TR－SLT－0035

音声対話処理のための発話を単位とした話題および発話行為タイプの推定

Detection of Topic and Speech Act Type on Utterance-by-utterance Basis for Spoken Dialogue Processing

浅見　克志　　竹澤　寿幸
Katsushi ASAMI　　Toshiyuki TAKEZAWA
菊井　玄一郎
Genichiro KIKUI

2003年3月17日

音声インタフェースでは，音声認識によって音声から変換されたテキストに基づく情報抽出や加工・変換処理が本質である．本論文では，その基本技術として，1発話毎に話題および発話行為を推定する手法を提案する．提案手法は，まず訓練段階において，話題および発話行為タグが付与された訓練データから，単語と話題および発話行為の関連性を示すスコアを求める．実行段階では，入力された発話に含まれる単語に関して，訓練段階で求めたスコアの和を話題毎および発話行為毎に求め，その大小の順位付けによって，入力発話の話題および発話行為を推定する．ここで用いるスコアは，訓練データ中の単語と話題および発話行為の出現パターンに関する相互情報量と，単語の出現パターンに関するエントロピーから求める．提案手法は，単語と話題および発話行為の1対1の関連性のみを利用し，複数単語の共起などを考慮しないことから，音声対話において多く観察される文長の短い発話にも有効に機能する．旅行会話でよく使われる発話表現を用いた実験の結果，誤認識を含む単語列に対して，話題推定の正解率は約83%，発話行為推定の正解率は約70%であった．

（株）国際電気通信基礎技術研究所
音声言語コミュニケーション研究所
〒619－0288「けいはんな学研都市」光台二丁目2番地2 TEL：0774－95－1301

Advanced Telecommunication Research Institute International
Spoken Language Translation Research Laboratories
2-2-2 Hikaridai “Keihanna Science City” 619-0288,Japan
Telephone:+81-774-95-1301
Fax　　:+81-774-95-1308

もくじ

1 まえがき

音声認識技術の性能向上により，市場に投入され一般消費者が入手可能な機器にも音声インタフェースが搭載されるようになってきている．しかしながら，実際には音声インタフェースがユーザに受け入れられているとは言いがたい状況である．これには，インタフェースの手段としての音声の特性が考慮されていない[1]，音声を用いないほうが所望のタスクを効率的に達成できる場合が多いなど，様々な要因が考えられるが，インタフェースが受理できる言い回しが限定されていることも一因と考えられる[2]．すなわち，ユーザはインタフェースに受理される言い回しを事前に把握しておかなければならない．しかし，人間が記憶できる言い回しには限りがある．一方で，ごく限られた言い回ししか受理できないのであれば，音声インタフェースのメリットがない．音声認識単体では，n-gram などの統計的言語モデルにより，様々な言い回しの発話音声をテキストに変換可能でも，音声インタフェースにおける音声認識の後処理では，その一部のみの解釈や変換がなされているのが技術の現状である．

ここで，「音声認識の後処理」とは，音声認識の出力であるテキスト単語列に対する意味解析処理を指す．意味解析処理としては，係り受け解析など構文解析の結果を利用した処理が考えられるが，音声インタフェースでは構造が単純な短い発話が多く現れると予想されること，様々な言い回しに対応が必要なことから，発話文の構造を極力考慮せず，発話に含まれる単語に基づいて，ユーザ要求を抽出することが望ましい．音声インタフェースの目的は，音声言語により表出されるユーザ発話の，具体的な処理を行なう機器・アプリケーションに対する指令への変換であるので，意味解析処理としてユーザ要求の抽出を行なうことは，合目的的である．

本研究では，音声認識の出力である単語列から，その発話の話題および発話行為を推定する．話題の推定を音声インタフェースに当てはめると，次のように考えられる．音声インタフェースのメリットとして，複数のアプリケーションを，音声指令により自由に切り替えながら利用できることが挙げられる[3]．この場合，ユーザの発話が，どのアプリケーションを利用するための発話なのか，判断する必要がある．話題の推定はこれに相当する．また，発話行為の推定は，ユーザがアプリケーションに何らかの動作を要求している，あるいはアプリケーションの動作に必要な情報を提示しているなど，ユーザの発話意図の推定に相当する．

話題および発話行為の推定は，発話の分類問題と捉えられる．新聞記事などの書き言葉による文書の分類では例えば[4]，ニュース音声など音声言語による文書の分類では例えば[5]～[8]などが報告されている．これらで分類の対象とされる単位は，記事やニュースであり，本研究で対象とする話し言葉における発話と比較して語数の多い，大きな単位となっている．また，コールセンタでの自然言語による Call steering タスクも同様の分類問題である（例えば[9]～[11]）．このタスクの分類対象は，話し言葉における発話であるが，銀行や通信会社などのコールセンタ業務を想定

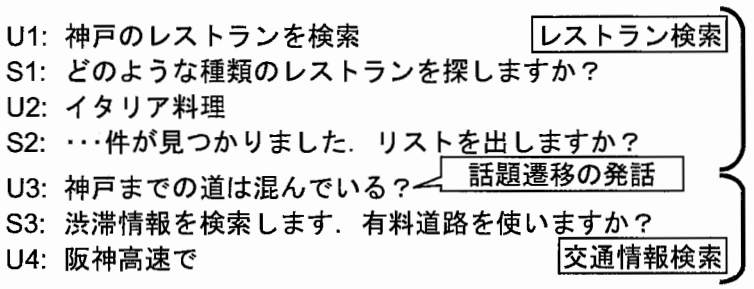

図 1　話題遷移・追従の例

し，ドメインが限定されているため，扱っているデータの語彙サイズは 2000 語程度である．

本研究では，平均 8 語程度の単語から構成される短い発話を対象とし，また語彙サイズが 20000 語程度（推定に利用する実効語彙サイズは約 11000 語）のデータを扱う．提案手法では，訓練フェーズであらかじめ規定された話題および発話行為と単語の関連度（関連性尺度によるスコア）を求めておき，推定実行フェーズでは推定対象発話に含まれる単語に関する関連度の大小関係によって，話題および発話行為を推定する．

2　音声対話処理のための話題および発話行為の推定の要件

音声インタフェースにおける対話処理への応用を前提に話題および発話行為の推定を検討する場合，次に挙げる 5 つの要件を考慮しなければいけない．

言い回しを限定しない．

前述ように，言い回しの限定は音声インタフェースのユーザビリティ向上に対する障壁である．言い回しの限定につながる，単語の共起や語順などへの依存は避ける必要がある．

(1)　幅広い話題を扱う．

前述のように，複数のアプリケーションを同時に扱い，任意のタイミングで任意のアプリケーションを操作できることは，音声インタフェースのメリットの 1 つである．「話題」は，音声インタフェースでは「アプリケーション」に相当すると考えられるので，このメリットを活かすためには，幅広い話題を扱えることが必要である．

(2)　話し言葉特有の，文長の短い発話に対応する．

会話での話し言葉で出現する発話の文長は，一般に短く，さらに音声インタフェースでは要求や情報を簡潔に伝えることが重視されるため，発話長はより短くなると考えられる．このため，単語数の少ない，短い発話から話題などを推定する必要がある．

(3)　1 発話毎に推定を行なう．

音声インタフェースでは，人間同士の対話ではあまり見られない，予期しないタイミングでの話題遷移が発生する可能性がある．例えば，図 1 に示すような対話で利用できるアプリケーションを想定する．図 1 に示す対話例では，アプリケーション側の S2 の問いかけに対し，ユーザはそれに応答せず別のアプリケーションの利用を要求している．このような話題の遷移に追従する

には，U3 の発話だけからアプリケーションの切り替えを要求されていることを検知しなければいけない．すなわち，対話の文脈などを利用せず，個々の発話に関する情報のみを利用して推定する必要がある．

(4) 音声認識誤りに対してロバストである．

音声認識では誤認識は避けられないが，音声インタフェースとしては，発話が完全に正しく認識されなくても，ユーザの指令がアプリケーションに正しく伝達されればよい．すなわち，発話の一部が誤認識されても話題および発話行為は正しく推定できることが望ましい．

本稿では，以上の 5 要件をできるだけ満足するという観点から，話題と発話行為の推定手法を提案する．

3 旅行会話基本表現集(BTEC)

本研究で用いた旅行会話基本表現集(BTEC)[12]について述べる．本研究では，話題および発話行為の推定手法の応用対象として，旅行会話ではなく音声インタフェースを想定している．旅行会話に現れる発話と，音声インタフェースで想定される発話では，内容的に共通する部分は少ない．しかし，今回使用した旅行会話表現集は

- 旅行，特に海外旅行で遭遇が予想される様々な場面（話題に相当）で使われる発話を大量に集めている．
- 複雑な会話で使われる発話ではなく，簡潔な基本表現発話を集めている．
- 1 発話あたりの単語数は 7.7 語と短い． (図 2)
- 個々の発話が独立し，一連の対話を形成していない．

という特徴があり，2 で述べた要件を満たす手法の検討に適している．

この表現集は，一般に書籍として入手可能な旅行会話表現集に見られるような表現の日本語・

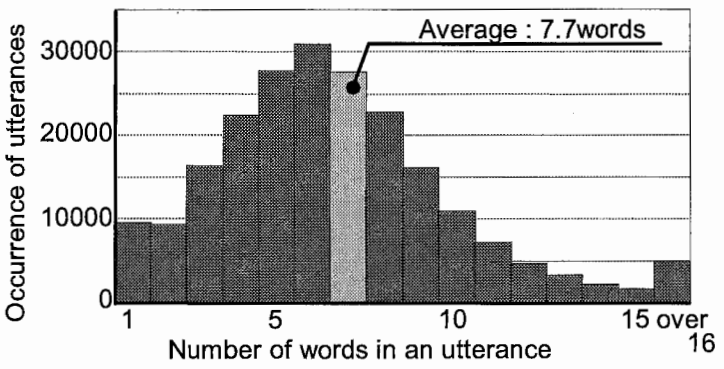

図 2 1発話あたりの単語数の分布

英語の対訳データ集になっており，約 20 万文から成る．本研究では，日本語文のみを使用している．

3.1 話題タグ

この表現集では，個々の発話に話題情報が付与されている．この話題情報は，一般的な旅行会話表現集に見られるような構成に基づいて付与されている．通常，旅行会話表現集は，利便性を考慮して空港，航空機内，出入国審査など場面に分けて構成されているが，場面分類の表記は，表現集により異なる．BTEC では，様々な場面分類体系を基に，Layer 0～2 の 3 階層で 10～252 種類の話題を表すタグを設定した．各発話に対して，3 階層それぞれのタグが付与されている．この例を表 1 に示す．本研究の話題推定では，Layer 0 (10 分類)および Layer 1 (20 分類)について推定する．

また，BTEC には「お支払いは現金ですかクレジットカードですか」（“Restaurant”と“Shopping”）など複数の話題に属し得る発話が含まれる．そのような発話に対して，訓練セットでは，便宜的に 1 回の出現につき，話題を 1 つに決め，各階層の話題タグを付与している．評価セットに関しては，発話毎に可能性のある話題を全て列挙している．

表 1　話題タグの例

Layer 0	Layer 1	Layer 2	Examples
ACTIVITY	beauty	<nil>	自然な色のマニキュアをしてください。
		ask questions regarding cosmetology	スウェーデン式ってどんなマッサージなの。
		receive cosmetic treatment	先のほうだけ切りそろえてください。
	shopping	<nil>	化粧品はどの列に置いてありますか。
		adjust to one's size	この靴の八か二分の一サイズを見せてください。
		trouble	お釣りが違っています。
	sightseeing	<nil>	いちばんの見どころは歴史的な建物です。
		buy a ticket	バルコニー席を二枚ください。
		enjoy sports	リフトの回数券はどこで買えますか。
BUSINESS	business	<nil>	私の名刺です。
		sell	貴社の支払い条件をおうかがいしたい。
		make a contract	この契約のご承認をいただきたいのですが。
	research	<nil>	
		arrange one's research environment	
		do an exchange with other researchers	
COMMUNICATION	basic	<nil>	約束します。
		ask/demand	急いでください。
		greet someone	お目にかかれてうれしく思います。
	communication	<nil>	ここに坐ってもいいですか。
		greet someone	皆様方をお迎えできましたことは私の喜びです。
		introduce one's country	首都は東京です。
CONTACT	contact	<nil>	東京に直通電話をかけたいのです。
		make a claim	違う番号につながれたのですが。
		make a phone call	ピーターさんをお願いできますか。

Layer 0	Layer 1	Layer 2	Examples
EATandDRINK	a snack	<nil>	バニラのアイスクリームをひとつお願いします。
		get fast food	ハンバーガーをひとつとコカコーラの大をひとつください。
		order	アメリカンモーニングセットにします。
	drink	<nil>	予約はしていませんが席はありますか。
		order	ウィスキーをシングルのオンザロックでください。
		pay	キャッシュオンデリバリーでいいの。
	restaurant	<nil>	そうですね。何か前菜がいいですね。
		ask a waiter/waitress a question	今日のお勧め料理は何ですか。
		enjoy one's meal	低カロリーのデザートはありませんか。
EXCHANGE	exchange	<nil>	両替はどこでできますか。
		exchange	この国の通貨に替えてください。
		look for the exchange counter	すみません。最寄りの銀行を教えてくださいませんか。
STAY	stay	<nil>	ホテルを紹介してください。
		check-in	チェックインをお願いします。
		make a claim	部屋の冷房が効きません。
STUDY	homestay	<nil>	起きなさい。
		apologize	謝らなくてはいけないことがあるんですが。
		become ill	あなた顔色が悪いわ。
	study overseas	<nil>	英語を勉強したいのですが。
		go to school	これが入学証明書です。
		live in a dorm	寮に住んでいるのですか。
TRANSPORT	airplane	<nil>	お冷やがほしいのですが。
		ask a flight attendant a question	いつサンフランシスコに着くのですか。
		prepare for landing	入国カードをもらえますか。
	airport	<nil>	パスポートを見せてください。
		carry one's baggage	タクシー乗り場まで荷物を運んでください。
		go through boarding procedures	ノースウエストのカウンターはどこですか。
	go home	<nil>	もしもし予約の再確認をしたいのですが。
		other	出発の二時間前には空港に来てください。
		reconfirm one's reservation	予約の再確認をしたいのですが。
TRANSPORT	move	<nil>	タクシー乗り場はどこですか。
		use a rental car	車を借りたいのですが。
		make a claim	メーターより多く請求しています。
TROUBLE	trouble	<nil>	財布をなくしました。
		ask directions	道に迷ってしまいました。
		buy some medicine	この近くに薬局がありますか。
10分類	20分類	252分類	

3.2 発話行為タグ

本研究で用いる発話行為タグは，音声翻訳研究国際コンソーシアムC-STAR（Consortium for Speech Translation Advanced Research）が，意味の表現形式として定めた中間言語[13][14] (注 1) に基づいている．この中間言語では，ほぼ単文相当の「意味単位」を発話行為，概念，引数で表現する．1つの発話は，1個以上の意味単位で構成される．発話行為タグに関しては60個以上が定義されて

（注 1）：この中間言語定義は，現在は欧米の研究機関・大学で構成されるNESPOLE!（Negotiating through Spoken Language in e-commerce）プロジェクトの中間言語IF（Interchange Format）に引き継がれている．

表 2 発話行為タグの例

Speech act tag	Example
give-information	１月３日から２泊です
request-information	電話番号をお願いします
request-action	予約をお願いします
verify-give-information	２泊でございますね
closing	それでは失礼いたします
accept	はい、それで結構です
affirm	はい、その通りです
reject	いいえ、要りません
negate	いいえ、違います

いるが，BTEC では，このうちの 25 種類が出現する．この例を表 2 に示す．発話行為タグ付きのデータは，表現集に含まれる各発話を作業者が読み，適切なタグを判断し付与することによって作成される．従って，作業量の面での制約が大きくなる．このため，発話行為タグは，BTEC の 10075 発話にのみ付与されている (注 2)．

3.3 話題および発話行為タグの分布に関する特徴

発話行為タグを複数の話者による実際の対話データ（例えばホテルのフロント係と客の模擬対話など）に付与すると[15]，発話行為の出現分布は多岐の発話行為にわたる．しかし，BTEC は独立した発話を収集したデータであるため，対話の進行によって現れると考えられる確認表現や受諾表現はほとんど出現せず，情報提示表現，情報要求表現，行為要求表現に集中する．図 3 に，ホテルのフロント係と客の模擬対話である音声言語データベース(SLDB)と，BTEC での発話行為タグの出現分布を示す．データ中の各発話行為タグの出現割合は，give-information, request-information, request-action で 85%程度に達し，他の発話行為タグは 1～2%以下である．このため，本研究の BTEC での発話行為推定では，give-information, request-information, request-action と「その他」タグを想定して，4 種類の発話行為タグの推定を行なう．

4 話題および発話行為推定手法

本稿で述べる手法では，入力は形態素情報が付与された単語系列，出力は話題タグ（Layer 0, Layer 1）および発話行為タグである．本章では，この推定手法を訓練フェーズと実行フェーズに分けて述べる．

(i) 訓練フェーズ

訓練データに含まれる全ての単語と話題タグおよび発話行為タグの組み合せの関連度を求める．

（注 2）発話行為タグ付与の指針については，ＴＲ－ＳＬＴ－００３４を参照．

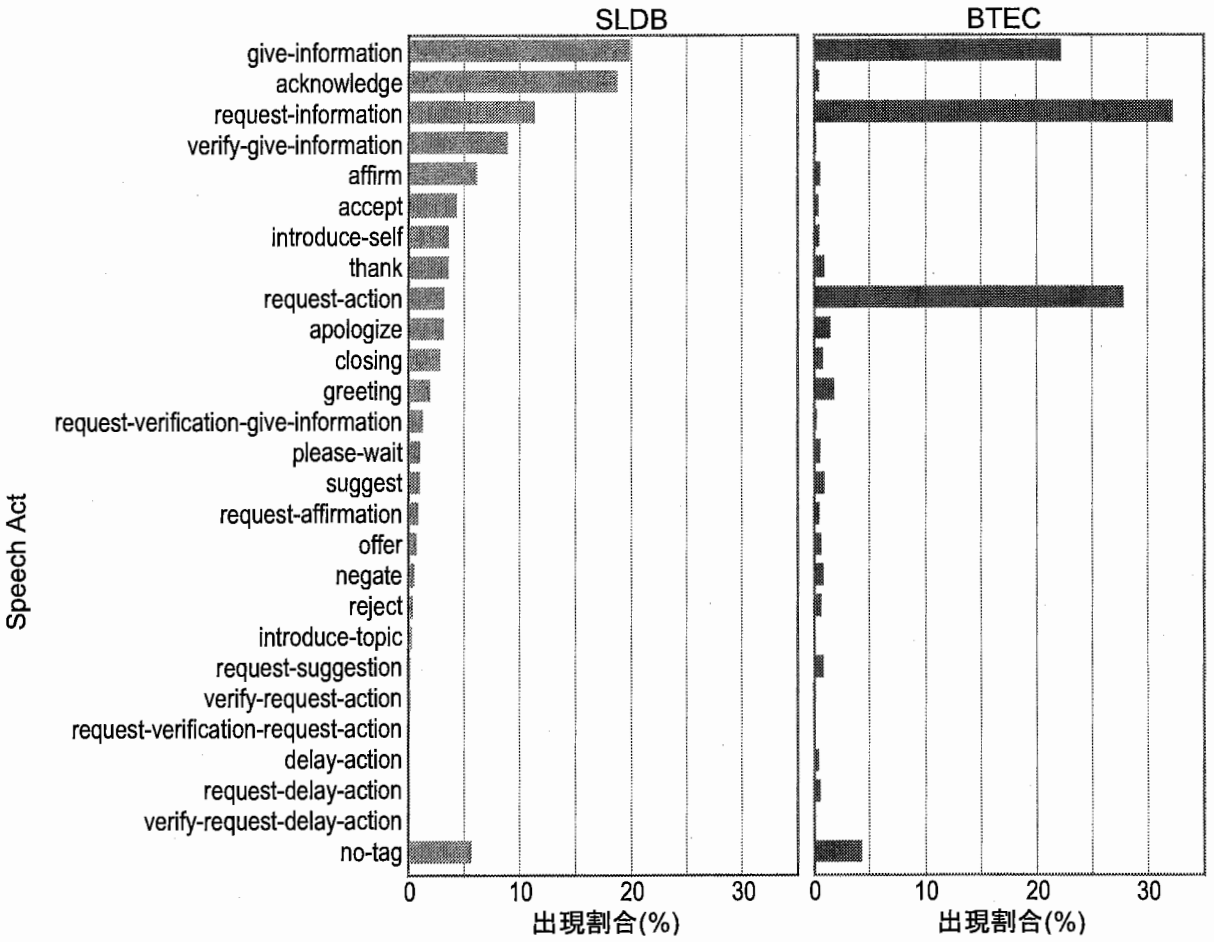

図 3　発話行為タグの出現分布

(ii) 実行フェーズ

訓練フェーズで求めた関連度を利用して，入力発話の話題および発話行為を推定する．

4.1 訓練フェーズ ー 関連度の計算

関連度とは，単語と話題および発話行為の関連の強さを示す数値である．本研究で扱うような関連性尺度としては，相互情報量が用いられることが多い[6][7]．すなわち，特定の話題または発話行為に依存して出現する単語は，その話題または発話行為との相互情報量が大きくなる性質を利用する．話題または発話行為 t_k と単語 w_i の相互情報量 $i(t_k; w_i)$は，式(1)で表される．

$$\begin{aligned} i(t_k;w_i) &= i(t_k) - i(t_k|w_i) \\ &= -\log_2 p(t_k) - \{-\log_2 p(t_k|w_i)\} \\ &= -\log_2 p(t_k) - \left\{-\log_2 \frac{p(t_k,w_i)}{p(w_i)}\right\} \end{aligned} \tag{1}$$

ここで，$p(t_k), p(w_i)$は，それぞれ話題または発話行為 t_k と単語 w_i の出現確率，$p(t_k|w_i)$は w_i が与えられたときの t_k の条件付確率である．しかし，本研究で対象としているのは，語数の少ない短い

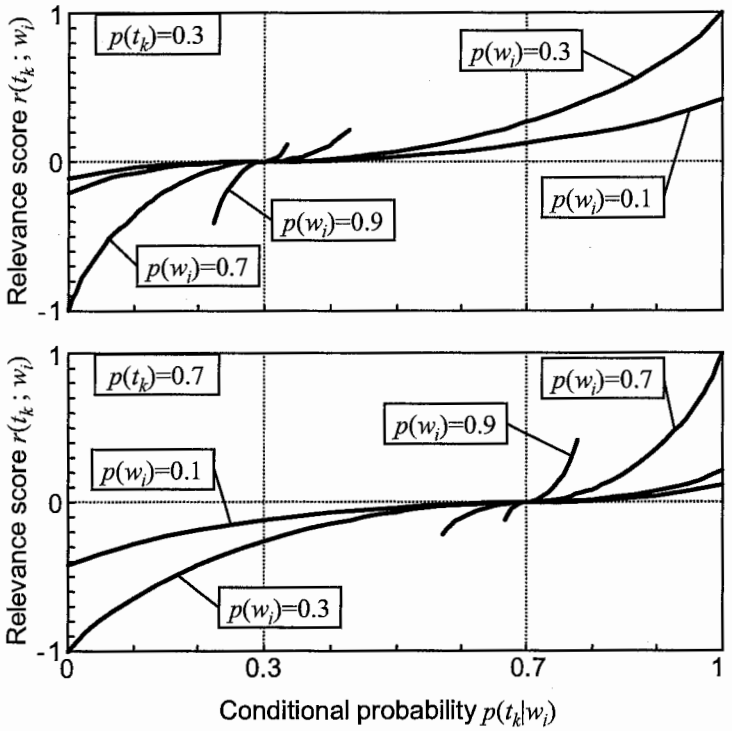

図 4　単語と話題または発話行為の出現確率と関連度

発話であるため，特定の話題や発話行為と関連の大きな単語であっても，その話題や発話行為との共起確率が，話題の出現確率に比べ，あまり大きくならない．結果として，式(1)で表される相互情報量における，出現した単語の重要度による差は小さくなる．すなわち，この相互情報量には，話題または発話行為 t_k の情報量 $i(t_k)$の影響が強く残り，単語と話題および発話行為の関連性を適切に表せない．

また，相互情報量が，単語の出現頻度を考慮していない点を補うために，ある話題における単語の重要度を示すとされる TF-IDF(Term Frequency, Inverse Document Frequency)を併用することが提案されている[7]．ここで，TF-IDF は式(2)のように適用される．

$$\mathrm{TF\text{-}IDF} = \mathrm{TF}(w_i, t_k) \cdot \mathrm{IDF}(w_i) \tag{2}$$

$$\mathrm{TF}(w_i, t_k) = \text{単語}w_i\text{が話題}t_k\text{の発話に出現する回数}$$

$$\mathrm{IDF}(w_i) = \log \frac{\text{訓練セット中の全発話数}}{\text{単語}w_i\text{が出現する発話数}}$$

しかし TF-IDF も，対象が短い発話であるため，有効ではない（詳しくは 5.2 節）．

式(1)で表される相互情報量も TF-IDF も，単語と話題や発話行為が共起する事象のみを捉えていると考えられるが，本研究では，共起以外の事象も含む，単語と話題，発話行為の出現パターンの一致度に基づく関連性尺度を導入する．対象とする発話が短いため，１発話内に同一の単語が複数回出現することは考え難く，仮にある話題と完全に関連している単語が存在すれば，その出

現パターンは話題の出現パターンと一致すると考えられる．

ここで「単語 w_i を含む発話が出現する／しない」という情報源 W と「話題または発話行為 t_k の発話が出現する／しない」という情報源 T を考える．この２個の情報源の関連性を単語と話題および発話行為の関連度と考える．

ここで，情報源 T，W の出力パターンが一致している場合，相互情報量 $I(T;W)$ と単語のエントロピー $H(W)$ が等しくなる．そこで，話題または発話行為 t_k と単語 w_i の関連度 $r(t_k;w_i)$ を次のように定義する．

$$r(t_k;w_i)=\begin{cases}\dfrac{I(T;W)}{H(W)} & \left(p(t_k|w_i)\geq p(t_k)\right)\\ -\dfrac{I(T;W)}{H(W)} & \left(p(t_k|w_i)< p(t_k)\right)\end{cases} \tag{3}$$

ただし

$$H(W)=-\sum_{w_i,\overline{w}_i}p(w)\log_2 p(w) \tag{4}$$

$$I(T;W)=\sum_{t_k,\overline{t}_k}\sum_{w,\overline{w}_i}p(t,w)\log_2\frac{p(t|w)}{p(t)} \tag{5}$$

式(3)は，単語 w_i から見た話題または発話行為 t_k との関連度を示し，$r(t_k;w_i)$ の値の範囲は[−1,1]である．図４に単語，話題または発話行為の出現確率と関連度の関係を示す．$p(t_k|w_i)=p(t_k)$ すなわち T，W が独立ならば，T，W に関連性は無く，$r(t_k;w_i)=0$ となる．一方，$p(t_k)=p(w_i)$ かつ $p(t_k|w_i)=1$ のとき，T，W は完全に出力パターンが一致しており，$r(t_k;w_i)=1$ となる．$p(t_k|w_i)=1$ であっても，$p(t_k)\neq p(w_i)$ ならば，$r(t_k;w_i)<1$ となる．また，$r(t_k;w_i)=-1$ となるのは，T，W の出力パターンが完全に反転している場合である．$p(t_k|w_i)<p(t_k)$ の領域での $r(t_k;w_i)$ は「単語 w_i が出現した場合，話題または発話行為が t_k ではない」関連度を表している．

式(3)のような，相互情報量のエントロピーによる正規化を，２つの情報源の情報エントロピー的観点による相関係数とする考え方もある[16]．

また，相互情報量 $I(T;W)$ のみでも，T，W の関連性をある程度表していると考えられる．しかし，$I(T;W)$ は「W の出力を知った場合に T の曖昧さがどの程度減少するか」のみを表しており，図５に示すように，関連性が小さくても相互情報量が大きくなる可能性があり，関連性尺度としては不十分である．

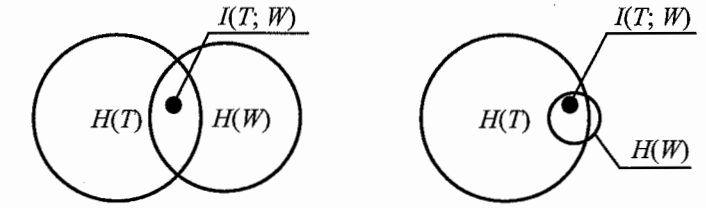

図 5 相互情報量と関連度の関係

なお，話題推定については，話題の階層構造に依存して，上位階層を全体空間として求める．つまり，Layer 0 では訓練データの全発話が全体空間となるが，Layer 1 では上位階層の Layer 0 で同じ話題タグが付与された発話の集合が全体空間となる．例えば、Layer 1 の話題タグ“Airport”に関する関連度を求める場合は，対応する Layer 0 の話題タグとして“TRANSPORT”が付与された発話の集合を全体空間とする（表 1 参照）．

4.2 実行フェーズ ― 話題および発話行為の推定

推定の基本的な考え方は，各話題および発話行為に対する入力発話のスコアを求め，最大スコアの話題および発話行為を選択する，あるいはスコアに応じそれらを順位付けすることである．各話題および発話行為に対する入力発話のスコアは，入力発話に含まれる単語の，各話題および発話行為との関連度の総和として求められる．

話題推定の実行フェーズは，式(6)から式(10)に示す，簡単な行列演算で表される．

Z を入力発話 S の特徴を示すベクトルとする．Z は，S が単語 w_i を含む／含まないを 1/0 で表す．

$$
\begin{aligned}
&Z = \left[\mu_Z(w_1)\ \mu_Z(w_2)\ \cdots\ \mu_Z(w_l)\right] \\
&\mu_Z(w_i) = \begin{cases} 1 & : S \text{ includes a word } w_i \\ 0 & : S \text{ does not include a word } w_i \end{cases} \\
&(i = 1, 2, \cdots, l)
\end{aligned}
\tag{6}
$$

次に，単語と話題との関連度を要素とする行列 R_T，発話行為との関連度を要素とする行列 R_S を考える．

$$
R_T = \begin{bmatrix} r(t_1; w_1) & \cdots & r(t_m; w_1) \\ \vdots & \ddots & \vdots \\ r(t_1; w_l) & \cdots & r(t_m; w_l) \end{bmatrix}
\tag{7}
$$

$$
R_S = \begin{bmatrix} r(s_1; w_1) & \cdots & r(s_n; w_1) \\ \vdots & \ddots & \vdots \\ r(s_1; w_l) & \cdots & r(s_n; w_l) \end{bmatrix}
\tag{8}
$$

発話－話題関連度ベクトル A_T，発話－発話行為関連度ベクトル A_S は Z と R_T，R_S の積で求められる．

$$\begin{aligned} A_T &= Z \cdot R_T \\ &= \left[\mu_Z(w_1) \cdots \mu_Z(w_l)\right] \begin{bmatrix} r(t_1;w_1) & \cdots & r(t_m;w_1) \\ \vdots & \ddots & \vdots \\ r(t_1;w_l) & \cdots & r(t_m;w_l) \end{bmatrix} \\ &= \left[r_{A_T}(t_1) \cdots r_{A_T}(t_m)\right] \end{aligned} \tag{9}$$

$$\begin{aligned} A_S &= Z \cdot R_S \\ &= \left[\mu_Z(w_1) \cdots \mu_Z(w_l)\right] \begin{bmatrix} r(s_1;w_1) & \cdots & r(s_n;w_1) \\ \vdots & \ddots & \vdots \\ r(s_1;w_l) & \cdots & r(s_n;w_l) \end{bmatrix} \\ &= \left[r_{A_S}(s_1) \cdots r_{A_S}(s_n)\right] \end{aligned} \tag{10}$$

1best の推定結果を求める場合は，次式から求められる．

$$\hat{t} = \underset{t_p}{\operatorname{argmax}}\ r_{A_T}(t_p) \qquad (p = 1, 2, \cdots, m) \tag{11}$$

$$\hat{s} = \underset{s_q}{\operatorname{argmax}}\ r_{A_S}(s_q) \qquad (q = 1, 2, \cdots, n) \tag{12}$$

ここで，m=30（Layer 0, Layer 1 の話題分類数の和），n=4（give-information, request-information, request-action, その他）である．ただし，話題推定では階層構造に依存するので，Layer 1 の話題に対応する発話－話題関連度の順位付けは，Layer 0 の発話－話題関連度の順位付けに依存する．例えば，Layer 0 の推定結果の第 1 位が“ACTIVITY”，第 2 位が“TRANSPORT”であるとする．このとき，Layer 1 では，関連度の計算結果の大小関係で第 1 位が“airplane”（Layer0 の“TRANSPORT”に従属），第 2, 3 位がそれぞれ“sightseeing”, “shopping”（いずれも Layer0 の“ACTIVITY”に従属）の場合，推定結果の第 1 位は，Layer 0 の推定結果に依存して，“sightseeing”となる．

4.3 音声認識誤りの影響の低減 － 同音語のマージ

単語は“表記形”，“読み”，“正規形”，“品詞”，“活用形”などの属性を有する．単語を識別する際に考慮するこれらの属性の選択が，音声認識誤りに対する話題推定のロバスト性に影響を与える．特に，“表記形”は意味情報まで含むので，同音異義語の音声認識誤りの影響に関係する．一方で，“表記形”の異なりを無視し，同音語をマージすることによって，若干の情報を失いつつも，音声認識誤りの影響を低減できると考えられる．この属性の組み合せと単語識別例を表 3 に示す．

表3 単語属性と単語識別例

	属性			
	表記形	読み	基本形	品詞
A	乗り換え	ノリカエ	乗り換える	動詞
B			乗り換え	名詞
C	乗換		乗換	
D	席	セキ	席	名詞
E	咳		咳	

“表記形”のみを考慮する場合
→{A, B}, {C}, {D}, {E}
“読み”のみを考慮する場合
→{A, B, C}, {D, E}
{ }内は同一単語として扱う.

4.4 サポートベクタマシンによる話題推定

ここまで述べてきた提案手法では，単語と話題あるいは発話行為の関連性を示す値を，極めて直接的に演算することにより推定を行なっている．これに対し，文書分類問題においては，より優れた分類アルゴリズムに関する議論が盛んであり，ブースティングやサポートベクタマシン(SVM)が現在有力とされている[17]．本稿では，SVM による話題推定と比較する．ここでは，今回の話題推定に対する SVM の適用の概要について述べる．

SVM の適用にあたって，まず発話文の特徴を示す属性として名詞類，動詞類を選択する．発話文ベクトルは，これらの属性を含む(0)／含まない(1)により表現する場合，例えば次のように表現される．

$$\mathbf{x}_i = (0, 1, 0, 0, 1, 1, 1, 0, \cdots)$$

入力文が分類対象の話題に「属する／属さない」は+1/–1 と表現して訓練する．

また，発話文ベクトルを，式(3)で求められる分類対象の話題と単語との関連度で表現する場合は，次のように表現される．

$$\mathbf{x}_i = (-0.000116, 0.001437, 0, -0.002639, 0.151007, 0.074635, 0.017724, \cdots)$$

また，SVM のカーネル関数には以下の多項式関数を使用した[4].

$$K_{poly} = (\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^d \qquad (d = 1) \tag{13}$$

5 評価実験・検討

表 4 に示す実験条件のもと，提案手法の性能を評価する実験を行なった.

最初に，評価セット(I), (II)の違いについて述べておく．3.1 で述べたように，BTEC では，１発話の１回の出現について１個の話題および発話行為が与えられている．評価セット(I)に対しては，これとは別に，１人の作業者が判断した，各発話で可能性のある複数の話題も与えられている．こ

表 4 話題推定の実験条件

	訓練セット	評価セット(I)	評価セット(II)
発話数	162320	1521	633
単語識別のための属性組み合せと語彙サイズ			
組み合せ			語彙サイズ
(a)	表記形＋読み＋基本形＋品詞		21436
(b)	表記形＋品詞		21050
(c)	読み＋品詞		18447
(d)	読み		17477
音声認識の単語誤り率（WER）(%)			8.0

表 5 発話行為推定の実験条件

	語彙サイズ	意味単位数	
		訓練セット	評価セット
BTEC	3588	8409	1666
SLDB	2271	14074	---------
BTEC+SLDB	5069	22483	---------

の作業者は英日方向版の同様の表現集 (注 3) 作成に従事しており，話題分類および発話と話題の関係について熟知している．

評価セット(II)は，評価セット(I)から，作業者による判断の結果，話題が 1 個に限定され，かつ，その話題が元来付与されていた話題と一致する発話のみを抽出したセットである．話題の曖昧性が小さいセットと言える．

以下に述べる話題推定実験では，まず，単語識別のための属性組み合せの影響について検討する．次に，関連性尺度として本研究で提案する関連度と，相互情報量[6][7][8]，TF-IDF と相互情報量の積[7]を用いた場合の推定結果を比較する．

従来，相互情報量などを用いる場合は，品詞フィルタによって内容語のみを使用することが多いが，このフィルタの有無についても比較する．実験で使用する品詞フィルタは名詞，動詞，形容詞，副詞のみを抽出する．

発話行為については，話題推定実験で良好な結果が得られた条件，すなわち式(3)で求められる単語の発話行為の関連度を用い，単語属性組み合せは表 4 の(c)「読み＋品詞」のもとで，推定実験を行なった．関連度の計算には，BTEC, SLDB, および両者の混合データを用いた．表 5 に発話行為推定の実験条件を示す．BTEC の評価セットは，話題推定の評価セット(II)と同じ発話から構成されるが，3.2 に述べたように，1 つの発話は 1 個以上の意味単位から構成されるため，発話数

（注 3）：日本を訪れる外国人旅行者向けの旅行会話表現集に出現するような基本表現を集めたコーパス．

と意味単位数は一致しない．また，SLDB，および両者の混合データによる場合は，クローズドテストのみ実施した．発話行為推定では，利用できるデータの規模が話題推定に比べ小さいため，語彙サイズの小さくなっている．BTEC，SLDB の重なり語彙は 790 であった．なお，3.3 に述べたように，BTEC での発話行為推定は 4 種類のタグを推定するが，SLDB を推定対象にする場合は，25 種類のタグを推定する．

評価は次に示す推定精度を基準とした．

$$\text{推定精度} = \frac{\left(\begin{array}{l}\text{評価セットに付与されている}\\ \text{話題または発話行為と}\\ \text{推定結果が一致した発話数}\end{array}\right)}{\text{評価セット内の全発話数}} \times 100$$

ここで，推定結果は 1best の話題および発話行為である（式(11), (12)参照）．ただし，提案関連度による話題推定に関してのみ，3best までの推定結果も示す．また，評価セット(I)を用いて評価する場合，各発話に対し列挙された複数の話題に，推定の結果得られた話題と一致するものが含まれていれば，「評価セットに付与されている話題と推定結果が一致した発話」としてカウントする．

SVM を除く，いずれの実験においても，訓練データ中に一度しか観測されなかった単語は，推定に使用していない．

また，各表中および各図中の *Trans.*は，評価セットの書き起しテキストから推定した結果，*Recog.*は，音声認識の出力から推定した結果を示す．

5.1 単語属性組み合せに関する比較

表 6 に提案関連度を用いた話題推定における，単語属性選択の影響を示す．この表に示される実験結果は評価セット(II)を用いたものである．表中の劣化率は，評価セットの書き起しテキスト（*Trans.*, WER=0%）を入力として話題を推定した場合に対して，誤りを含む音声認識結果（*Recog.*, WER=8.0%）を入力とした場合に，推定精度が劣化した割合である．劣化率は，単語情報をより詳細に扱う単語属性組み合せほど，大きくなる．この結果から，単語属性の表記形に含まれる意味的な情報を無視することにより，若干の情報を失いつつも，話題推定に対する音声認識誤りの影響が低減されていると考えられる．また，4.2 節で述べた推定手法では，個々の単語と話題および発話行為との関連度を利用するため，話題および発話行為推定に対する影響の小さい単語，すなわち話題や発話行為に関わらず，どのような発話にも出現しうる単語が音声認識誤りによって失われても，話題および発話行為推定としての問題は小さいと考えられる．表 6 では，単語を最も詳細に扱う属性組み合せ(a)でも，劣化率は音声認識の WER に比べ小さいことから，単語属性の組み合せだけでなく，推定方法自体にも，音声認識誤りの影響を低減する効果があると考えられる．

表 6 単語属性組み合せと話題推定精度

話題階層	Layer 0			Layer 1		
	Trans.	*Recog.*	劣化率	*Trans.*	*Recog.*	劣化率
組み合せ(a)	85.8	80.7	5.9%	81.5	76.5	6.2%
組み合せ(b)	85.6	81.5	4.8%	81.3	77.2	5.1%
組み合せ(c)	85.8	83.3	2.9%	81.3	78.4	3.6%
組み合せ(d)	85.0	82.8	2.6%	80.3	77.9	3.0%

表 7 関連性尺度と話題推定精度

話題階層	Layer 0		Layer 1	
	Trans.	*Recog.*	*Trans.*	*Recog.*
提案関連度	85.8	83.3	81.3	78.4
相互情報量	45.0	43.4	33.9	31.1
TF-IDF･相互情報量	50.6	49.1	23.4	23.0

単語属性組み合せ(c)による.

5.2 関連性尺度の比較

表 7 に関連性尺度に関して推定精度を比較した結果を示す．表 7 から，相互情報量や TF-IDF と相互情報量の積に比べ, 提案関連度により推定する場合の精度が格段に良好であることがわかる．相互情報量による場合の推定誤りを分析した結果，訓練セットで最も出現頻度が小さい話題を推定結果として出力している場合が多かった．すなわち，式(1)の第 1 項の影響が大きく，推定誤りが発生していると考えられる．

一方，TF-IDF と相互情報量の積による場合の推定誤りを分析すると，訓練セットで最も発話数が多かった話題を推定結果として出力している場合が多かった．これは，本研究での推定対象が「短い発話」であることに由来すると考えられる．TF-IDF は，出現頻度が高く，かつ，特定の発話に偏って現れるほど値が大きくなる性質を持つ．しかし，本研究で扱う短い発話では，ある単語の出現頻度と，その単語を含む発話数は，ほぼ一致する．このため，TF-IDF 値は単語の出現頻度の増加と共に単調増加し，また発話数の多い話題ほど単語の出現頻度が大きくなるので，推定結果が発話数の多い話題に誘導されたと考えられる．

これらに対し，提案関連度は話題や単語の出現頻度に関係なく，話題と単語の出現パターンの一致度合いを示す数値であるため，より適切に話題と単語の関連性を表示し，良好な推定結果が得られていると考えられる．

5.3 品詞に基づく推定に使用する単語選択の影響

表 8 に品詞フィルタの有無に関する推定精度の比較結果を示す．品詞による内容語の選択と組み合わせて使用されることが多い相互情報量などでは，品詞フィルタの効果は非常に大きい．し

表 8 品詞フィルタの有無と話題推定精度

フィルタの有無→	Layer 0			Layer 1		
	あり	なし	改善率	あり	なし	改善率
提案関連度	81.7	83.3	−1.9%	77.0	78.4	−1.8%
相互情報量	48.4	43.4	11.6%	41.7	31.1	33.9%
TF-IDF･相互情報量	67.3	49.1	37.0%	46.9	23.0	103.8%

単語属性組み合せ(c)，音声認識の結果からの推定による．

かし，フィルタを使う場合でも，その推定精度は提案関連度による場合に比べ，低い．

一方，提案関連度による推定では，品詞フィルタは，わずかに逆効果になっている．品詞フィルタにより除去される機能語類は，相対的に高頻度で出現すると考えられるが，それでも一般的には出現率が 0.5 を超えることはない．従って，式(4)で $H(W)$は単調増加となるため，高頻度語に対する関連度 $r(t_k;w_i)$の絶対値は小さくなる．このように，提案関連度には，高頻度語の影響を除去する効果が含まれていると考えられる．さらに，「品詞」という定性的な基準による単語選別により，提案関連度の定量的な基準で有効に機能していた情報の一部が除去されるため，逆効果が現れていると考えられる．

図 6〜図 13 に 5.1〜5.3 に示した実験の，すべての条件の組み合せによる実験結果を示す．表 6〜表 8 では，代表的な条件による結果を抜粋して示してあるが，図 6〜図 13 に示す実験結果おいても，これまでに述べたような傾向が現れている．ただし，絶対値としては，条件により大きな変動が見られる．相互情報量および TF-IDF･相互情報量による推定精度は，評価セット(I)を用いる場合のほうが，評価セット(II)を用いる場合を大幅に上回っている．これは，評価セット(I)が複数の話題に分類される発話を多く含んでおり，その複数の話題に推定結果と一致する話題が含まれていれば，推定結果が正解と評価するため，評価セット(II)に比べ，推定精度の評価が緩やかになっているためと考えられる．一方，提案関連度による推定精度は，評価セット(I)による場合のほうが，むしろ若干低いが，その変動幅小さい．どのような条件でも，提案関連度による推定精度の変動幅は小さく，安定している．

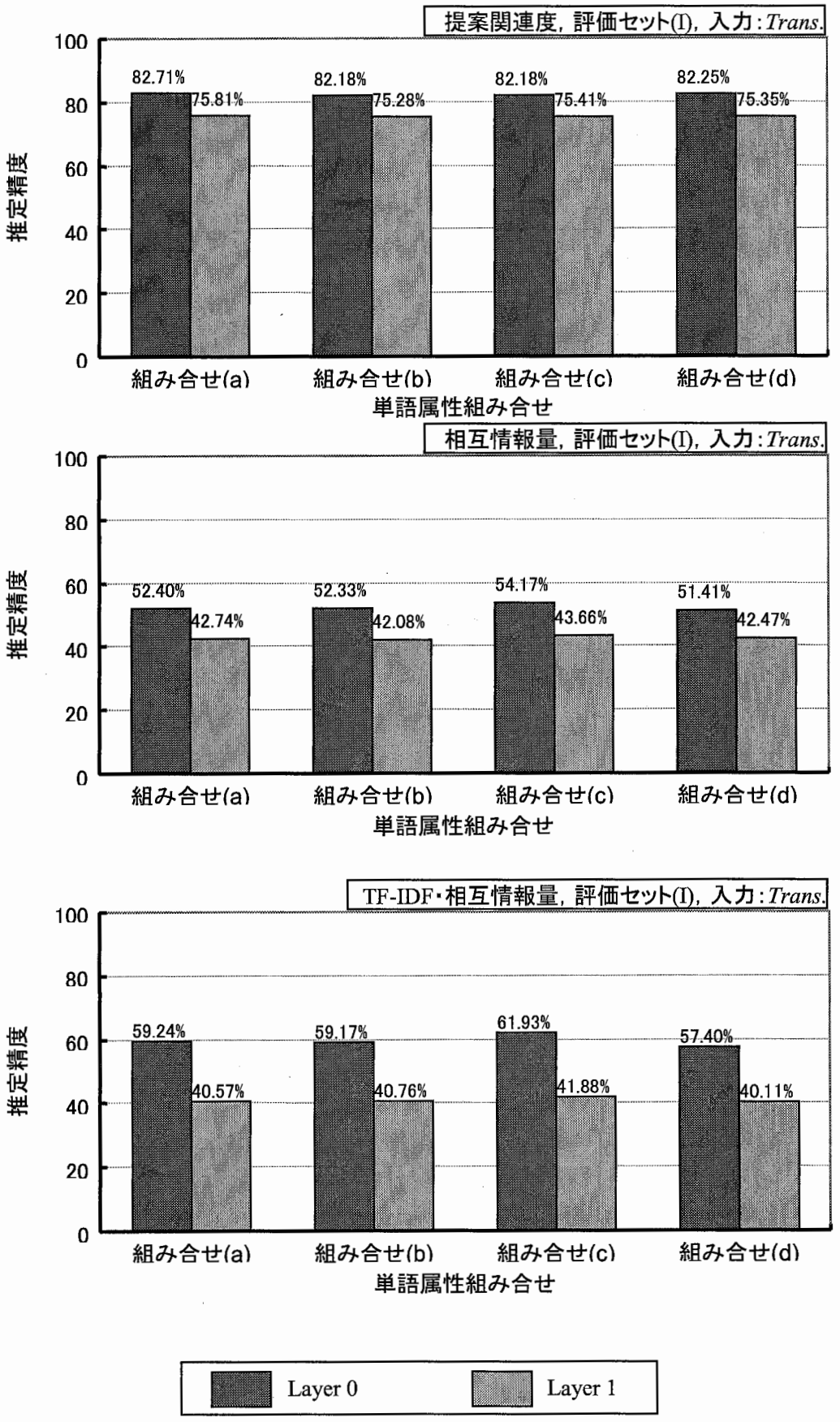

図 6 {評価セット(I), 書き起し入力, 品詞フィルタなし}の条件下での推定精度

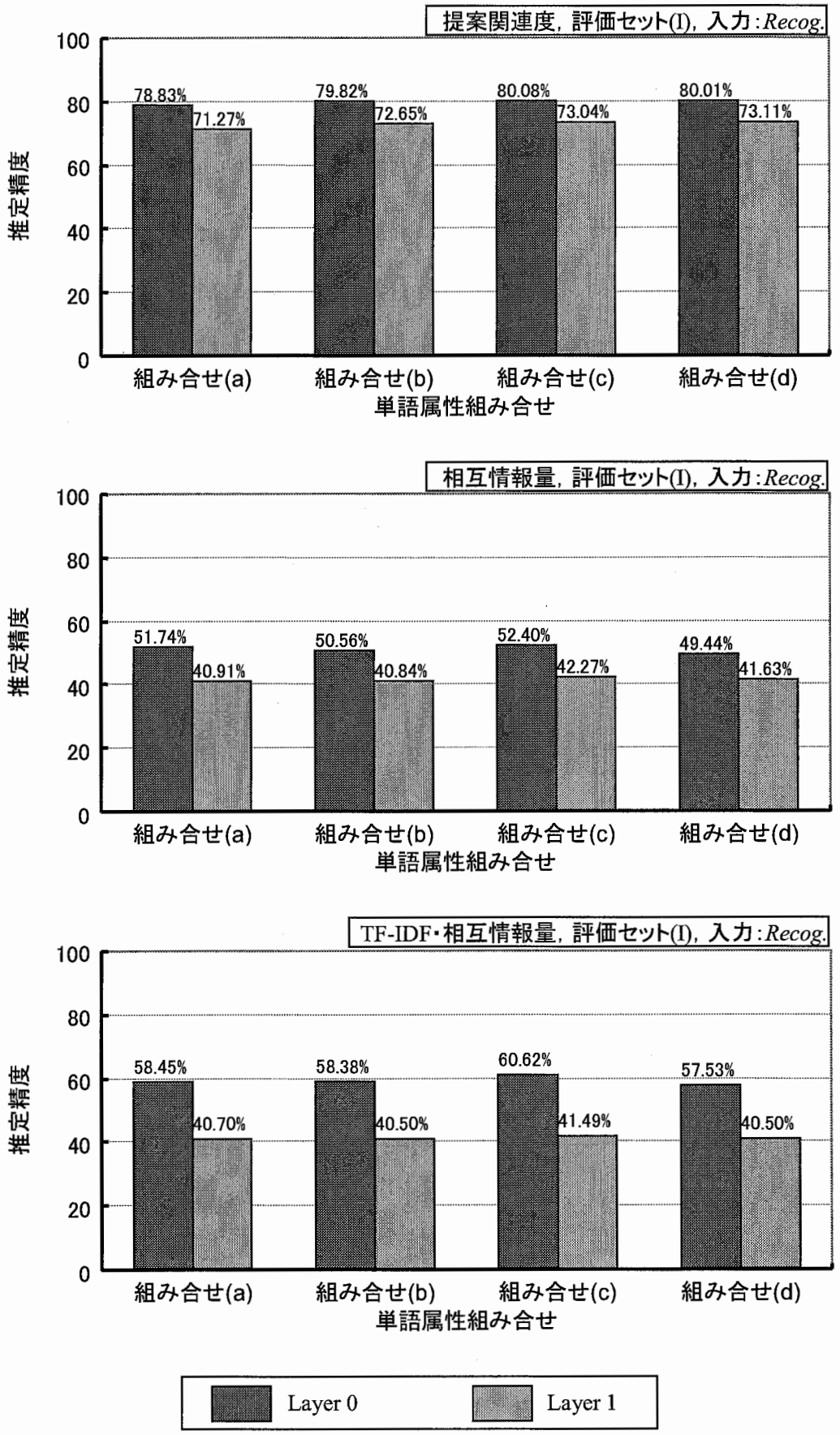

図 7 {評価セット(I)，音声認識結果を入力，品詞フィルタなし}の条件下での推定精度

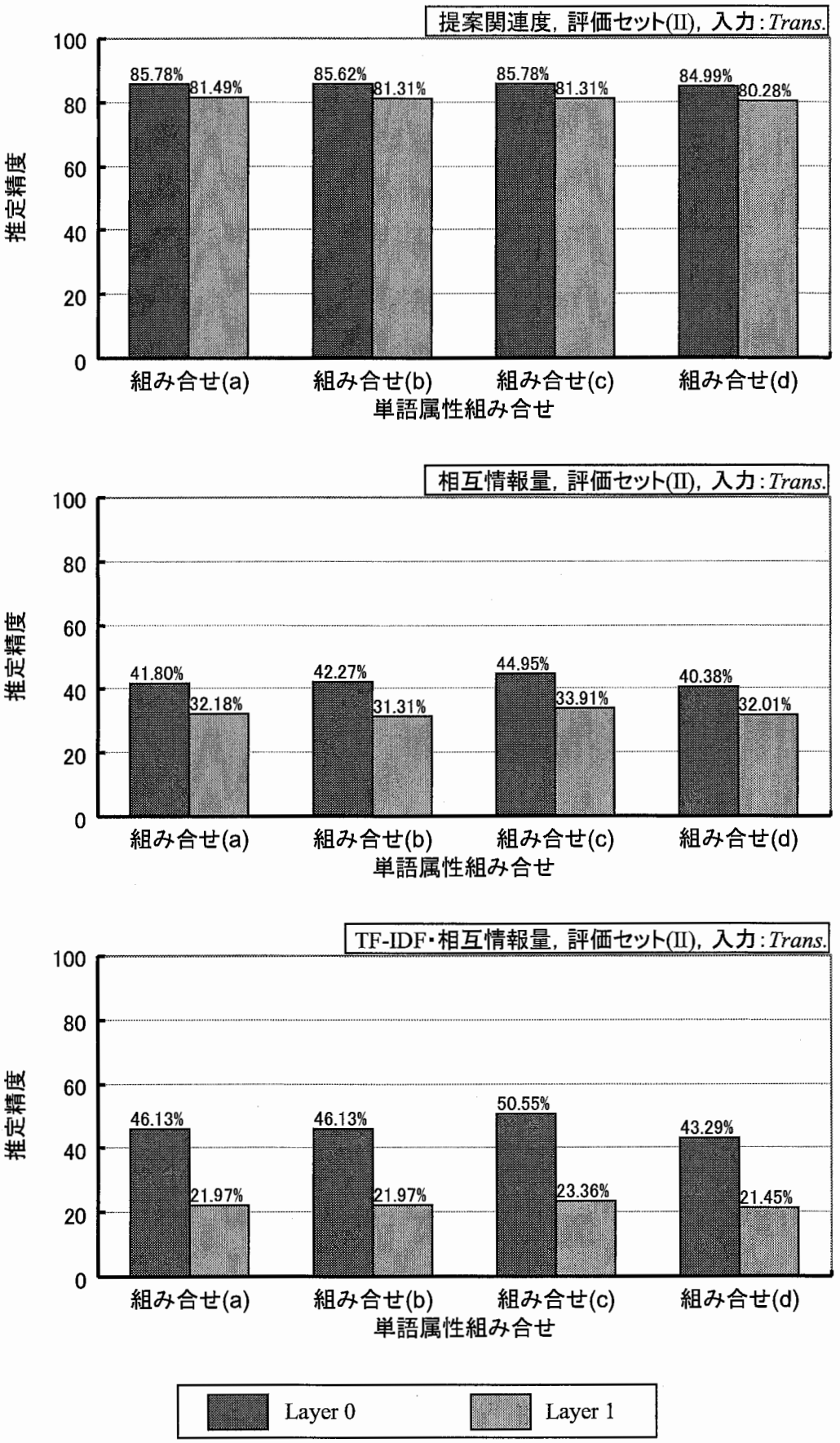

図 8 {評価セット(II)，書き起し入力，品詞フィルタなし}の条件下での推定精度

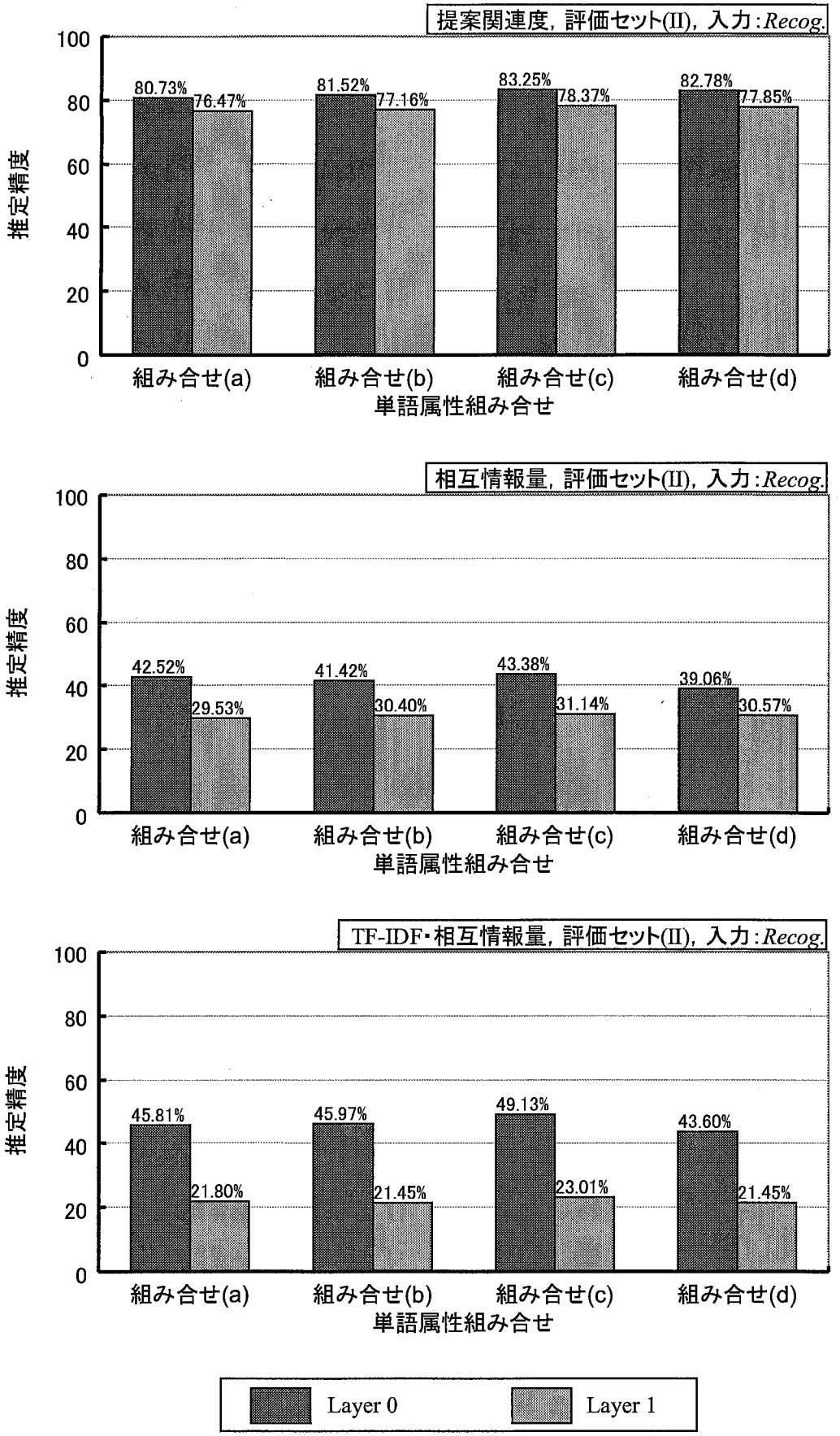

図 9 {評価セット(II)，音声認識結果を入力，品詞フィルタなし}の条件下での推定精度

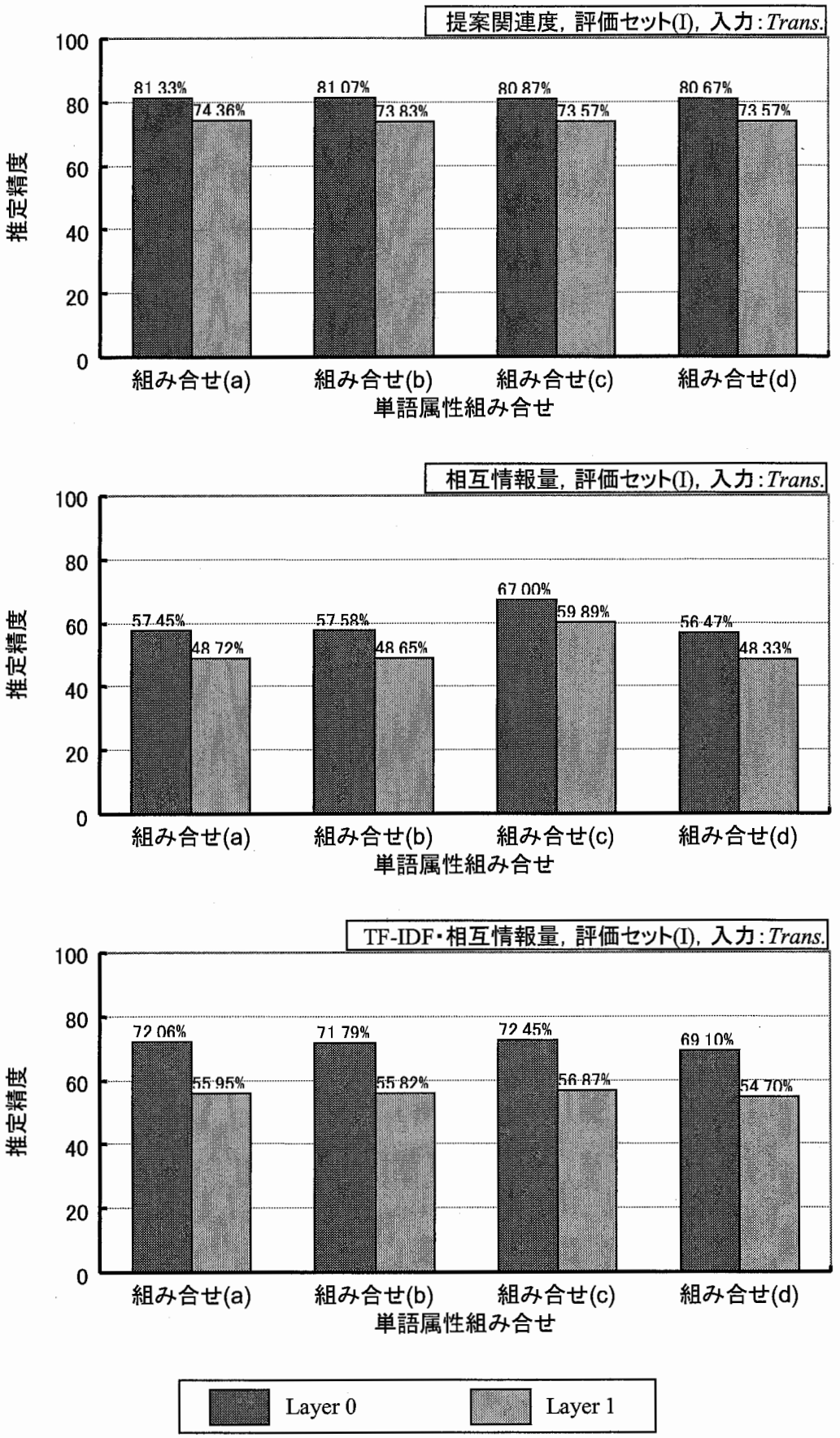

図 10 {評価セット(I)，書き起しを入力，品詞フィルタあり}の条件下での推定精度

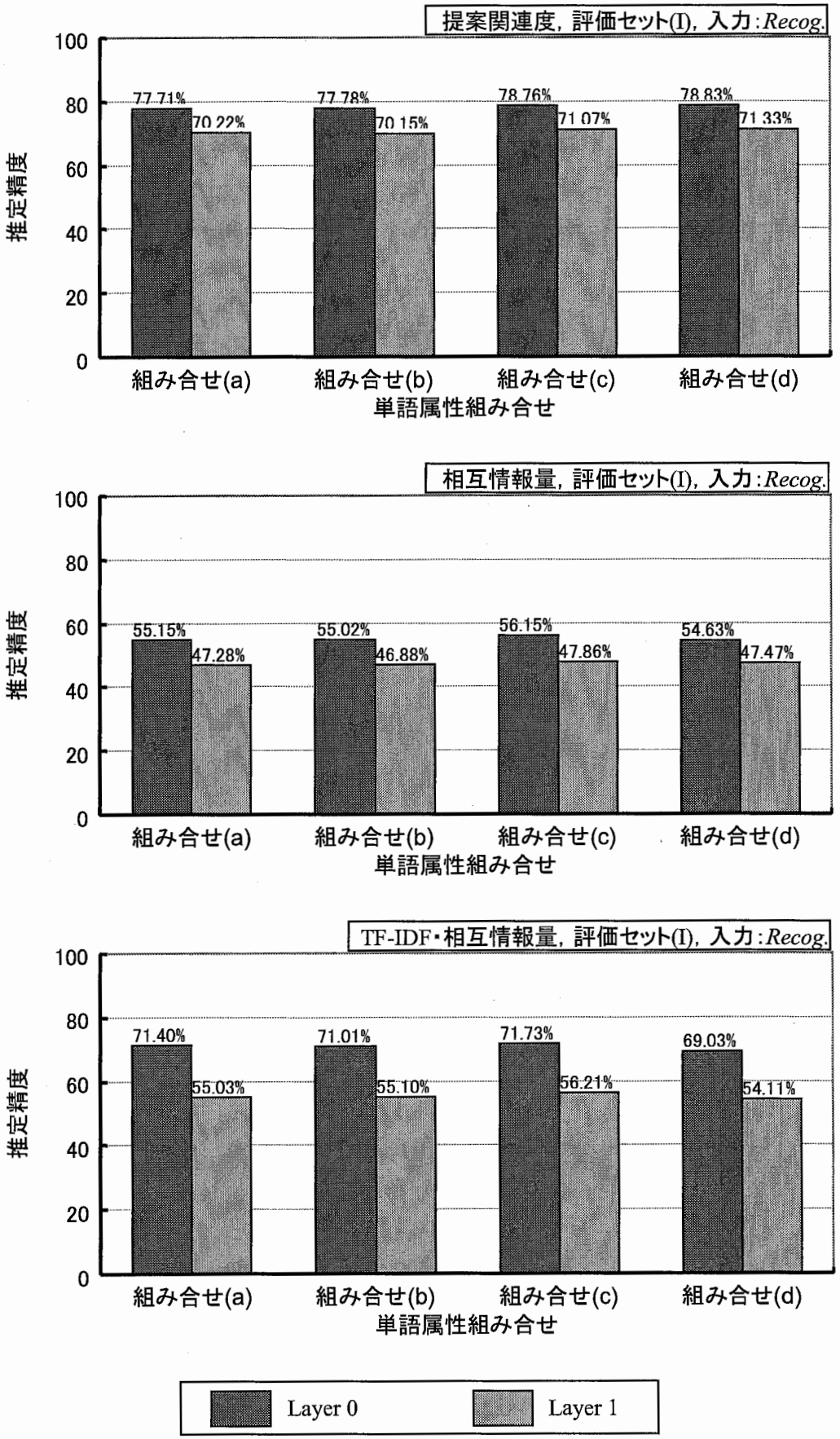

図 11 {評価セット(I), 音声認識結果を入力, 品詞フィルタあり}の条件下での推定精度

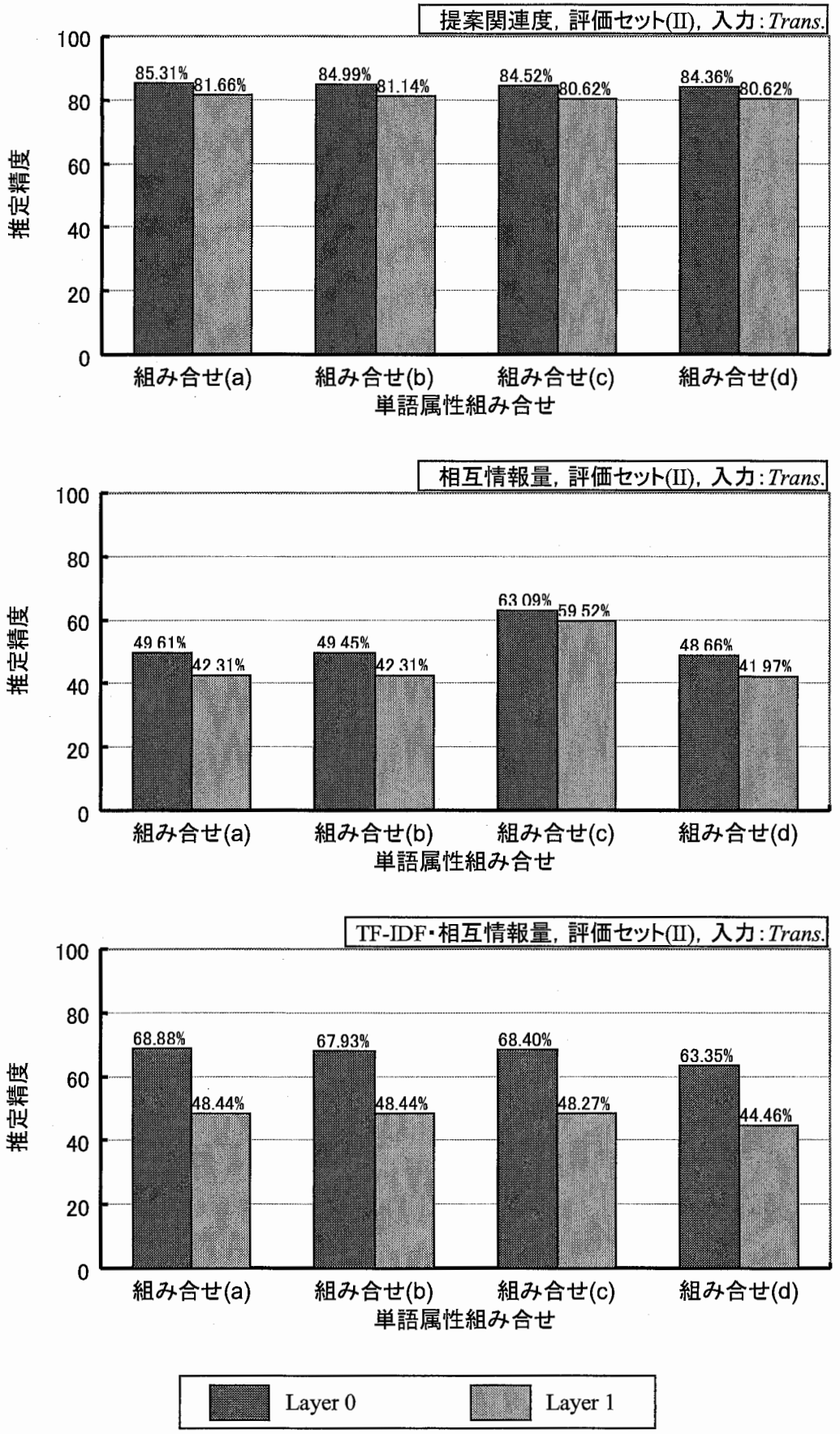

図 12 {評価セット(II)，書き起しを入力，品詞フィルタあり}の条件下での推定精度

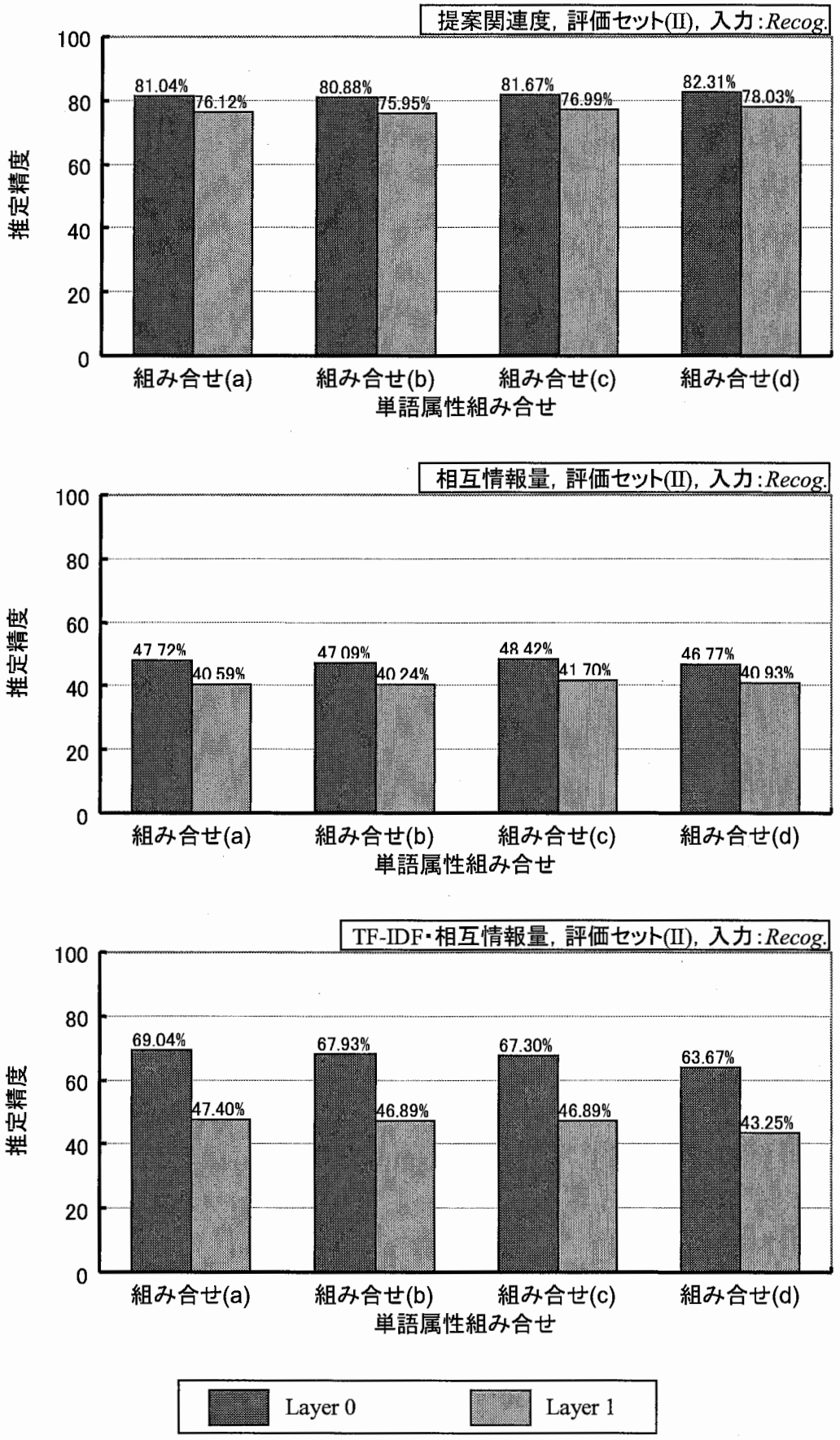

図 13 {評価セット(II), 音声認識結果を入力, 品詞フィルタあり}の条件下での推定精度

表 9 SVM による話題推定の推定精度

評価セットの種類	入力ベクトルの要素	Layer 0		Layer 1	
		Trans.	*Recog.*	*Trans.*	*Recog.*
評価セット(I)	単語が出現する(1)/しない(0)	61.6	59.0	51.4	48.7
	単語と話題の関連度([-1,1]の実数)	21.5	20.9	16.0	15.8
評価セット(II)	単語が出現する(1)/しない(0)	75.7	72.9	66.0	62.1
	単語と話題の関連度([-1,1]の実数)	37.0	35.4	27.9	27.4

表 10 SVM による話題推定の再現率, 適合率(評価セット(I))

入力ベクトルの要素		評価セット(I)			
		Layer 0		Layer 1	
		Trans.	*Recog.*	*Trans.*	*Recog.*
単語が出現する(1)/しない(0)	再現率	16.8%	16.3%	9.3%	9.0%
	適合率	74.8%	73.3%	61.2%	58.9%
単語と話題の関連度([-1,1]の実数)	再現率	6.4%	6.2%	2.9%	2.8%
	適合率	93.5%	91.5%	84.3%	84.8%

表 11 SVM による話題推定の再現率, 適合率(評価セット(II))

入力ベクトルの要素	入力単語列の種類	評価セット(II)			
		Layer 0		Layer 1	
		Trans.	*Recog.*	*Trans.*	*Recog.*
単語が出現する(1)/しない(0)	再現率	71.6%	69.2%	51.0%	48.8%
	適合率	79.8%	76.3%	60.6%	56.3%
単語と話題の関連度([-1,1]の実数)	再現率	37.6%	38.1%	19.8%	20.5%
	適合率	97.7%	95.8%	86.0%	85.6%

5.4 SVM による話題推定

SVM による話題推定実験の結果を，表 9～表 11 に示す．この実験では，単語属性の組み合せ(c)を用いている．また，表 10 および表 11 の再現率，適合率は話題毎に求めた値の平均値である．

入力ベクトルを「単語が出現する(1)/しない(0)」で表現する場合，表 9 に示す推定精度を，提案関連度を用い，4.2 で述べた方法による推定精度（図 6～9）と比較すると，SVM による推定結果は，評価セット(I)を用いる場合で 20～25 ポイント，評価セット(II)を用いる場合で 10～15 ポイント劣る．しかし，相互情報量，TF-IDF･相互情報量による場合を概ね上回る結果である．これに対し，入力ベクトルを「単語と話題の関連度([-1,1]の実数)」で表現する場合の推定精度は大幅に劣化する．ここでの関連度は，式(3)で求められる関連度である．

ここで，表 10，11 に示す再現率および適合率を見ると，入力ベクトルを「単語が出現する(1)/し

ない(0)」で表現し，評価セット(II)を用いる場合，Layer 0 で 70～80%, Layer 1 で 50～60%の値で再現率，適合率のバランスが取れており，推定精度とも対応している．他方，入力ベクトルを「単語と話題の関連度([-1,1]の実数)」で表現する場合と，評価セット(I)を用いる場合は，再現率の値が低く，適合率の値が高くなっている．特に，入力ベクトルを「単語と話題の関連度([-1,1]の実数)」で表現する場合，その差が大きくなっている．これには，次の 2 つの理由が考えられる．

まず，1 つ目は評価セットの違いによる影響である．表 10 に示す再現率は，話題毎に計算した結果の平均値であるが，評価セット(I)に含まれる発話は，複数の話題に属するものが多い．このため，各話題で再現率を計算する際の分母となる発話数の和と求めると，話題の重複分だけ評価セット(I)の発話数よりも多くなる．すなわち，再現率の計算における分母が大きくなるため，各話題の再現率の値は小さくなる．これは，評価値の計算上の問題であり，SVM による話題推定の精度そのものの問題ではない．表 10 の，入力ベクトルを「単語が出現する(1)/しない(0)」で表現する場合の再現率，適合率のアンバランスは，主にこの影響によると考えられる．

2 つ目は，SVM による分類モデルの影響である．入力ベクトルを「単語と話題の関連度([-1,1]の実数)」で表現する場合，単語の出現／非出現だけでなく，単語と話題との関連度まで考慮し，詳細に識別学習をすることにより，過学習のような状態になっていると考えられる．すなわち，評価セットに含まれる発話を，話題毎に「極めて確信度が高い発話」とそうでない発話に分類している．これにより，適合率の値が非常に高く，逆に再現率の値が低くなっていると考えられる．この場合，推定誤りは，「誤った話題への分類」ではなく，「どの話題にも分類されない」誤りが多い．表 10 において，入力ベクトルを「単語と話題の関連度([-1,1]の実数)」で表現する場合の再現率，適合率の極端な違いは，前述の評価セットの影響と，過学習状態の影響が重畳しているものと考えられる．また，評価セット(II)は，話題に関する曖昧性の小さな発話で構成されているため，表 11 に示す再現率，適合率には，主に過学習状態の影響が現れていると考えられる．

一般に，SVM は汎化能力が高いとされるが，この実験の結果は，分類対象の表現を，過度に詳細にしてしまうと，必ずしもその能力が活かされないことを示していると考えられる．

また，SVM は分類対象を正例／負例に分類するが，今回扱ったデータは，評価セット(I)のように，複数の話題に属する発話が多数存在する．同時に，学習データは大規模であり，評価セット(II)のように，曖昧性の小さな発話だけに絞り込むことは困難である．すなわち，現状の話題分類体系では，SVM の適用は必ずしも適切ではない．これは，評価セット(II)を用いた場合でも，提案関連度を用い 4.2 で述べた手法による推定精度に及ばない原因と考えられる．SVM の適用を第一義とするならば，SVM で分類しやすい話題分類体系を構築する，学習データを曖昧性の小さな発話に絞り込む，等の処理を施した上で適用しなければならないと考えられる．

表 12 3best までの話題推定精度

話題階層	Layer 0		Layer 1	
	Trans.	*Recog.*	*Trans.*	*Recog.*
1best	85.8	83.3	81.3	78.4
2best	94.8	92.4	89.3	86.2
3best	95.9	94.8	91.2	88.8

提案関連度，単語属性組み合せ(c)による．

表 13 発話行為の推定精度(BTEC の評価セットによるオープンテスト)

	Trans.	*Recog.*	劣化率
BTECモデル	71.5%	70.2%	1.8%
SLDBモデル	67.0%	66.6%	0.6%
BTEC+SLDB混合モデル	70.5%	69.2%	1.9%

表 14 発話行為の推定精度(クローズドテスト)

評価データ		BTECモデル	SLDBモデル	BTEC+SLDB混合モデル	SLDB拡張モデル
BTEC	1best	80.9%	69.0%	82.4%	---------
SLDB	1best	30.7%	71.5%	70.1%	76.8%
	2best	46.3%	84.1%	84.0%	86.1%
	3best	60.6%	88.5%	90.0%	89.7%

5.5 提案関連度による 3best までの話題推定結果

表 12 に提案関連度，単語属性組み合せ(c)による話題推定の 3best までの推定結果を示す．音声認識誤りを含む入力から話題を推定する場合でも，3best までの累積ならば Layer 0 で 94.8%, Layer 1 で 88.8%の精度が得られている．ここで例えば，10 種類のアプリケーションを扱い，かつ，それらを任意のタイミングで自由に利用できる音声インタフェースを想定する．この実験結果は，そのようなインタフェースにおいて，ユーザ発話の，若干の誤りを含む認識結果のみを用いて，ユーザの要求するアプリケーションを約95%の精度で３つに絞り込めることを意味する．これは，本研究で提案する話題推定手法が，音声インタフェースのユーザビリティ向上に有効であることを示すものである．

5.6 発話行為の推定

表 13 に，BTEC の評価セットを用いたオープンテストによる発話行為の推定精度を示す．話題の推定実験の結果を踏まえ，発話行為推定の条件は，提案関連度を用い，単語の属性組み合せ(c)，品詞フィルタ無しとしている．BTEC モデル，SLDB モデルは，それぞれ BTEC，SLDB から関連度を計算し，BTEC+SLDB 混合モデルは，両者をまとめて関連度を計算している．BTEC モデルを用いた場合の推定精度を，同条件による Layer 0 の話題の推定精度と比較とすると，分類数が少

ないにもかかわらず，10 ポイント以上低い値となっている．この理由として，元々BTEC が発話行為の分布に関する偏りが大きく，BTEC を対象として発話行為推定を行なうこと自体が妥当ではないこと，発話行為を推定する手がかりとして，発話中に含まれる単語だけでは不十分であること，等が考えられる．

表 14 にクローズドテストの結果を示す(注 4)．この表で，モデルと評価データが同じ場合と異なる場合を比較すると，BTEC モデルで SLDB を評価データとする場合のほうが，その逆の場合に比べ，推定精度の落ち込みが大きい．これは，BTEC での発話行為の出現分布の偏りが大きく，SLDB で出現する発話行為に，推定できないものが多く含まれることを示している．一方，SLDB での発話行為の出現分布は，図 3 に示すようにスムーズであり，BTEC で出現する発話行為もカバーしているので，極端な落ち込みがないと考えられる．また，BTEC+SLDB 混合モデルを用いる場合は，BTEC，SLDB ともに，モデルと評価データが同じ場合と同程度の推定精度となっている．混合による相乗効果はないと考えられる．

また，SLDB 拡張モデルは，各発話の話者役割（ホテルのフロント係側，客側）と直前の相手発話の発話行為を，単語と同様に扱い関連度を求めたものである．BTEC は表現集形式であり，一連の対話を形成していないため，このような情報は利用できない．しかし，発話行為の推定では，話者役割や，対話における直前発話の発話行為の種類も有益な情報になると考えられ，表 14 に示す SLDB 拡張モデルによる推定精度は，SLDB モデルに比べ，1best で約 5 ポイント向上している．SLDB 拡張モデルは，話者役割や直前発話の発話行為を，単純に単語とみなしているが，これらの情報と実際の単語の間には「客の立場なら使うが，フロント係側なら使わない単語」などの制約が存在するものと考えられる．推定手法は異なるが，これらの制約も考慮して推定した場合，80〜90%の推定精度が得られている[15][19]．直前発話の発話行為や，話者役割などの情報は，音声インタフェースでは必ずしも利用できないが，本稿で提案した推定手法を基礎として，可能な限りこれらの情報を考慮することにより，発話行為の推定精度を向上できると考えられる．

5.7 音声インタフェースにおける話題推定の応用

音声インタフェースにおいて考えられる，話題推定の具体的な応用例を挙げておく．
音声インタフェースでは，複数ドメインの処理が可能なシステムが必要であり，マルチドメイン音声対話システムが提案されている[3][20]．その例を図 14 に示す．ここでのドメインとは，例えばカーナビゲーションシステムであれば，ルート設定や施設検索，ネット接続による情報検索などの“機能”である．話題推定の応用を考えるとき，“話題”を“機能”と置き換える．すなわち，ユーザの発話は，「どの“機能”を利用したくて発せられたものか？」を推定する．

(注 4) 厳密には，BTEC モデルで SLDB のデータを推定対象にする場合，および SLDB モデルで BTEC のデータを推定対象にする場合は，クローズドではない．

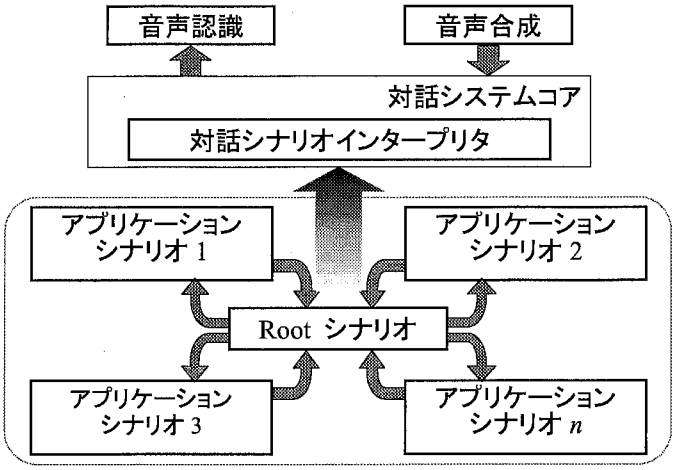

図 14 対話システム例[20]

図 14 で, アプリケーションシナリオとはそれぞれ特定の機能を利用するためのユーザとの対話戦略を記述するスクリプトであり, Root シナリオとは, アプリケーションシナリオ間の遷移を司るスクリプトである. これらのスクリプトは, 共通のシナリオインタープリタで処理され, ユーザとの対話が実行される. ユーザが何らかの機能を使用中でも, 予期できないタイミングで他の機能の使用を要求する場合も考えられる. 例えば, 施設情報の検索中に突然, 渋滞情報の検索を要求する, 等である. これに対し, このシステムはいつでもシナリオの遷移ができるようになっている. あるアプリケーションシナリオの処理中に, そのシナリオで対処できないユーザ発話が入力されると, その発話は Root シナリオ処理に送られ, Root シナリオでは, その発話の処理に適したアプリケーションシナリオを選択し, そのシナリオの処理に遷移する. このシナリオ選択の基準は, [20]では予め定義した要求語によっているが, この部分に, 1 発話毎の話題推定が可能である本稿の提案手法を応用できる. 網羅性の面では, 事前定義の要求語を用いるより, 提案手法の有用性が高いと考えられる.

6 むすび

本研究では, 音声インタフェースに利用可能な意味解析処理のひとつとして, 発話に含まれる単語から, 語順や共起を考慮せずに, 話題や発話行為を推定する手法を提案した.

まず, 音声インタフェースに利用可能な話題および発話行為の推定手法が満足すべき要件として, (1)言い回しを限定しない, (2)幅広い話題を扱う, (3)話し言葉特有の短い発話に対応する, (4)1 発話毎に推定する, (5)音声認識誤りに対してロバストである, の 5 点を述べた.

提案手法は, 単語と話題および発話行為との関連性を示す尺度として, 情報源「ある単語が出現する／しない」と情報源「ある話題または発話行為が出現する／しない」の相互情報量を単語の情報源のエントロピーで正規化した数値を導入し, 入力に含まれる単語から, 話題および発話行為を推定する.

また，本研究で使用した BTEC は，各発話が短い，一連の対話を形成していないなどの特徴が音声インタフェースで想定される発話に共通する．

この BTEC を利用した話題推定実験では，提案手法は良好な推定精度を示し，冒頭に挙げた要件に対し，有効な手法であることが示され，音声インタフェースのユーザビリティ向上に役立つと考えられる．

さらに，SVM を用いた話題推定実験も行なった．SVM を用いた場合の推定精度は，最も良好な結果を得られる条件でも，提案手法による結果に比べ若干劣る．これは，BTEC には，複数の話題に分類される発話が多く含まれ，SVM の適用に不向きであることが考えられる．また，一般に SVM は汎化能力が高いとされるが，入力データの表現によっては，過学習状態になる可能性も示唆された．

発話行為推定に関しては，単語のみを手がかりとする現状では，推定精度は十分ではない．また，BTEC は発話行為の分布に偏りがあるため，発話行為の推定に関する検討に不向きとも考えられる．しかし，提案手法を基礎として，考慮可能な情報を加味することにより，精度向上が期待できると考えられる．

文　献

[1] 小林哲則，“音声対話研究の原状と動向，”人工知能学会誌，17 巻，3 号，pp.266-270，May 2002.

[2] 竹澤寿幸，菅谷史昭，“音声インタフェース，”ヒューマンインタフェース学会誌，vol.3，No.2，pp.103-106, May 2001.

[3] 長森誠，河口信夫，松原茂樹，外山勝彦，稲垣康善，“マルチドメイン音声対話システムの構築手法，”情処学研報，SLP-31-7，pp.45-52，2000.

[4] 平博順，春野雅彦，“Support vector machine によるテキスト分類における属性選択，”情処学論，Vol.41，No.4，pp.1113-1123，April 2000.

[5] 櫻井光康，有木康雄，“キーワードスポッティングによるニュース音声の索引付けと分類，”信学技報，SP96-66，pp.37-44，1996.

[6] 横井謙太郎，河原達也，堂下修司，“キーワードスポッティングに基づくニュース音声の話題同定，” 情処学研報，SLP-6-3，pp.15-20，1995.

[7] 鷹尾誠一，緒方淳，有木康雄，“ニュース音声に対する検索方法の比較，”情処学研報，SLP-29-17，pp.97-102，1999.

[8] K. Ohtsuki, T. Matsuoka, S. Matsunaga, S. Furui, “Topic extraction based on continuous speech recognition in broadcast news speech,” IEICE Trans. Inf. & Syst., vol.E85-D, no.7, July 2002.

[9] A.L. Gorin, “Processing of semantic information in fluently spoken language,” Proc. ICSLP-96, pp.1001-1004, 1996.

[10] C. Wu, Q. Zhou, H-K.J. Kuo, A. Saad, D. Attwater, P. Durston, M. Farrell, F. Scahill, “Natural

language call steering for service applications," Proc. ICSLP-2000.

[11] H-K.J. Kuo, C-H. Lee, I. Zitouni, E. Fosler-Lussier, E. Ammicht, "Discriminative training for call classification and routing," Proc. ICSLP-2002, pp.1145-1148, 2002.

[12] T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, S. Yamamoto, "Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world," LREC 2002, pp.147-152, 2002.

[13] L. Levin, D. Gates, A. Lavie, and A. Waibel, "An interlingua based on domain actions for machine translation of task-oriented dialogues," Proc.ICSLP-98, pp.1155-1158, 1998.

[14] http://www.is.cs.cmu.edu/nespole/db/index.html

[15] 谷垣宏一，匂坂芳典，"中間言語表現の生成を目的とした統計的音声理解方式，" 信学論（D-II），vol. J83-D-II, no.11, pp.2428-2437, Nov. 2000.

[16] 堀部安一，情報エントロピー論，森北出版，1989.

[17] 永田昌明, 平博順, "テキスト分類－学習理論の見本市－, "情報処理, vol.42-1, pp.32-37, 2001.

[18] K. Asami, T. Takezawa, G. Kikui, "Topic detection of an utterance for speech dialogue processing," Proc.ICSLP-2002, pp.1977-1980, 2002.

[19] 浅見克志，竹澤寿幸，菊井玄一郎，"自動生成されたルールによる発話行為の推定，" 情報処理学会第 63 回全国大会，vol.2，pp.199-200，2001.

[20] 笹木 美樹男, 浅見 克志: 車載マルチメディアにおける対話エージェントについて, シンポジウム「カーナビ・携帯電話の利用性と人間工学」, pp.165-170，2000.