics and Informatics, pp.162-167 (2004)