音声言語情報処理 8－4
（1995．10．20）

スケジューリングタスクの自由発話音声の言語的性質

中筋 知己　　山本 英明　　樽松 明

電気通信大学

〒182 東京都調布市調布ヶ丘 1-5-1

E-mail : nakasuji@apple.ee.uec.ac.jp

あらまし　話し言葉による音声対話システムや音声翻訳システムを実現するには、頑強な音声認識と自由発話の言語処理が必須である。このためには、人間が話す音声や言語の基本的研究のベースとなる自由発話のデータベースが重要である。我々は、二人の話者が都合のよいスケジュールをきめるというスケジューリングタスクについて、人間対人間の音声対話データベースを収集している。本稿では、まず、データベースの収集方法について述べる。次に、書き起こしおよび形態素解析の結果から、自由発話音声対話の言語的性質を主に品詞の頻度情報を用いて分析する。また、品詞の細分類についても検討した結果、品詞バイグラム文法のパープレキシティの減少を得た。

キーワード　音声対話, 音声データベース, スケジューリングタスク, 相互情報量, パープレキシティ

Linguistic Characteristics of Spontaneous Speech on the Scheduling Task

Tomomi Nakasuji　　Hideaki Yamamoto　　Akira Kurematsu

The University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo 182, Japan

E-mail : nakasuji@apple.ee.uec.ac.jp

Abstract　The importance of database of spontaneous speech database has been highly emphasized as basis of fundamental researches on spoken language to realize a robust man-machine spoken dialogue system and a speech translation system. We are collecting human-to-human dialogues data of scheduling task in which two speakers arrange a meeting. This report describes how we are building this database. Then analysis on linguistic characteristics of spontaneous spoken dialogues in the database were made mainly in terms of the statistics of the part of speech. Lastly, we attempted clustering of the part of speech, by which the perplexity of part-of-speech bigram grammar decreased.

Keywords　Spoken dialogue, Speech database, Scheduling task, Mutual information, Perplexity

1 はじめに

音声や言語の研究を遂行するためにの基礎となるデータベースの重要性は広く認識され、日本語におけるこのための努力がこれまで重ねられてきている [1], [2], [3]。話し言葉のデータベースについては、従来、主として模擬会話による対話データ、あるいは利用者とシステムとのやりとりの収録がおこなわれているが、まだデータの規模は小さい。ATRにおいては、音声翻訳のための音声対話技術の基礎として、旅行会話のタスクで自然な会話の対話データの収集を進めている [4]。

今後、音声による対話の研究においては、従来の音声や言語の個別のデータだけでは不十分であり、音声対話としての研究開発に役立つ人間対人間の自由発話の会話音声データベースの構築が不可欠であるといえる [5] [6] [7] [8]。音声対話データベースは、これをどのように利用するかによって内容が変わってくる。自由発話の音声認識・理解、対話の理解、対話システムにおける発話メカニズムなどの研究を想定する場合には、データベースは、統計的データをもとにした対話プロセスのモデル化が可能になるように構築されなければならない。これには、音声対話現象を広い範囲でカバーするデータの量の拡大と、対話現象の基本となる表現や談話プロセスを包含した質の良さが必要である。

本稿では、スケジューリングタスクにおける音声対話データベースの構築のためのデータの収集とその言語的性質について述べる。

2 スケジューリングタスクによる音声対話データの収集

2.1 スケジューリングタスク

スケジューリングタスクは、二人の話者がそれぞれ自分の予定スケジュールのカレンダーをもち、打ち合わせや相談について、都合のよい時間や場所を決めるというものである。カレンダーは1カ月分のもので、あらかじめ13種類のカレンダーのシナリオの中から適当なものを選ぶ。このシナリオは、大学における職員の講義や会議に関する予定が時間と内容を記述したもので、2週間から1カ月の予定表が記入されている。ところどころに空いた時間があり、互いに会話によりスケジュールを決めていく（図1）。

話者として、手近な協力者である大学の学生および職員を選んだ。対話の設定は、本人が自分のスケジュールとみなして会話する場合と、教官の秘書という役柄で秘書どうしが会議の日程を決めるという場合を設定した。会話内容の設定は、話者の自由である。会話内容をあまり制約すると、考えることが狭くなって話し方が偏ってしまう。一方、枠組みのない全く自由な会話にすると、話題が広がりすぎて音声対話処理に有用な言語情報が得られないということになる。したがって、カレンダーによるスケジュールという枠組みのなかで自由に話してもらうようにした。カレンダーの制約以外には、話し方については常識的な話し方をしてもらうこと以外には自由に発話する。

会話を開始する前に、話者に対して、この会話を機械による音声認識や音声翻訳の研究に使うことを説明して、極端にくずれた会話にならないように協力してもらうようにしている。また、できるだけ相手が話している間に割り込みはしないようにつとめるよう指示した。

A：えー月曜日はいかがですか？
B：月曜日はー。
A：7日の月曜日。
B：7日ですか。
あのー会議が9時から11時までありまして。
えーそのあとあのー、打合せが、12時から2時。
A：はあはあ。

図 1: 対話例

2.2 データ収集の方法

二人の話者の音声の収録方法としては、(1) ワークステーションに接続されたADコンバータを通してコンピュータに直接録音する方法と、(2) DATテープに録音する方法とがある。ADコンバータから直接録音する方法は、録音レベルがリアルタイムで確認できないため、録音時のレベル調整に手間どる。DATテープに録音する方法では、レベルはモニターで知ることができることと、ポータブルに持ち運びが簡単なため、収集場所が広範囲にできるというメリットがある。

今回のデータ収集における二人の話者の音声の収録は、DATレコーダにデジタル録音した。マイクロフォンは、接話型マイクロフォン(SENNHEISER)を用い、2人の音声を2チャンネルに別々に録音した。録音のレベル設定はマニュアルで行なった。

手ぶりなどの非音声モーダルによるコミュニケーションを避けるため、相手の顔が見えないようにしてレシーバから相手の音声が聞こえる状況で同じ部屋で会話した。二人の対話の場合、相手が話し終わらないうちに話し出すという重畳音声が避けられない。2チャンネルに別々に録音し、計算機処理の際に、タイムスタンプの時刻をつけておくことで、重畳音声も扱うことができる。

3 文字表記とタグ付け

データベースをもとに、音声的あるいは言語的な統計処理を行うには、データに必要な範囲のタグ情報を付与することが必要である。音声データのラベル表現には、音声セグメントラベル、韻律情報、言語情報などがある。これらの付加情報を、品質をチェックして表記するには、人手による大量な作業を必要とする。データ量の規模の拡大が求められることを考えると、必要最小限の付加情報をできるだけ自動的なタギング手法を用いて付与していくことが必要になる。

3.1 書き起こしテキストの表記

(1) 文字表記

収録された音声データの内容を記述するもので、かな漢字による表記とローマ字およびかな表記を行っている。ローマ字は、日本以外の国でこのデータベースを利用する際の便宜を考えている。ローマ字およびかなによるテキストの表記では、形態素解析でなされる区切りを参考にして文節に近い区分で区切る。ローマ字表記には、電子協による音声データベース表記ガイドラインに準拠した。

(2) ノイズの表記

書き起こしには、人によるノイズと雑音などの種々のノイズに相当する音声以外の音も表記する。人によるノイズには、咳、吸息、吐息、笑い声、口で鳴る音などがある。雑音には、キーのクリック音、ペンでたたく音、ドアの音などがある。また、会話の途切れの沈黙、言い直し、言い間違いもそれぞれマークをつける。会話の重なりが生じた箇所では、重なり始めた音と終りの音に記号を付加する。あいづちの場合はあいづちの入った場所に記号を付加する。ノイズ表記のリストを表1に示す。

表1: ノイズ表記のリスト

表記	ノイズ種別	出現回数
/h#/	吸息、吐息	248
/ls/	口で鳴る音	26
/lg/	笑い声	25
/cg/	咳	7
/z#/	鼻水をすする音	5
#pencil-#	鉛筆が転がる音	26
#pencil-use#	紙に鉛筆で何か書く音	0
#paper-#	紙のめくれる音	34
#paper-tap#	紙束で机をたたく音	2
#chair-#	椅子のきしむ音	0
#key-tap#	キーボードをたたく音	8
#pi#	ビープ音	16
#kaiwa#	A、B以外の会話	15
#door#	ドアの開く音	1
##	不定	5

3.2 形態素解析

かな漢字による表記をもとに、音声認識のための言語モデルを作成するには、形態素解析 [9] が有用である。日本語テキストの形態素解析には、大きな辞書と計算機による言語処理に適した文法が必要である。形態素解析のための辞書は、大辞林 [10] をベースとした。品詞の種類は 27 種類とした。すなわち、固有名詞、サ変名詞、形容名詞、普通名詞、数詞、代名詞、本動詞、補助動詞、形容詞、副詞、連体詞、接続詞、感動詞、間投詞、助動詞、格助詞、準体助詞、係助詞、副助詞、並立助詞、接続助詞、終助詞、連体助詞、引用助詞、接頭語、接尾語、その他である。

音声対話データのタグ付けのための辞書定義は、まだ定着しているとはいえない。話し言葉においては、普遍的な言語の定義が難しい。特に表現がゆれたり乱れたりするため、文法的な情報のみでは、形態素種別を人間でも安定に判定するのは必ずしも容易ではない。意味情報も加味していく必要がある。

音声対話データベースをもとにして、形態素カテゴリーの統計的性質を言語モデルとして利用することを考えると、形態素カテゴリーの種類はあまり少ないと統計量が有効に働かない。また、あまり細かすぎると精度や安定性が保てなくなる。したがって、形態素解析プログラムの品詞体系カテゴリーを適度に細かくした分類が適当と思われる。これについての検討は後述する。

なお、話し言葉の会話の音声認識や言語処理をする上で、助詞や助動詞の品詞カテゴリーを、学校文法のカテゴリー分けよりより長いもの（連接されたもの）をとるほうが好ましいという点もある。

例: でしょ（助動詞）う（助動詞）
−> でしょう（助動詞）

4 自由発話音声データの言語的性質

スケジューリングタスクにおける自由発話音声対話について、自由発話の言語モデルを作成するための情報を得るために言語的性質を分析してみた [11] [12]。13 対話のデータについての分析である。データの量が少ないため、まだ、定性的な考察が含まれている。定量的分析についてはデータ量を増大して行っていく。

4.1 会話の文体

会話の文体は、会話の相手が誰であるかによって大きく変化する。たとえば、親しい友人との会話の場合、発音が曖昧になったり、文法がいい加減になったりする。一方、目上の人物や初対面の人と話す場合は、発音、文法ともに比較的正確になる。収集した対話データの中から例を示す。（図 2, 図 3）

A：あーね。行こ行こ言って全然行ってないよ。
B：そうそうそう。
A：うーんうーん。
B：なかなか。だからどうせ　火、金なんて、学校きてるでしょ。

図 2: 友人との会話

B：#h-#　水曜日はセミナーがはいってまして。
A：あー、そうなんですか。#pencil-#
B：午後の 4 時まで空いてませんので #h-#。#pencil-#
A：あ、じゃー無理ですね。
B：この日もちょっと、うーん。
（#pencil-#: 鉛筆が転がる音、#h-#: 息を吸う音、吐く音）

図 3: 秘書どうしという設定での会話

4.2 分析結果

13対話で発話数は全部で1019である。ここで、発話とは、同じ話者が続けて話した内容とした。（発話数よりも文の数の方が多い。）1発話あたりの平均単語数は、約7単語である。

4.2.1 品詞の出現頻度情報

スケジューリング対話における主な統計量を表2に、品詞出現率を図4に、品詞ごとの異なり単語数の割合を図5に示す。

表2: スケジューリング対話データの主な統計量

	総数	異なり数
対話	13	
発話	1019	
単語	6716	750
単語(記号含む)	8624	754
間投詞	767	38
普通名詞	943	206
本動詞	446	117

また、対話の進行に伴う単語異なり数・累積単語出現数の増加の様子を図6に示す。図中の縦線は、対話の切れ目である。

文頭における品詞の上位5位は、感動詞、普通名詞、副詞、接続助詞、数詞である。また、文末に来る品詞の上位5位は、感動詞、終助詞、接続助詞、助動詞、格助詞である。話し言葉に特有な傾向である。文頭、文末に来る品詞の中で出現頻度が高い上位5位で、出現単語数は全体の12%である。また、これら上位5位までで全発話数の約80%の文頭、文末を占めている。

品詞の出現傾向は、10%を超える品詞は、普通名詞、助動詞、格助詞、間投詞の順である。ついで、5%～10%のものが本動詞、接続助詞、接尾語の順である。そのほかの品詞はそれぞれ数%以下である。自由発話であるので、間投詞や接続助詞が多いこと、動詞が比較的少ないこと、助詞や助動詞が多いことがあげられる。

日本語における主要な助詞「は、が、の、を、に」の出現は、総計713で、これらが出現単語数に占める割合は約11%である。その内訳は、「は(25.6%), が(20.1%), の(41.9%), を(4.21%), に(7.99%)」である。

文を構成する主な要素は、品詞出現率から、名詞、助動詞、格助詞であることがわかる。

4.2.2 話し言葉特有な語、表現

文末が完結していないで、名詞や、途中でとまっているものがある。また、現在の若い人の間で話される話し言葉として、短縮語や、本人同士がわかるローカルな用語があらわれる。これらは、未定義語となる。間投詞や感動詞や副詞の使い方には、個的な癖がある。

4.2.3 間投詞に関する分析

間投詞は、一般に文のはじめにあって、感動、呼びかけ、応答などの意を表す。（なお、ここでは、感動詞も間投詞に含む。）間投詞の出現割合は、多い対話で12.5 %、少ない対話で5.02 %、平均8.89 %であった。これは個人差によるところが大きい。しかし、間投詞の異なり語数は38個に収れんしている。

間投詞をポーズとの位置関係によって分類した結果を、表3に示す。（ここではテキスト中での句読点をポーズとした。） 多くの場合、間投詞はポーズを前後にはさみ、独立して用いられている。実際の出現例として、多い順に挙げると、「はい(36.9%), えー(13.0%), あ(12.9%), えーと(9.91%), え(5.22%), あー(2.87%), あのー(2.35%),...」となる。間投詞・ポーズに関しては詳しい研究[13]がある。

表3: 間投詞とポーズとの位置関係

前後にポーズがつくもの	91.9%
前にのみポーズがつくもの	5.35%
後ろにのみポーズがつくもの	2.48%
前後にもポーズがつかないもの	0.26%

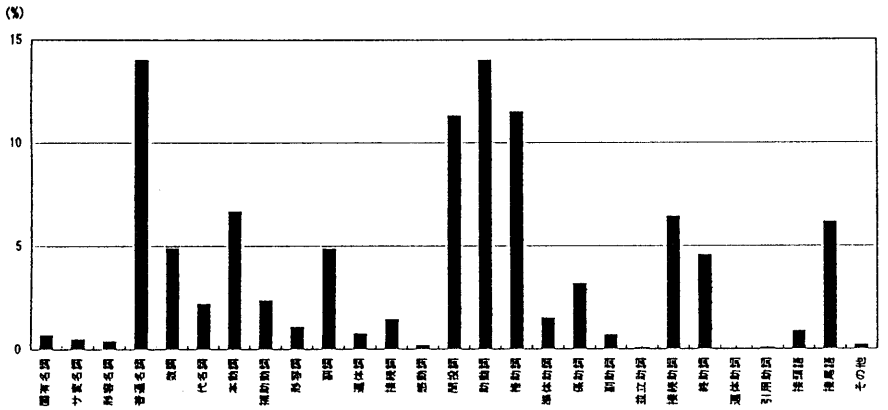

図 4: スケジューリング対話における品詞出現率（総数:6716）

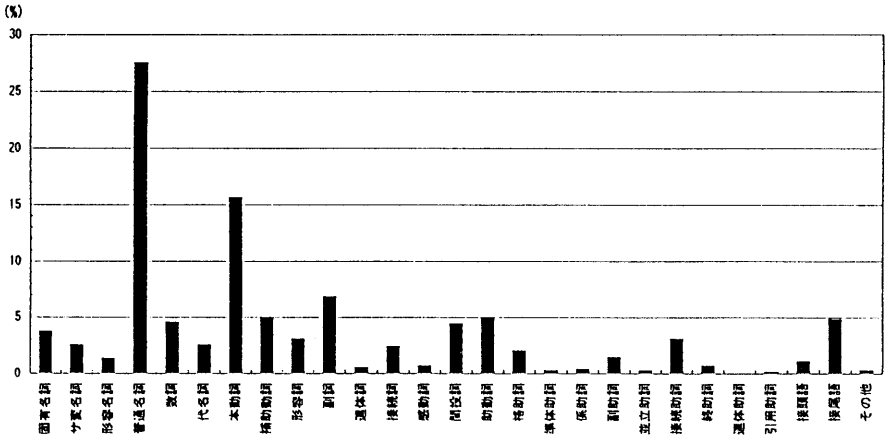

図 5: スケジューリング対話における品詞ごとの異なり単語数の割合（総数:750）

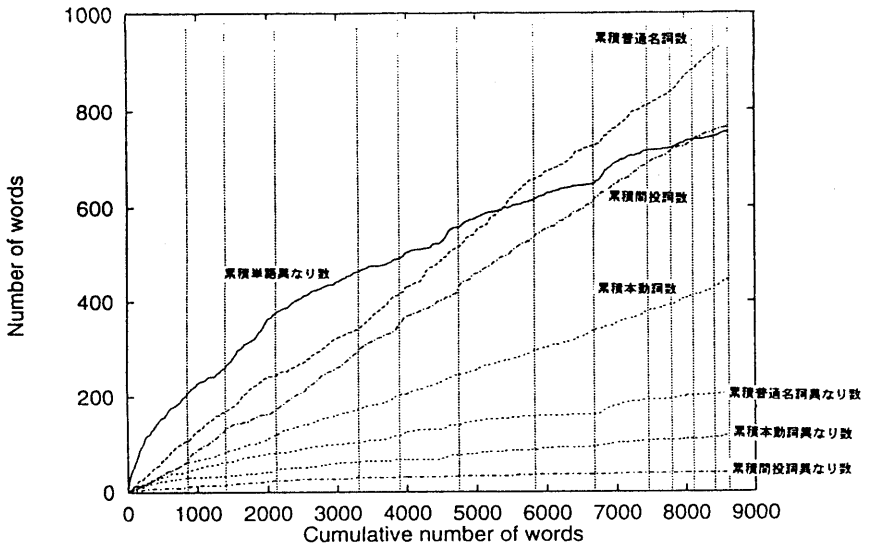

図 6: スケジューリング対話の累積語数

4.2.4 単語の細分類についての検討

音声認識における次単語の予測に、単語 bigram（二つ組）の利用が考えられるが、bigram を得た学習用テキスト以外の入力には対処できない。したがって、制約を緩くした品詞 bigram の利用を考えるが、1つの品詞カテゴリに多くの語が含まれると、予測の効率が悪くなる。これに対処するため、品詞の細分類について検討するが、ここでは普通名詞、本動詞、間投詞・感動詞について、相互情報量をもとにしたクラスタリングを試みた [14]。相互情報量は式 (1) で算出される。

$$I(w_1;w_2) = \log \frac{P(w_1w_2)}{P(w_1)P(w_2)} \tag{1}$$

ただし、$P(w_1w_2)$ は単語 w_1 と単語 w_2 が連続する場合の出現確率、$P(w_1)$、$P(w_2)$ は、それぞれ w_1、w_2 の独立の出現確率である。この場合、相互情報量が大きいものほど、同時に出現しやすいことがいえる。

クラスタリングの方法としては、まず最初に1クラスタに1単語を対応させ、任意の2クラスタをまとめた場合の平均相互情報量を計算する。そして、平均相互情報量の減少が最も少ない2つのクラスタを結合するという過程を繰り返すというものである。ここでいう平均相互情報量とは、関係するすべての単語ペアについて相互情報量を計算し、平均をとったものとする。なお、その単語がクラスタリングされている場合、クラスタに含まれるすべての単語について評価を行ない、平均をとる。

以上のようにして普通名詞 206 個、本動詞 117 個をそれぞれ 20 個のクラスタに、間投詞・感動詞 38 個を 5 個のクラスタに分類した。普通名詞に関しては、分類語彙表 [15] による分類も行なった。

評価として、情報理論的な意味での平均分岐数であるパープレキシティ [16] を式 (2),(3) として算出する。ここでは単語 bigram の出現確率まで考慮する。

$$F_p(L) = 2^{H(L)} \tag{2}$$

$$\begin{aligned} H(L) &= -\sum_{w_1,w_2} P(w_1)P(w_2|w_1)\log_2 P(w_2|w_1) \\ &= -\sum_{w_1,w_2} P(w_1w_2)\log_2 P(w_1w_2) \\ &\quad + \sum_{w} P(w_1)\log_2 P(w_1) \end{aligned} \tag{3}$$

ただし、$P(w_2|w_1)$ は、w_1 の次に w_2 がくる確率であり、$P(w_1w_2)/P(w_1)$ に等しいとする。

その結果、表 4のように、品詞を細分類しない場合よりもパープレキシティが減少していることがわかる。また、相互情報量のみによる分類では、データの量が十分ではないので、精度的に問題があるようだが、分類語彙表を用いた場合とあまり差が出なかった。つまり、分割の仕方よりも分割の数に依存するところか大きいものと思われるが、これは、さらなる検討を要する。また、自由対話のパープレキシティは大きい値をとることがわかった。

表 4: スケジューリング対話における bigram 比較表

	750 単語 + (pause)	27 品詞 + (pause)	品詞細分類 (相互情報量)	品詞細分類 (+ 分類語彙表)
品詞 bigram 数	–	252	679	682
単語 bigram 数	2,577	386,124	143,554	147,112
パープレキシティ	8.09	144	91.4	96.0

5 まとめ

本稿では、音声対話データベースとしてスケジューリング対話をとりあげ、その言語的性質として、特に品詞の頻度情報について分析を行なった結果を述べた。また、次単語予測の効率化のために、相互情報量をもとにした品詞の細分類を試みた。perplexityの減少により、音声認識の言語モデルとして品詞の細分類は有用であるといえる。今後の課題は、音声自由会話のデータ量をさらに拡大し、十分な統計量を得て、自由発話対話システム実現のための言語モデルを構築することである。

謝辞

この研究は文部省重点領域研究「音声対話」の一環として行なわれたものである。本稿をまとめるにあたり、有益なご助言を頂いた電気通信大学情報工学科高木一幸助手に感謝の意を表します。また、本研究室の皆様にも感謝いたします。

参考文献

[1] 江原, 井ノ上, ほか,"ATR対話データベースの内容", ATRテクニカルレポート, No.T-I-186, (1990)

[2] 板橋秀一,"文部省「重点領域研究」による言語データベース", 日本音響学会誌, Vol.48, No.12, pp.894-898, (1992)

[3] 小林, 板橋, 速水, 竹沢,"日本音響学会研究用連続音声データベース", 日本音響学会誌、 Vol.48, No.12, pp.888-893, (1992)

[4] 匂坂芳典, 浦谷則好,"ＡＴＲ音声、言語データベース", 日本音響学会誌,Vol.48 , No.12, pp.878-882, (1992)

[5] 竹沢 寿幸, 田代 敏久, 森元 逞, "自然発話の言語現象と音声認識用日本語文法", 情処研報 95-SLP6-5 (1995-05)

[6] 東 郁雄, 荒木 雅弘, 堂下 修司, "音声対話データの収録と意味的情報の統計的分析", AI全大 20-02 (1995-07)

[7] 森川 恵美, 中里 収, 田中 修一, 白井 克彦, "協調問題解決における対話モデル", AI全大 19-01 (1995-07)

[8] 森元, 村上,"音声対話における言語現象", 日本音響学会誌,Vol.50,No.7, pp.55 8-562, (1994)

[9] Noriyoshi URATANI, Toshihisa TASHIRO, Hi-sako YAMADA,Kaori MATSUMOTO, "Users Manual for Japanese Morphological analysis in the ATR Spoken Language Database", TR-IT-0009,1993.

[10] 大辞林, 三省堂, 1985.

[11] 山本 英明, 樽松 明, "音声自由会話の特徴分析", 音講論 3-P-17(1995-03)

[12] 中筋 知己, 樽松 明, "スケジューリングタスクにおける自由発話音声の言語的性質", 音講論 1-Q-28(1995-09)

[13] 中川 聖一, 小林 聡,"自然な音声対話における間投詞・ポーズ・言い直しの出現パターンと音響的性質", 日本音響学会誌, Vol.51,No.3, pp.202-210, (1995)

[14] Peter F. Brown et al. (IBM T.J. Watson Reserch Center), "Class-Based n-gram Models of Natural Language", Computational Linguistics 18-4 (1992)

[15] 国立国語研究所：分類語彙表, 秀英出版, 1964

[16] 中川 聖一,"確率モデルによる音声認識", 電子情報通信学会, 1988.