

音声言語処理と 自然言語処理(増補)

中川 聖一 編著

小林 聡	峯松 信明
宇津呂武仁	秋葉 友良
北岡 教英	山本 幹雄
甲斐 充彦	山本 一公
土屋 雅稔	

共著

コ ロ ナ 社

ま え が き

本書では音声言語と文字言語を対象とし、工学的応用（音声認識、音声合成、機械翻訳、検索など）を目的とした基礎技術について解説している。なお、一般には文字言語は自然言語と呼ばれており、本書もこれになった。

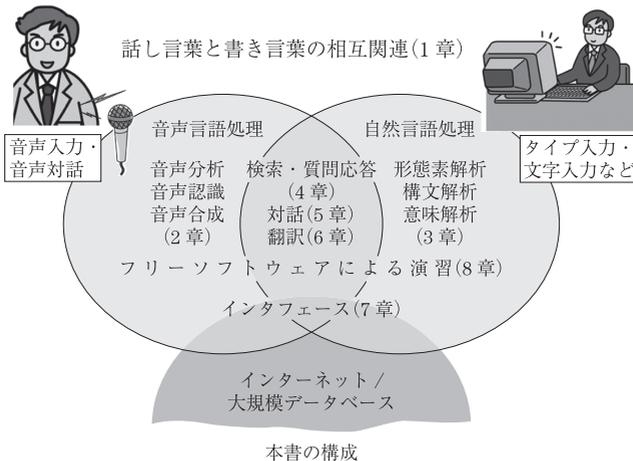
音声とは、送信者から受信者へ意味のある意思伝達を声を媒介として行うときの声そのものである。これから、動物の発声する声はわれわれ人間にとっては単なる音、声にすぎないが、動物にとっては立派な音声であるといえる。また、言語とは何かを媒介として意味のある意思伝達を行う行為、行うための規則である。なお、辞書的な意味では、「音」 耳で聞くもの、「声」 発声器官を使って口から出る「音」、「言（こと、ことば）」 「声、文字」などを使って意図を伝えるものと定義される。音声を媒介とするものを音声言語、文字を媒介とするものを文字言語という。また、定義から視覚による言語も考えられる（身振りによる意思伝達、画像による意思伝達。例えば読唇、手話、象形文字）。さらには、プログラミング言語をはじめとする人工言語もある。

人間の歴史過程からいえば、音声言語は10～20万年前に自然発生的に発明された（解剖学的知見から推定、チンパンジーはいくら学習しても人間の音声は発声できない）。音声言語は、効率性と冗長性を反映した誤り訂正符号で、出現頻度の高い単語ほど発声時間が短いことなど情報理論的考察と符合する。これと比べれば、文字言語の発生（発明）はずっと新しい。文字言語の歴史は5000年程度、われわれ日本語に限れば1500年程度にすぎず、人類の数百万年以上の歴史と比べて微々たる期間である。しかし、文字言語の発明によって人類は急速な進歩を遂げた。これは、文字によって音声言語をシンボル化することができるようになり、それを記録として残すことができ、知識の伝承と正確な意思伝達、社会秩序の維持が可能になったためだと思われる。このことから、音

声言語を文字言語に自動変換できれば、計り知れない効用があることが想像できる。

以上からも明らかのように、音声言語と自然言語には密接な関連があることがわかっていただけだと思う。最近、とみに話し言葉に近いブログの解析や音声翻訳、音声によるウェブ検索など、音声言語処理と自然言語処理の統合分野での実用化が目立ってきた。今まで以上に、両分野に精通した技術者が望まれている。

本書は、音声言語処理と自然言語処理を有機的に関連付けたわが国で最初の大学の教科書である。学部でこの分野の講義がない場合は、大学院の教科書としても十分使用できる内容である。ただし、音声と言語の両分野を2単位分で学ぶのは量が多すぎるので、1章～4章と8章の演習が中心となる。音声言語処理に重点を置いた講義では、1章、2章、5章、7章および8章の演習を中心に、自然言語処理に重点を置いた講義では、1章、3章、4章、6章および8章の演習を中心に学ぶことをお薦めする。この分野に興味をもたれる方が少しでも増えれば著者らの望外の喜びである。



2013年1月

中川聖一

増補版に寄せて

最近のビッグデータと深層ニューラルネットワークに基づく AI ブームは、画像処理や音声処理、機械翻訳、将棋・囲碁のゲーム探索などにおいて画期的な成功をもたらし、今回のブームは本物と言われている。しかし、これに至るまでにニューラルネットワークのブームと期待外れは過去何回もあった。

第 1 次ブームは、1960 年代から 1970 年にかけて、Rosenblatt のパーセプトロンの学習時代で、機械学習の理論的研究と小規模のパターン認識の研究が数多くなされた（第 1 次 AI ブームはこれより 5 年ほど早い）。この時代は、人間の脳機能に模したモデルで、何がどこまでできるか知的好奇心を満たす研究が主であり、小規模問題への適用段階であった。第 2 次ブームは、1980 年代から 1990 年代にかけて、Rumelhart の誤差逆伝搬学習の時代で、脳モデルと認知心理学現象との対応や実世界のパターン認識等への応用研究が盛んに研究された。しかし、同時期に進行していた確率・統計手法に優る成果が得られず、ブームは自然に去った。2010 年代から現在まで続いている第 3 次ブームは、1980 年代の千倍 ~1 万倍以上のコンピュータ処理能力の大幅な向上と大量の学習データ、大規模（深層）ネットワークの学習アルゴリズムの改善により、1980 年代に証明されていたニューラルネットワークの潜在的能力を大規模のネットワークで引き出せるようになり、大きな発展を遂げている。

本書の第一版の執筆に取り掛かったのは、2011 年であり（発刊は 2013 年）、当時は確率・統計的手法が主流であり、深層ニューラルネットワークはそれほど成果が上がっておらず、本書の内容には含めなかった。ところが、最近の著しい発展・成果を見るにつけ、教科書とは言え、深層ニューラルネットワークによる音声処理技術・自然言語処理技術を無視するには行かなくなってきた。そこで、音声認識と音声合成、分散表現等による自然言語処理、機械翻訳に焦点

を当てて、ニューラルネットワーク技術を必要最小限であるが追加・解説することにした。

ニューラルネットワークの基本構造である順伝搬型のネットワークを多層にした深層ニューラルネットワークの構造と学習方法、および他の代表的な構造である再帰型ニューラルネットワークと畳み込みニューラルネットワーク、応用範囲の広い自己符号化器は、まとめて2章で解説している。このほか2章では、ニューラルネットワークの音声認識、言語モデル、音声合成への適用方法について解説している。3章では、ニューラルネットワークによる単語の分散表現、文の構文解析と意味解析に相当する依存構造解析、文の分類問題への適用方法について解説している。近年、単語の分散表現は、文の分散表現や文章（パラグラフ）の分散表現に拡張され、自然言語処理の基本技術になりつつある。6章では、ニューラルネットワークによる機械翻訳技術を解説している。これは系列-系列変換を基本とする手法で、記号の系列処理に関しても、ニューラルネットワークの能力の高さを示すものである。紙数の関係で本書では解説できなかったが、4章の検索・応答システム、5章の対話システム、7章の入力インタフェースにおいてもニューラルネットワーク技術の導入が進んでいる（7.3.2項に最近応用展開が目覚ましいスマートスピーカーの話題を追加している）。いずれも基本的には2章、3章、および6章で解説した手法に基づいている。8章のフリーソフトウェアによる演習においては、ニューラルネットワークによる音声認識、機械翻訳（分散表現含む）の演習を追加した。

旧版で詳述した音声の発生・生成現象や認識・合成メカニズム、自然言語の解析・理解メカニズムを確率・統計的にモデル化する従来の技術をよく理解した上で、ややもするとブラックボックス化したニューラルネットワーク技術を学ぶことにより、本書が音声言語処理や自然言語処理は勿論、他の多くの情報処理の技術開発の礎になれば幸いである。

2018年7月

目 次

1. 音声と言語の諸相

1.1 音声科学と音声工学	1
1.2 言語科学と言語工学	3
1.3 音声学，音韻論と言語学	5
1.3.1 音声学，音韻論	5
1.3.2 言 語 学	7
1.4 話し言葉と書き言葉	10
1.4.1 話し言葉の特徴	10
1.4.2 書き言葉の特徴	12
1.5 言 語 の 獲 得	13
1.5.1 音韻，音韻の構造，音韻体系の獲得	13
1.5.2 文法の獲得と第 2 言語の学習	14
章 末 問 題	15

2. 音声言語処理のモデル

2.1 音声の音響的分析とそのモデル	16
2.1.1 音声生成のメカニズムとそのモデル	16
2.1.2 音声に含まれる情報とその音響的対応物	17
2.1.3 音声の記号化	22
2.1.4 音声の音響的分析	23

2.1.5	時間構造の異なる2つの特徴ベクトル系列の対応付け	31
2.2	音声の認識とそのモデル	33
2.2.1	音声認識の難しさ	33
2.2.2	音声認識問題の数理統計的な定式化	36
2.2.3	音響モデル	37
2.2.4	言語モデル	47
2.2.5	仮説探索(デコーディング)	51
2.3	音声の合成とそのモデル	54
2.3.1	テキスト音声合成の難しさ	54
2.3.2	音韻処理	56
2.3.3	韻律処理	58
2.3.4	HMM音声合成方式による波形生成	59
2.4	深層ニューラルネットワークに基づく音声認識と音声合成	65
2.4.1	多層化されたニューラルネットワークとその音声処理への応用	66
2.4.2	誤差逆伝搬法と自己符号化器を使った事前学習	69
2.4.3	GMM-HMMからDNN-HMMへ	71
2.4.4	さまざまなネットワーク構造	73
2.4.5	DNN/RNN/LSTMを用いた言語モデル	76
2.4.6	DNN/RMM/LSTMを用いた音声合成	78
	章末問題	81

3. 自然言語処理のモデル

3.1	形態素解析	85
3.1.1	形態素解析の枠組み	85
3.1.2	統計的モデルに基づく形態素解析	93
3.1.3	仮名漢字変換	94

3.2 構文解析	97
3.2.1 句構造解析	97
3.2.2 係り受け解析	109
3.3 意味解析	113
3.3.1 意味素とシソーラス	113
3.3.2 格解析	115
3.3.3 語義曖昧性解消	119
3.3.4 語彙知識の獲得	122
3.4 文脈解析	127
3.4.1 照応解析	127
3.4.2 修辞構造解析	131
3.5 ニューラルネットワークによる自然言語処理	132
3.5.1 単語の分散表現	133
3.5.2 依存構造解析	135
3.5.3 文の分類	137
章末問題	141

4. 検索・質問応答システム

4.1 文字列照合	144
4.1.1 完全一致文字列照合のオンライン手法	145
4.1.2 近似文字列照合のオンライン手法	148
4.1.3 文字列照合のオフライン手法	151
4.1.4 近似文字列照合と索引付け	156
4.2 文書検索	157
4.2.1 文書のベクトル表現	158
4.2.2 ベクトル空間モデル	159

4.2.3 確率的言語モデルによる文書検索	162
4.3 質問応答	166
4.4 音声と情報検索	170
4.4.1 音声ドキュメント検索の問題設定	171
4.4.2 音声ドキュメント検索の課題と手法	172
章末問題	174

5. 対話システム

5.1 談話と対話	175
5.1.1 談話とは	175
5.1.2 対話と会話	176
5.1.3 対話の公準	176
5.1.4 談話, 対話の構造	177
5.1.5 対話行為	178
5.2 対話システム	179
5.2.1 対話システムとは	179
5.2.2 対話システムの構成	181
5.2.3 対話の主導権	182
5.3 対話制御	184
5.3.1 対話制御とは	184
5.3.2 有限状態オートマトンによる状態表現を用いた対話制御	185
5.3.3 意味表現に基づいた応答生成による対話制御 — ケーススタ ディ—	187
5.3.4 POMDP による対話制御	191
5.4 マルチモーダル対話	195
5.4.1 マルチモーダルな状態	195

5.4.2 マルチモーダル対話システム	196
章 末 問 題	199

6. 翻訳システム

6.1 機械翻訳の歴史と代表的なアプローチ	200
6.2 規則に基づく機械翻訳	206
6.2.1 規則に基づく機械翻訳の概要	206
6.2.2 単 語 変 換	209
6.2.3 構 造 変 換	210
6.2.4 規則に基づく手法の問題点	211
6.3 コーパスに基づく機械翻訳と統計的機械翻訳	214
6.3.1 単語単位の統計的機械翻訳	214
6.3.2 フレーズ単位の統計的機械翻訳	217
6.4 ニューラル機械翻訳	220
6.4.1 系列変換モデルによる機械翻訳	220
6.4.2 注 意 機 構	223
6.5 音 声 翻 訳	224
6.5.1 テキスト機械翻訳と音声機械翻訳	225
6.5.2 音声認識結果の整形	226
6.5.3 統計的機械翻訳を用いた音声翻訳	227
6.5.4 ニューラル機械翻訳による音声翻訳	228
章 末 問 題	230

7. テキスト，音声入力インタフェース

7.1 ヒューマンインタフェース	231
------------------------	-----

7.2	テキスト入力インタフェース	234
7.3	音声入力インタフェース	239
7.3.1	テキスト入力的手段としての音声インタフェース	240
7.3.2	意図・情報伝達的手段としての音声インタフェース	242
7.4	マルチモーダル入力インタフェース	249
	章末問題	251

8. フリーソフトウェアによる演習

8.1	音声分析, ラベリング	252
8.2	音声認識	253
8.3	音声合成	255
8.4	形態素解析	256
8.5	係り受け解析	258
8.6	全文検索	258
8.7	統計的機械翻訳	259
8.8	深層学習フレームワーク	259
	演習課題	261
	引用・参考文献	262
	章末問題解答	271
	索引	283

1

音声と言語の諸相

本章では、本書のキーワードである「音声学」、「言語学」、「音声言語」、「自然言語」、「話し言葉」、「書き言葉」、「文法、意味、談話」、「言語獲得」、および応用分野などについて全体像と相互関係を述べ、2章以降の理解の準備とする。

1.1 音声科学と音声学^{1)†}

音声というものが、その発生および受理に関して人間が主体であり、生理的な機構を介して行われるという事実によることから、これらのメカニズムを解明することは音声科学と呼ばれている。

音声の基本単位である音素、音節、モーラなどの知覚・生成能力となると、音声生得説は揺るぎない事実であろう。例えば、乳幼児を対象とした研究で、6ヶ月の幼児にも母音や子音の対比、ピッチの違い、話者の違い、音声学的な構造の違いを聞き分ける能力があることがわかっている。これから、生まれながら言語学的対立を聞き分ける能力があるのか、言語と無関係な聴覚機構の特性によって識別が行われているのか定かではないが、いずれにしても、先天的に音声言語の知覚・識別機構が備わっていると考えられる。もちろん、高機能な知覚メカニズムはニューラルネットワークが学習的に形成（自己組織化）されることにより、後天的に獲得されると考えられる。最近では、ニューラルネットワークの一部は可塑的で（外部の入力より構造が変化する）、残りは遺伝子によって先天的に決められているとする説が有力である。

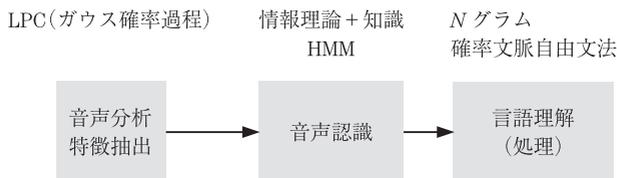
[†] 肩付き数字は、巻末の引用・参考文献の番号を表す。

2 1. 音声と言語の諸相

一方、音声の本質を科学的に解明する必要性はあるが、工学的なモデル化による構成的解明や音声の工学的な応用も価値がある。例えば、われわれ日本人は英語子音の [l] と [r] の区別ができないことや英語の多種類の母音の発音が苦手なことなど、言語に依存する生後の学習による知識獲得も存在し、ここに工学的モデルの必要性、有用性が出てくるように思われ、音声の分析、合成（生成）、認識（知覚）に優れた工学的モデルが提案されている（2章）。また、代表的な応用分野として音声の狭帯域通信（音声のデータ圧縮）、音声合成（speech synthesis）、音声認識（speech recognition）、話者認識（speaker recognition）、音声対話（5章）がある。これらはいずれも音声分析を基礎技術としており、これらを総称して音声工学と呼ぶ。もちろん、音声合成は音声生成機構、音声認識は音声知覚機構が解明できれば実現できるが、その道は険しい。そこで、これらの知見を参考にしながら工学的なモデルに基づいて実現を図ることになる。音声工学の総合的技術が音声翻訳（speech translation）であろう。音声翻訳の実現には、音声認識・合成研究以外に機械翻訳（machine translation, 6章）や対話理解（5章）という自然言語処理の研究が必要である。

優れた工学的モデルは優れた認知（知覚）モデルに発展したり、（聴覚）生理モデルに示唆を与えるであろう。人間の脳から学ぶことと脳を模倣することとは別である。このことから音声言語処理を工学的に実現するモデルとして脳が

■工学的アプローチモデル



■科学的アプローチモデル

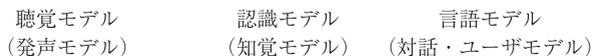


図 1.1 音声言語処理システム²⁾

最適だとは必ずしもいえない。しかし、その基本原理である並列分散処理に学ぶべき点が多い。図 1.1 に音声言語処理システムの工学的アプローチモデルと科学的アプローチモデルを示す（2 章では工学的モデルを解説している²⁾）。

1.2 言語科学と言語工学

音声に科学的側面と工学的側面があるように、言語にも科学的側面と工学的側面があり、それぞれ言語科学と言語工学と呼ばれている。例えば、日本語の文法についていえば、日本語を内省的に分析し、文生成や文構造を体系的に記述する方法を見いだそうとするのが言語科学である。科学的アプローチは、典型的な文や特異な文を例に取り、その生成や構造を説明し、その仮説を認知科学や脳科学の知見で検証しようとする。一方、工学的アプローチは、特異な文はさておき、より多くの典型的な文を記述できる文法を求めようとするもので、文法の良し悪しの評価基準が明確である。図 1.2 に両アプローチの違いを示す。

■工学的アプローチモデル

句構造文法	コーパスベース	部分観測マルコフ
N グラム	格フレーム	決定過程
係り受け文法		



■科学的アプローチモデル

生成文法	意味ネットワーク	談話表示理論
認知文法	非単調論理	修辞構造理論

図 1.2 自然言語処理システム

人間は、無から言語を獲得しているのではない。チョムスキー（Chomsky）による生成文法（generative grammar）以前の構造言語学では、乳幼児の時期に周囲で交わされる、あるいは自分かけられる音声を覚え、それによって言語を使えるようになって考えられていた。しかし、この仮説では子供が経験す

る音声言語に比べ、後に子供が用いる音声言語が複雑であることの説明ができない。しかし、チョムスキーは言語獲得装置が脳内に生得的に存在することによって、言語の獲得が可能となると考えた^{3), 4)}。言語獲得装置は、獲得可能な文法のタイプを定めた普遍文法 (universal grammar) と、普遍文法と経験の相互作用のあり方を定めた言語獲得原理からなる。特に、普遍文法については、原理とパラメータによって表現されていると考える。これにより、聞いたことがない文を生成したり、あるいは聞いたことがない文が文法的に正しいか否かの判断ができることなどの説明が可能となる。普遍文法のパラメータの獲得は、生得的に備わっている音韻知覚能力が、学習により言語に依存した音韻体系を獲得していくのと似ている。

一方、工学的なアプローチでは、生得的に備わっているとされる文法も工学的に大規模データから獲得しようとする。普遍文法からパラメータを獲得することにより、言語現象の 95 % をきれいに記述、説明できるとしても満足せずに、期待値として観測現象の 99 % を処理できる手法を目指すものである。最も端的な例が機械翻訳のアプローチであろう (6 章)。言語科学的なアプローチは、規則に基づいて得られたソース言語を解析し、部分的な構造木をターゲット言語の構造木へ変換ルールを用いて変換し、部分的な構造木列を、ルールで並べ替えて文に構成し、変換ルールを用いて表層形に変換する手法だと考えられる (6 章)。このときに用いる知識は上述の各種変換ルール以外に、単語間の意味関係、単語の内包的・外延の意味などがある。機械翻訳システムは、これを工学的に実現しようとしてきたが、知識ボトルネックに陥り、限界に達している。最近では、音声認識手法と同じく、大規模データ (例えば、翻訳例文を 100 万ペア使用) に基づく統計的・確率的手法で翻訳するのが主流となっている (6 章)。このようなアプローチは、到底人間が実行している手法 (科学的アプローチ) とは考えられないが、従来法よりもよい翻訳性能が得られつつある。

言語工学の応用分野としては、仮名漢字変換 (3 章)、テキストマイニング (3 章)、質問応答システム (4 章)、検索 (4 章)、要約、機械翻訳 (6 章)、対話インタフェース (5 章, 7 章)、語学学習などがある。

1.3 音声学，音韻論と言語学

1.3.1 音声学，音韻論

音声学における音とは、人間が言語音として用いるものを、物理的（または調音の）性質から網羅的に分類したものである。これに対して音素（phoneme）は、特定の言語の話者が弁別する必要があるか否かによって分類したものである。ひとつ以上の音素が結合し音節（syllable）を形づくる。音節を構成する音素、あるいは音節の並びの条件もまた言語によって違いがあるが、基本的には中心となる母音の前後それぞれに0個以上の子音（consonant）が結合したものである。日本語の場合、母音（vowel）をVと書けば、ア行音は/V/となる。カ行以降の音節は、子音をCと書けば/CV/となる。また、/キャ/などは、拗音をyと書けば、/CyV/となる。/ン/(/N/)は、先行音節と結び付いて1音節を形成する。

なお、日本語には/アッサリ/と/アサリ/のように、音素の連鎖の中の無音（促音）の有無によって、異なる単語となるものがある。促音は、物理的には無音であるが、単語の区別に寄与している。無音であるため、音声学的な音でも音韻論的な音素でもない。同様に、日本語には長音も存在する。長音は、音声学的な音としても、音韻論的な音素としても、継続長の違いによって区別されるものである。例えば、/フクロオ/(鼻)と/フクロ/(袋)の区別は、最後の/o/という音素の継続長の違いによる。促音と長音は音声学的な音や音韻論的な音素、およびそれらに基づく音節では表記できないものの、日本語においては単語の弁別の役割を担っている。そこで、日本語の分析においては、通常の音素に加えて、促音、長音も含めてモーラ（拍）という単位が用いられることがある。また、先に/ン/は先行音節と結び付いて1音節を形成すると述べたが、モーラを単位とする場合には/ン/も独立したモーラとする。このようなモーラとは、大まかに日本語話者にとって1音と感じられる単位である。例えば、/to o kyo o/(東京)は、/to o/および/kyo o/の2音節語であるが、モーラで数えれば、/to/、

索引

【あ】	後ろ向き確率	41	音声対話システム	181	
曖昧性	102	後ろ向き中心	130	音声知覚機構	2
アクセント	6, 55	埋め込み	77	音声中の検索語検出	172
アクセント核	57	埋込み文	12	音声ドキュメント検索	171
アクセント句	58	【え】		音声認識	2, 171, 225, 227
アクセント結合	55	衛星	131	音声翻訳	2, 200, 206, 224
誤り率最小化基準	54	枝刈り	51	音声理解タスク	239
【い】		遠隔音声認識	244, 246	音節	5
言い直し	35	【お】		音素	5, 23
言い淀み	35	応答	192	音素環境	33
異音	23, 34	応答生成	190	音素記号	22
依存構造	110	応答生成部	182	オンライン手法	145
依存構造解析	97	オートマトン	37	【か】	
意図構造	177	オフライン手法	145	ガーデンパス文	12
意図理解	190	オープンドメイン質問応答	166	外界照応	128
意味解析	113, 189	重み付き有限状態		開始記号	99
意味距離	233	トランスデューサ	53	解析/生成ピラミッド	207
意味素	113, 209	音圧	18	回答候補抽出	168
意味素性	7, 188	音韻	5	会話	176
意味表現	189	音韻処理	56	ガウス分布	38
意味役割	119	音響モデル	34, 37	過学習	46
意味役割付与	119	音響モデル適応	35	係り受け解析	109
インタフェース設計に		音源	16	係り受け規則	110
かかわる4原則	233	音声科学	1	係り受け文法	9
インテンシティ	18	音声学	5	書き言葉	10, 225, 226
咽頭	16	音声記号	22	格	115
イントネーション	6	音声区間検出	242	核	131
インプットメソッド	236	音声言語処理	2	格解析	117
韻律	6, 198	音声学	2	格フレーム	8, 117, 188
韻律処理	56	音声工学	2	格文法	115, 188
【う】		音声合成	2, 225	確率文脈自由文法	105
ウェブ資源	241	音声生成機構	2	隠れマルコフモデル	37
		音声生得説	1	下降型	100
				仮説探索	51

活性化関数	66
活用語幹	86
活用語尾	86
仮名漢字変換	94
仮名漢字変換方式	236
仮名漢字モデル	96
含意	127
漢字仮名混じり文	55
感情表示	197
間投詞	10

【き】

機械学習	112, 121
機械翻訳	4, 200, 225
木構造辞書	47
擬似適合性フィードバック	162, 166
記述長最小化基準	46
基準母音	22
規則に基づく機械翻訳	203, 206
期待回答タイプ	168
機能語	158
基本周期	18
基本周波数	17, 18, 29
基本周波数パターン生成	
過程モデル	59
逆コサイン変換	29
強化学習	194
共起	122
共振(共鳴)周波数	19
協調原理	177
局所距離	32
近似文字列照合	145, 173

【く】

グーグル日本語入力	
ソフトウェア	96
クエリ拡張	161
句構造	97, 110
句構造解析	97
句構造文法	99
句点	12

【け】

形式文法	98
形態素	7, 85
形態素解析	56, 85, 158
形態素解析アルゴリズム	88
形態素コスト	90
形態素数最小法	90
系列変換モデル	220
結索性	9, 177
決定木	45
ケプストラム係数	28
ケフレンシー	28
原言語	203, 215
言語科学	3
言語獲得装置	4, 14
言語行為論	178
言語工学	3
言語構造	177
言語処理部	181
言語的情報	16, 33
言語モデル	36, 47, 96, 215, 225
言語モデル重み	52

【こ】

語彙	7, 159
語彙化確率文脈自由文法	108
語彙サイズ	159
語彙知識獲得	122
口腔	16
交差エントロピー	69
公準	176
構造変換	208, 210
高速フーリエ変換	26
恒等写像	67
勾配消失問題	69
構文解析	97
構文解析済みコーパス	107, 122
後方照応	128
語義	119
語義曖昧性解消	119

語基化	158
呼気段落	55
誤差逆伝搬法	69, 222
50音表割当て型	238
コーパスに基づく	
機械翻訳	204, 214
固有振動数	19
固有表現抽出	168
混合ガウス分布	39
混合主導対話	184
コンフュージョン	
ネットワーク	228

【さ】

再帰型ニューラル	
ネットワーク	73
最急降下法	69
最小コスト法	90, 95
最長一致法	90
最尤状態系列	61
最尤推定	164
最尤推定法	41
最尤推定量	94, 97, 107
最尤特徴量系列	61
最良優先探索	52
逆茂木型の文	12
索引	151
索引付け	151, 171
三角行列	103
残差信号	29
三重母音	21

【し】

子音	5, 17
時間同期ビーム探索	51
時間分解能	25
識別モデル	54
シグモイド関数	66
事後確率	36
自己相関関数	29
自己符号化器	70
辞書	113
辞書の定義文	120

普遍文法	4
不要語リスト	158
フラットスタート	44
フーリエ級数	18
フーリエ変換	24
フレーズ対応	218
フレーズペア	219
フレーズ翻訳モデル	217, 228
フレーム	30
フレーム問題	181
フロントエンド処理	241
文	98
分散型音声認識	240
分散表現	77, 133, 221
文書	157
文書拡張	161, 174
文書検索	157, 168
文書ベクトル	159
分節音	6
文節数最小法	90
文節まとめ上げ	110
文法	9
文脈依存文法	99
文脈解析	127
文脈自由文法	9, 99, 111
文脈照応	128
文脈理解	189
分類語彙表	114
分類シソーラス	114
分類問題	168
【へ】	
ベイズの識別則	215
ベイズの定理	53
ベクトル空間法	165
ベクトル空間モデル	159
ベクトル値確率変数	162
編集距離	148
【ほ】	
ボイスサーチ	171
ボイヤー-ムーア法	146

母音	5, 16
母音図	19
報酬	193
母語	13
補正パープレキシティ	50
ボトムアップ	100
ボトルネック特徴量	72
翻訳モデル	215, 225

【ま】

マイクロフォンアレイ	244, 246
前向き確率	39
前向き中心	129
窓関数	24, 26
マルコフ過程	37
マルチモーダル	
インタフェース	231
マルチモーダル対話	195

【み】

右枝分かれ構造	11
見出し語化	158
未知語	14, 35, 91, 239
未定義語	91

【む】

結び	43
無声	26
無声音	19

【め】

名詞間の類似度	124
メタデータ	170
メディア	225
メル化対数パワー	
スペクトル	28
メル尺度	28
メル周波数ケプストラム	
係数	29
メンタルモデル	243

【も】

目的言語	203, 215
文字言語	7
文字列照合	145
文字 N グラム索引	152
モダリティ	195
モノフォンモデル	45
モーラ	5

【や】

訳語選択	120
ヤマト	201

【ゆ】

有限状態オートマトン	185
ユーザ主導対話	183
ユーザビリティ	242
ユーザビリティ工学	234
有聲	26
有声音	19
有声休止	11
優先中心	129
ユニグラム	49

【よ】

用言間の関係	125
用例に基づく機械翻訳	205
予測文字入力方式	236

【ら】

ラウドネス	18
ラティスデコーディング	228

【り】

離散フーリエ変換	25
離散 HMM	38
隣接ペア	177

【れ】

例示子	196
連結学習	43
接続可能性	87

接続コスト 90
 連続 DP マッチング 150

【ろ】

ロープナー賞 198

【わ】

話者ダイアライゼーション 242

【A】

A スター探索 52
 AE 70
 AI スピーカ 246
 ALPAC レポート 202
 A-D 変換 23

【B】

BM25 161
 BOW 158
 BP 法 69, 222

【C】

CKY 法 103
 CNN 75

【D】

DAE 72
 DF 160
 DNN 65
 DNN-HMM 65
 DP マッチング 32
 DTW 31

【E】

ELIZA 179
 EM アルゴリズム 41, 216

【F】

factoid 型質問 167

【G】

GMM 39
 GMM-HMM 65

【H】

HMM 合成 56

【I】

IBM モデル 215, 216
 IDF 160, 165, 168
 inside-outside アルゴリズム 107

【K】

KL ダイバージェンス 166

【L】

left-to-right 型 HMM 38
 LSTM 74, 222
 LVCSR 48, 241

【M】

Mozc 96
 MSD-HMM 64

【N】

N グラム 9, 47
 noisy channel model 36, 204, 227
 non-factoid 型質問 167

【O】

one-hot ベクトル 76, 221

【P】

POMDP 191
 put-that-there 197

【R】

RBM 70
 RMS 18
 RNN 73, 220

【S】

senone 71
 SHRDLU 180
 softmax 関数 67, 223

【T】

TF 160, 168
 TF-IDF 重み付け 160

【V】

VAD 242
 VoiceXML 247

【W】

wavenet 80
 WIMP インタフェース 231

中川 聖一（なかがわ せいいち）（編集，1章）

1976年京都大学大学院工学研究科博士課程修了（工学博士）。京都大学工学部情報工学科助手，豊橋技術科学大学情報工学系講師，助教授を経て，1990年豊橋技術科学大学教授。2014年定年退職（名誉教授）後，豊橋技術科学大学特任教授，特命教授を経て，2017年より中部大学工学部情報工学科教授。著書に「確率モデルによる音声認識」（単著，電子情報通信学会），「音声・聴覚と神経回路網モデル」（共著，オーム社），「情報理論の基礎と応用」（単著，近代科学社），「Speech, Hearing and Neural Network Models」（共著，IOS Press），「音声」（共著，岩波書店），「パターン情報処理」（単著，丸善），「Spoken Language Systems」（編著，IOS Press），「情報理論—基礎から応用まで—」（単著，近代科学社）がある。

小林 聡（こばやし さとし）（1章）

1999年静岡大学大学院工学研究科博士後期課程満期退学（博士（工学），静岡大学，2000年）。豊橋技術科学大学助手を経て，2004年より鳥根大学助教授。2015年退職後，山梨県でプログラミング教育に従事。著書に「トラ技コンピュータ増刊 はじめてのテキスト処理言語 AWK」（共著，CQ出版社）がある。

宇津呂 武仁（うつろ たけひと）（3章）

1994年京都大学大学院工学研究科博士課程修了（博士（工学））。奈良先端科学技術大学院大学助手，豊橋技術科学大学講師，京都大学講師，筑波大学助教授・准教授を経て，2012年より筑波大学教授。著書に「文字と音の情報処理」（共著，岩波書店），「音声認識システム」（共著，オーム社），「自然言語処理」（共著，オーム社）がある。

北岡 教英（きたおか のりひで）（5章）

1994年京都大学大学院工学研究科修士課程修了。日本電装株式会社（現デンソー）勤務，2000年豊橋技術科学大学大学院博士課程修了（博士（工学））。豊橋技術科学大学助手・講師，名古屋大学助教授・准教授を経て，2014年より徳島大学教授。著書に「ITSと情報通信技術」（共著，裳華房），「韻律と音声言語情報処理」（共著，丸善），「メディア情報処理」（共著，オーム社）がある。

峯松 信明（みねまつ のぶあき）（2章）

1995年東京大学大学院工学系研究科博士課程修了（博士（工学））。豊橋技術科学大学助手，東京大学助教授・准教授を経て，2012年より東京大学教授。著書に「音声認識システム」（共著，オーム社），「人と共存するコンピュータ・ロボット学」（共著，オーム社），「韻律と音声言語情報処理」（共著，丸善），「Digital resources for learning Japanese」（共著，Bononia University Press）がある。

秋葉 友良（あきば ともしよ）

（4章，6.4節）

1995年東京工業大学大学院総合理工学研究科博士後期課程修了（博士（工学））。通産省工業技術院電子技術総合研究所研究員を経て，2004年より豊橋技術科学大学助教授（現在，准教授）。

山本 幹雄（やまもと みきお）（6章）

1986年豊橋技術科学大学大学院修士課程修了（博士（工学）），豊橋技術科学大学，1992年。株式会社沖テクノシステムズラボラトリ勤務，豊橋技術科学大学教務職員・助手，筑波大学講師・助教授・准教授を経て，2008年より筑波大学教授。著書に「音声認識システム」（共著，オーム社），「特許情報処理：言語処理的アプローチ」（共著，コロナ社）がある。

甲斐 充彦 (かい あつひこ) (7章)
1996年豊橋技術科学大学大学院博士課程修了
(博士(工学))。豊橋技術科学大学助手、静岡大
学講師を経て、2000年より静岡大学助教授(現
在、准教授)。
著書に「ITSと情報通信技術」(共著、裳華房)、
「Spoken Language Systems」(共著、IOS
Press)がある。

土屋 雅稔 (つちや まさとし)
(8章、演習課題)
2004年京都大学大学院情報学研究科博士課程単
位認定退学(博士(情報学))、京都大学、2007
年)。豊橋技術科学大学助手・助教を経て、2014
年より豊橋技術科学大学准教授。

山本 一公 (やまもと かずまさ)
(8章、演習課題)
2000年豊橋技術科学大学大学院博士後期課程修
了(博士(工学))。信州大学助手、豊橋技術科学
大学助教・准教授を経て、2017年より中部大学
准教授。

音声言語処理と自然言語処理 (増補)

Spoken Language Processing and Natural Language Processing

© Seiichi Nakagawa 2013, 2018

2013年3月28日 初版第1刷発行
2018年9月20日 初版第3刷発行 (増補)



検印省略

編著者 中川 聖一
発行者 株式会社 コロナ社
代表者 牛来 真也
印刷所 三美印刷株式会社
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10
発行所 株式会社 コロナ社
CORONA PUBLISHING CO., LTD.
Tokyo Japan

振替 00140-8-14844・電話(03)3941-3131(代)
ホームページ <http://www.coronasha.co.jp>

ISBN 978-4-339-02888-1 C3055 Printed in Japan

(新宅)



JCCOPY <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつと事前に、出版者著作権管理機構(電話 03-3513-6969, FAX 03-3513-6979, e-mail: info@jcopy.or.jp)の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。