

車載情報機器との音声言語インタフェースにおける 文体統一の効果

森田 浩康[†] 林 真弘[†] 伊藤 敏彦[†] 甲斐 充彦[†] 小西 達裕[†] 伊東 幸宏[†]
桂川 桂子[‡] 大野 健[‡]

[†] 静岡大学 〒432-8011 静岡県浜松市城北 3-5-1

[‡] 日産自動車株式会社 〒237-8523 神奈川県横須賀市夏島町 1 番地

E-mail: [†] cs8087@cs.inf.shizuoka.ac.jp

あらまし 本稿では、車載情報機器を音声操作する際に機器ごとに入力できる文体や語彙が異なるため、ユーザが異なった入力方法を習得しなければならず負担になってしまうという問題に対して、一貫性のある入力を用いた音声言語インタフェースを提案する。そこで、車載情報機器の特徴と、音声入力する際の問題点を整理し、入力文体の統一という手法について述べる。車載情報機器には様々な機能が搭載されているが、まずは訪問地検索にタスクを限定してシステムを実装した。一般的な対話形式との比較など、2つの評価実験を行い、我々の提案する手法が有効であることを実証した。

キーワード 音声言語インタフェース、音声対話システム、文体統一、入力形態の一貫性、車載情報機器

The effect of the style unification in a spoken language interface with car navigation system

Hiroyasu MORITA[†] Michihiro HAYASHI[†] Toshihiko ITOH[†] Atsuhiko KAI[†]

Tatsuhiko KONISHI[†] Yukihiro ITOH[†] Keiko KATSURAGAWA[‡] and Takeshi OONO[†]

[†] Shizuoka University 3-5-1, Johoku, Hamamatsu, Shizuoka 432-8011, Japan

[‡] Nissan Motor Co. Ltd. 1, Natsushima-cho, Yokosuka, Kanagawa 237-8523, Japan

E-mail: [†] cs8087@cs.inf.shizuoka.ac.jp

Abstract When a human interface accepts voice input, the vocabulary at sentence styles to be used are different from those for another device accepting voice input. The increase of such devices forces users to learn different input methods. In this paper, we propose a spoken language interface using a consistent input method which can be applied to every voice input device. We examine problems of voice input in car navigation systems and describe the technique for unification of sentence styles. We have implemented a system for destination search and conducted two experiments for evaluating the system.

Keyword spoken language interface, voice dialog system, style unification, consistency of an input form, car navigation system

1. はじめに

今日 IT の発展に伴い、ETC サービス等を搭載した車載情報機器の普及が進んでいる。車載情報機器は訪問地検索や経路探索のほかに、インターネットやオーディオなどの様々な機能を備えており、単なるナビゲーションシステムとしてではなく、ユーザに様々な情報を与える情報機器として使用されている。今後も高機能化が進み、今以上に多数のユーザに使用されていくものであると予想される。

車載情報機器の操作方法としては、リモコンやタッチパネルが一般的である。リモコンやタッチパネルで

の操作は、車載情報機器だけに限らず、テレビやビデオ、または銀行の ATM など様々なものに使用されており、ほとんどのユーザが使用した経験があるはずである。これらの操作方法に加えて、近年音声入力も普及し始めている。なぜなら、運転中にシステムを安全に操作できる手段の一つであるからである。だが、ユーザが音声で機器を操作する機会は少なく、リモコンやタッチパネルに比べると浸透していない。しかし、車載情報機器に限らず様々な機器においても音声入力を用いる場面が増加することが想定されるため、ユーザにとって利用しやすい音声言語インタフェースが必

要である。

音声入力において、ユーザが自由に発話した入力をすべて正確に認識できることが最も望ましい。しかし、現状の音声認識技術では自然な発話を 100% 認識することは不可能である。したがって、現状の技術を用いてできる限り高い認識率を保つためには、入力できる文体や語彙をある程度限定しなければならない。そのような制約の中でこれまでに車載情報機器の音声インタフェースに関する様々な研究[1][4]が行われているが、機器ごとに入力可能な文体や語彙が異なっている。そのため、ユーザは機器ごとに利用可能な文体や語彙を変えなければならない、大変負担となってしまう。このような、操作方法習得に対するユーザの負担を軽減し、利用しやすい音声言語インタフェースを構築するために、一貫性のある入力文体や語彙を用いた入力方法を提案する。車載情報機器には多数の機能が存在するが、まずは訪問地検索にタスクを限定してシステムを実装し、評価する。

2. 車載情報機器における音声入力

本章では、まず、車載情報機器の特徴をまとめる。そして、問題点を整理し、本研究で目標とするシステム像を述べる。

2.1. 現状における車載情報機器の特徴

現在市販されている最新の車載情報機器には、ナビゲーションシステム機能、ドライブプラン作成機能、オーディオ機能など様々な機能が搭載されており、年々進出し続けている。また、運転中の情報提供が第一目的であるため、早いレスポンスが求められるが、CPU の性能の向上や、メモリ容量の増加により、性能自体も向上している。

ほとんどの場合が車内で使用され、取り付けられる位置もほぼ決まっているため、画面の大きさに制限があり、6~8 インチと狭い。また、運転中の安全確保のために、表示可能文字数にも制限がある。これらの制限の中で、画面上により多くの情報量を含ませるために、ほとんどの車載情報機器ではリスト形式の表示を用いている。

車載情報機器の操作方法としては、リモコンやタッチパネルが一般的である。そのため、ユーザは画面に表示されているリストから適切なものを選択するだけでよい。しかし、運転中は運転操作の妨げになってしまうため、リモコンやタッチパネルでの操作は禁止されている。停車中に車載情報機器を操作する場合には問題ないが、急な経路変更のように運転中にも車載情報機器を操作するような状況も想定されるため、近年音声での入力が重要視されている。音声で操作する場

合は画面に表示されているものを読み上げればよい。また、近年では音声認識技術の向上により単語での入力だけでなく、「〇〇に立ち寄る」などの発話も可能となり、「静岡県 浜松インター」のように、区切ることなく連続して発話することも可能となっている。また、音声入力を使用する過程で車載情報機器がユーザの声の特徴を学習し、継続して使い込むほど音声認識率が向上する機能を兼ね備えたものなども存在する。

しかし、使用される状況が走行中の車内であるため、車のエンジン音や車内の会話、または、オーディオの音など、様々なノイズが存在する。また、音声入力に慣れていないユーザが多く、入力の際に迷ってしまい、認識不可能な文体や語彙での入力をしてしまうこともある。

これらの要因により、音声操作はリモコンやタッチパネルでの操作に比べると確実性に欠けてしまう。また、認識率が悪く誤認識が多いと、ユーザは音声入力に抵抗をもってしまい、使いにくいものだと思われてしまう可能性もある。したがって、運転中でも操作することが可能ではあるが、現状では音声入力がユーザにとって最も有効な操作方法であるとは言えない。そのため、音声で入力する際のユーザの負担となっている原因を明らかにし、対策を検討する必要がある。ノイズの影響による認識率の低下も想定されるが、我々は入力できる文体、語彙の制限に伴いユーザが音声入力を行う際に迷ってしまう点に着目する。次節ではこれらの問題点をより詳細に述べる。

2.2. 問題点の整理

2.2.1. 機器ごとに異なる入力文体と語彙

前節にも述べたが、現状における車載情報機器での音声入力では、発話可能な語彙や文体が制限されている。そのため、機器ごとに入力可能な文体や語彙が異なっているのが現状である。例えば、現在地点を基準としてそこから近くのコンビニを探したい場合に、ある機器では「現在地の近くのコンビニ」、他の機器では「コンビニ探す」のように、ユーザはその機器が認識できる文を発話しなければならない。そのため、ユーザは機器ごとに異なった入力方法を習得しなければならない、大きな負担となってしまう。したがって、認識不可能な文体や語彙での発話が増加してしまい、認識率が低下してしまうのである。

機器ごとに発話可能な文体や語彙を使い分けなければならない上、誤認識されてしまったら、ユーザビリティは大きく低下してしまう。この機器ごとに入力できる文体や語彙が異なり、操作方法習得を困難にしている問題に対して、入力できる文体や語彙に一貫性を持たせることでその負荷を軽減させる。また、一貫

性を持たせるだけでなく、ユーザビリティの向上のため、高い音声認識率を保つことができる音声インタフェースが必要である。

さらに、これらに加えて、ユーザにとってできる限り馴染み深い表現を用いることが大切である。普段使っている言葉に近ければ近いほどユーザが使用できる文体や語彙を理解しやすい上、スムーズに発話できるため誤認識の防止にもつながる。

2.2.2. 車載情報機器の画面表示

現在の車載情報機器では、画面の大きさなど様々な制限がある中で最大限発話可能な文体や語彙を表示するためにほとんどの車載情報機器でリスト表示を用いている。入力の際、ユーザに自由に発話させることが理想であるが、現状では不可能である。そのため、ユーザがリスト表示を見て適切なものを発話させることを想定している。

自然言語での入力では、ユーザはある意図を思いのままシステムへ伝えればよいが、リスト入力では、ユーザは自分の意図をリストからの選択形式に置き換える必要がある。しかし、同一の意図を示すようなリスト操作は、一意になるとは限らない。例えば、「浜松市のレストラン」を探すためのリスト操作の最初には「検索方法」という項目から「地名」を選択しなければならないようなリスト操作もあれば、「市区町村名」という項目があって、そこから「浜松市」を選択すればよいようなリスト操作もある。また、目的の施設を探すために、ジャンルを指定する必要があるようなリスト形式では、ユーザの意図する施設がシステムの提供するどのジャンルに分類されているかが分からなくてユーザが戸惑うような場合もある。例えば、ある人は「図書館」が「公共施設」に分類されるべきであると考えられるかもしれないし、またある人はこれが「文化施設」に分類されるべきであると考えられるかもしれない。

つまり、ユーザがある意図をもったときに、その意図をシステムに伝えるための手段としてユーザが想定するリストの形式は個人によって異なるため、システムが提示するリストの形式と、ユーザの想定したリストの形式が必ずしも一致するとは限らないということである。そのような場合には、ユーザはシステムが提示したリストを理解しなければならない。また、リストを見ても入力可能な文体や語彙をイメージすることはできないため、入力に戸惑ってしまう。

したがって、ユーザの想定するリストとシステムが提示するリストとを一致させるために、ユーザの意図を一意的リストに変換する変換規則が必要である。また、ユーザの音声入力を手助けするために、画面を一瞬見るだけで発話可能な文体、語彙をイメージさせる

ような工夫も必要である。

2.3. 目指すシステム像

現状の車載情報機器と、それらの問題点を考慮して、本研究で目標とするシステム像を述べる。まず、現状の車載情報機器には様々な機能が搭載されているが、我々は、まずは訪問地検索にタスクを限定してシステムを実装する。訪問地検索を行う GUI は、リスト形式を用いる。本研究で実現するシステムでは、以下の三つを重視して音声言語インタフェースを構築する。

- －習得に負荷がかからない一貫性のある文体、語彙
- －認識可能な文体、語彙をイメージさせるリスト表示
- －音声認識率の安定

前節で述べた問題点は、ユーザが初めてその入力方法をとる場合にその操作方法習得の困難性によって生じるものである。逆に言えば、その入力方法を一度習得してしまえば、解決できる問題がほとんどであると言えることが出来る。そのため、ユーザにとって違和感がない上、入力文体や語彙に一貫性を持たせることで、ユーザは入力の際に戸惑うこともなくなり、機器操作習得の負荷を軽減することができると考えている。それに加えて、音声認識率が向上すれば、ユーザにとって音声入力を使いやすいものになり、リモコンやタッチパネルでの操作同様に信頼性が高い操作方法であることが認識され则认为している。次章では上記三つを実現するための手法を述べる。

3. 文体の統一

ユーザが音声入力を行う際に、機器ごとに入力できる文体や語彙が異なるという問題や、ユーザの想定したリストとシステムが提示するリストとが一致しないという問題点が入力方法習得の負荷の原因となっていることは前章で述べた。そこで、我々は、ユーザの意図を発話可能な文体や語彙に変換する規則を提案し、一貫性のある入力が必要であると考えた。そこでまず、訪問地検索だけでなく、車内機器の操作などを含めた車載情報機器で操作するすべての状況を調べると、次の2つの文体を組み合わせることで表現可能であることが分かった。

－対象 を 操作

－属性 は 値

例えば、「静岡市の飲食店を探す」という場合には、「訪問地 を 探す」＋「ジャンル は 飲食店」＋「市区町村名 は 静岡市」と表現する。また、「エアコンをつけて 28 度にしたい」という場合は、「エアコン を つける」＋「温度 は 28 度」と表現できる。

このため、リスト構造がこの二つの文体に沿う形式になっていれば、すべての操作はこれら二つの文体に

変換しさえすればその意図をシステムに伝えることができる。文体とリスト構造を一致させるためには、「対象」と「属性」に当たる項目を縦に並べ、「操作」と「値」に相当する項目をそれぞれ「対象」や「操作」の右側にリスト形式に表示する。操作手順としては、例えば、「ジャンルは飲食店」のように「対象を操作」「属性は値」に相当する操作を、音声でそのまま読み上げる。また、「静岡駅の近くの飲食店を探す」といった場合には、目的の施設である飲食店と、その近隣施設としての目印である駅の二つの施設が同時に話題の対象となりうる。このような場合には、それぞれの対象の対象が、地名やジャンルなどの属性を個別に持つこととなり、単に「地名は静岡市」とだけ指定してもどちらの対象にその制約がかかるのかを判断できなくなる。そこで本システムでは、このように、同一の属性をもつ複数の対象を明確に区別するためにタブによる表示形式を採用した。タブ表示の例を図1に示す。例えば、「静岡駅の近くの中中華料理店」は、目的の施設である「中国料理店」とその近隣施設である「静岡駅」に関する条件を、異なるタブに表示し、二つの施設に関する話題の転換をタブの選択によって表現する。タブの選択も「ページ」は「近隣施設」という、「対象を操作」の形式で表現することができる。このように、同一の属性をもつ対象間で話題の転換が起こるような複雑な意図でも、「対象を操作」「属性は値」の二つの文体で表現することができる。また、これをリストの選択形式をGUI表示することでユーザにその構造を示すことが可能である。

このようにユーザの意図を一意的リストへ変換する規則を用いることで、ユーザの想定するリストとシステムが提示するリストとを常に一致させることが可能となる。音声入力の際に入力できる文体や語彙に迷うという問題も、統一文体にあわせてリストをそのまま読み上げるだけで解決できる。これにより、入力文体や語彙に一貫性を持たせることも可能になる。

また、統一文体は、入力の際の迷いを解消するだけでなく、一度理解してしまえば、他の統一文体を用いた機器を操作する際に負荷を生じることなく操作方法を習得できるという効果も期待できる。

4. 評価実験

本章では、我々が実現した統一文体「属性は値」「対象を操作」を用いたシステムの評価実験を行った。二つの実験を行ったため、それぞれについて述べる。

4.1. 統一文体の有用性の実証

まず、統一文体を一度理解してしまえば、他の統一文体を用いた機器を操作する際に負荷を生じることな

く操作方法を習得できるという効果の実証を行う。そこで、統一文体に慣れた被験者と、初心者である被験

図1 統一文体を用いたシステムのGUI

者に他の統一文体を用いた機器を使用させ、慣れた被験者が負荷なくシステムを操作できることを実証する。

4.1.1. 実験方法

被験者は、工学部系学生20名とした。これを、10名ずつ2つのグループに分けた。まず、グループAには、統一文体を用いたエアコン設定システムを30分ほど使用させ、統一文体に慣れさせた。エアコン設定システムとは、「エアコンをつける」「温度は28度」のような発話が可能なシステムである。グループBには、自然言語での入力が可能なエアコンシステムで20文正確に認識されるまで発話させ、音声操作に慣れさせた。自然言語での入力が可能なエアコンシステムとは、「寒いから温度を上げたい」など限定した発話のみ認識させるように作成したシステムである。その後、すべての被験者に本システムのマニュアルを読ませ、「浜名湖パルパルの近くのホテル」など、いくつか訪問地検索を行わせた。システムを使用する際にはリセット以外のすべての操作を音声で行わせた。

4.1.2. 実験結果

表1に実験結果を示す。一般的に、入力文体に慣れている被験者に比べて、慣れていない被験者は、想定していた文体や語彙と異なる発話を行うことが多い。したがって、認識率が低下し、訂正発話の回数が増加

するため、総発話回数が多くなると想定していた。しかし、表 1 より音声認識率を比較するとグループ A が 77.0%，グループ B が 73.6%と大きな差は見られなかった。また、総発話回数に関しても、グループ A とグループ B に大きな差は生じなかった。この結果より、想定していた慣れに関する効果は実証できなかった。

しかし、今回の実験結果では、マニュアルを読ませただけの文体に慣れていない被験者でも文体に慣れた被験者と同等の認識率や発話回数で訪問地検索を行っている。この結果より、統一文体は、初心者でも十分使いこなせ、マニュアルを読んだ程度で身につけることができるため、習得の際に負荷がかからない表現であることが分かった。

また、タスク達成時間を調べてみると、タスク 1 ではそれほど大きな差は生じなかったが、タスク 2 ではグループ A の方が短い時間で訪問地検索を行っている。タスク 1 に比べ、タスク 2 の方が難易度が高いため、文体に慣れていない被験者にとっては、難しい訪問地検索を行う際に多少迷いがあったのではないかと推測される。しかし、統一文体に慣れた被験者は悩むことなく訪問地検索を行えている。このため、統一文体に慣れる効果として、難しい検索にも対応でき、統一文体への変換速度が向上することが分かった。

表 1 慣れに関する実験結果

	グループ A	グループ B
音声認識率		
全タスク	77.0%	73.6%
総発話回数		
タスク 1	5.8	5.7
タスク 2	18.4	19.6
タスク達成時間		
タスク 1	1:13	1:47
タスク 2	3:39	6:01

4.2. 一般的な対話形式との比較

統一文体が一般的な対話形式と比較しても有効であることを実証する。以下に実験手順と結果を示す。

4.2.1. 比較システムの概要

本システムと比較するシステムは、本研究室で少ない対話数で訪問地検索を終了させることと、極力自由な発話を行えることに重点を置き設計、実装したシステムである(図 2)。ユーザは、システムからの応答に対し、画面に表示されているリストから適切な項目を音声で選択する。対話回数を減少させるために、最初は大きく分類するための質問をし、徐々に詳細化を進めていく。その詳細化の過程で、例えば、「飲食店」の

後に「ガスト」と発話された場合、ガストに関しては誤認識される可能性があるが、飲食店に関する発話がされたことで、「飲食店」は確定できる。このように、詳細化が進むにつれ、一つ上の階層を確定していく方針を用いる。このため、無駄な対話を防ぐことができ、少ない対話数での対話を実現することが可能である。

もうひとつの特徴として、誤認識された際に画面表示は誤ったまま次の画面へと進んでしまうが、ユーザは誤認識された発話を繰り返すことができ、簡単に対話を修正することが可能になっている。誤認識された発話のいい直しを実現するために、常に一つ前の対話で使用した辞書と文法を保持している。そして、画面に表示しているリストを認識するための文法辞書だけでなく、一つ前の対話で使用した文法、辞書も待ち受け状態にしておく。このような工夫を用いることで、常に訂正発話も認識できる状態で対話を進めることができるため、誤認識された際の対処がスムーズに行える。

認識可能な文体、語彙だが、極力自由な発話をさせることに重点を置いているため、単語での入力も認識可能にした。ユーザが発話可能な文体は、システムの応答に対し左右どちらかに表示されているものをそのまま単語で入力するか、二つのことを入力する際には、「〇〇で××」か「〇〇 ××」である。ユーザは画面に表示されているものを見て発話すればよいので、入力の際に迷うこともなく、短い時間で訪問地検索を終えることができるシステムである。

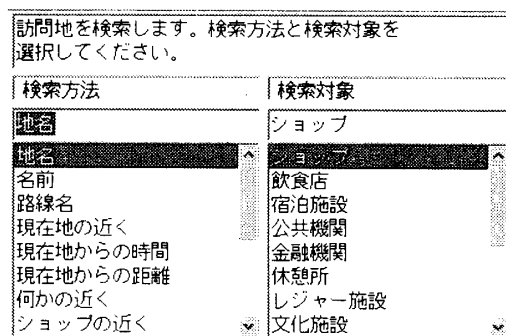


図 2 比較システムの GUI

4.2.2. 実験手順

被験者は、工学部系学生 10 名とした。まず、エアコン設定システムを 30 分ほど使用させ、統一文体に慣れさせた。次に、比較システムのマニュアルを読ませ、訪問地検索を行わせた。最後に、統一文体を用いたシステムのマニュアルを読ませ、訪問地検索を行わせた。各システムで訪問地検索に用いたタスクだが、最小操作回数が同じものを使用し、具体的な名称のみを変えたため、タスクは多少違うが、難易度は同程度である。

4.2.3. 実験結果

評価は7段階評価のアンケート結果による主観評価と、音声認識率などの客観評価を行った。主観評価の結果を表2に示す。以下、比較システムをシステムA、統一文体を用いたシステムをシステムBとする。ほとんどの項目で、システムBの方が高い評価を得た。特に、入力の際に迷ったか(迷わなかった=7)という質問に関して、評価に差が生じた。この結果から、統一文体は使いやすい表現であり、入力の際の迷いを解消できたことが分かった。

客観評価に関して、実験結果を表3に示す。受理可能な発話数の割合に関して、システムAが80.6%に対し、システムBでは95.8%であった。システムAでは認識できない文体や語彙での発話が多かった。しかし、システムBでは認識できない文体や語彙を用いて入力されることはほとんどなかった。したがって、統一文体を用いることで、想定外発話数を減少させることが可能である。次に受理された発話が正しく認識されているかどうかを確認するために、音声認識率も調べた。システムAの語彙数は279語である。また、システムBの語彙数は273語である。したがって、語彙数に大きな差はない。音声認識率を比較すると、システムAでは59.7%だったのに対し、システムBでは74.6%であった。語彙数が同等にもかかわらず、音声認識率に差が生じている。したがって、統一文体は高い認識率が残せる表現であることも分かる。また、タスク達成時間に関してもシステムBの方が短い時間で訪問地検索を行っている。そして、システムの応答が終了してから被験者が発話するまでの時間もシステムBの方が短い。システムAでは、ユーザが意図した語が表示されていないことがあると、入力の際に迷い、考え込むことがあった。しかし、システムBでは、ユーザの意図したリストとシステムが提示するリストとが一致しており、入力の際に迷うことなく発話できたと推測される。

これらの結果から、統一文体は一般的な対話形式と比較しても、利用しやすい表現であり、有用なものであることが実証できた。

表2 一般的な対話システムと比較した際の主観評価

	システムA	システムB
使いやすさ	5.0	5.8
画面表示の分かりやすさ	5.6	5.9
入力に迷ったか	2.6	4.2
入力文体に関する違和感	5.1	5.7
入力回数の多さ	3.7	3.6

表3 一般的な対話システムと比較した際の客観評価

	システムA	システムB
受理可能な発話数の割合		
全タスク	80.6%	95.8%
音声認識率		
全タスク	59.7%	74.6%
タスク達成時間		
タスク1	2:29	1:13
タスク2	4:30	3:25
システム応答から発話するまでの時間		
タスク1	6.61	5.20
タスク2	8.46	5.77

単位：秒

5. まとめ

本論文では、車載情報機器を音声で操作する際に、機器ごとに認識できる文体や語彙が異なり、ユーザは機器ごとに異なった入力方法を習得しなければならない大変負担になってしまうという問題に対して、「属性は値」「対象を操作」という統一文体を用いて入力に一貫性を持たせることを提案した。そして、システムを実装し、評価実験を行い、統一文体の有用性を実証した。統一文体は初心者にも簡単に使える表現であり文体習得には負荷がかからないことと、一般的な対話形式と比較しても、使い勝手や速度の面で十分有効であることが実証できた。

本研究では、車載情報機器に多数搭載されている機能のうち、訪問地検索にタスクを限定して実現したが、訪問地検索だけでなく、車載情報機器の操作すべてを統一文体で操作することを目指す。また、車載情報機器だけでなく、音声入力を用いるすべての機器においても統一文体というものが有効であると考えている。

文 献

- [1] 屋野竹秀, 笹島宗彦, 河野恭之, “音声対話タスクのための高速なキーワードラティスパーザBTH”, 人工知能学会論文誌 Vol.17, No.6, pp.658-666, November 2002
- [2] 伊東幸宏, 小西達弘, 伊藤敏彦, 桂川景子, “ドライブプランニングシステムの自然言語インタフェース”, 人工知能学会誌, Vol.17, No.3, pp.285-290, May 2002
- [3] Keiko Katsuragawa, Michinori Niwa, Takuya Yanagi, Masaki Watanabe, Toshihiko Itoh, Tatsuhiro Konishi, Yukihiro Itoh, “Constructing an Interactive Natural Language Interface for a Drive Planning System” SNLP COCODA, pp.237-242, May 2002
- [4] 桐山伸也, 広瀬啓吉: 話題知識を利用した文献検索音声対話システム 日本音響学会平成12年度秋季研究発表会講演論文集, Vol.1, 1-5-25, pp.49-50, September, 2000