

文脈解析

— 述語項構造・照応・談話構造の解析 —

工学博士 奥村 学 監修

博士(情報理工学) 笹野 遼 平 共著
博士(工学) 飯田 龍

コロナ社

刊行のことば

人間の思考、コミュニケーションにおいて不可欠なものである言語を計算機上で扱う自然言語処理という研究分野は、すでに半世紀の歴史を経るに至り、技術的にはかなり成熟するとともに、分野が細かく細分化され、また、処理対象となるものも、新聞以外に論文、特許、WWW上のテキストなど多岐にわたり、さらに、応用システムもさまざまなものが生まれつつある。そして、自然言語処理は、現在では、WWWの普及とともに、ネットワーク社会の基盤を支える重要な情報技術の一つとなっているといえる。

これまでの自然言語処理に関する専門書は、自然言語処理全般を広く浅く扱う教科書（入門書）以外には、情報検索、テキスト要約などを扱う、わずかの書籍が出版されているだけという状況であった。この現状を鑑みるに、読者は、「実際にいま役に立つ本」、「いまの話題に即した本」を求めているのではないかと推測される。そこで、これまでの自然言語処理に関する専門書では扱われておらず、なおかつ、「いま重要と考えられ、今後もその重要性が変わらない」と考えられるテーマを扱った書籍の出版を企画することになった。

このような背景の下生まれた「自然言語処理」シリーズの構成を以下に示す。

1. 自然言語処理で利用される、統計的手法、機械学習手法などを広く扱う
近年の自然言語処理は、コーパスに基づき、統計的手法あるいは機械学習手法を用いて、規則なり知識を自動獲得し、それを用いた処理を行うという手法を採用することが一般的になってきている。現状多くの研究者は、他の先端的な研究者の論文などを参考に、それらの統計的手法、機械学習手法に関する知識を得ており、体系的な知識を得る手がかりに欠けている。そこで、そのような統計的、機械学習手法に関する体系的知識を与える専門書が必要と感じている。
2. 情報検索、テキスト要約などと並ぶ、自然言語処理の応用を扱う
自然言語処理分野も歴史を重ね、技術もある程度成熟し、実際に使えるシステム、技術として世の中に少しずつ流通するようになってきている

ii 刊 行 の こ と ば

ものも出てきている。そのようなシステム、技術として、検索エンジン、要約システムなどがあり、それらに関する書籍も出版されるようになってきている。これらと同様に、近年実用化され、また、注目を集めている技術として、情報抽出、対話システムなどがあり、これらの技術に関する書籍の必要性を感じている。

3. 処理対象が新しい自然言語処理を扱う

自然言語処理の対象とするテキストは、近年多様化し始めており、その中でも、注目を集めているコンテンツに、特許（知的財産）、WWW上のテキストが挙げられる。これらを対象とした自然言語処理は、その処理結果により有用な情報が得られる可能性が高いことから、研究者が加速度的に増加し始めている。しかし、これらのテキストを対象とした自然言語処理は、これまでの自然言語処理と異なる点が多く、これまでの書籍で扱われていない内容が多い。

4. 自然言語処理の要素技術を扱う

形態素解析、構文解析、意味解析、談話解析など、自然言語処理の要素技術については、教科書の中で取り上げられることは多いが、技術が成熟しつつあるにもかかわらず、これまで技術の現状を詳細に説明する専門書が書かれることは少なかった。これらの技術を学びたいと思う研究者は、実際の論文を頼らざるを得なかったというのが現状ではないかと考える。

本シリーズの構成を述べてきたが、この構成は現在の仮のものであることを最後に付記しておきたい。今後これらの候補も含め、新たな書籍が本シリーズに加わり、本シリーズがさらに充実したものとなることを祈っている。

本シリーズは、その分野の第一人者の方々に各書籍の執筆をご快諾願えたことで、成功への最初の一步を踏み出せたのではないかと思っている。シリーズの書籍が、読者がその分野での研究を始める上で役に立ち、また、実際のシステム開発の上で参考になるとしたら、この企画を始めたものとして望外の幸せである。最後に、このような画期的な企画にご賛同下さり、実現に向けた労をとって下さったコロナ社の各氏に感謝したい。

2013 年 12 月

監修者 奥 村 学

ま え が き

本書は、「述語項構造解析」，「共参照・照応解析」，「談話構造解析」という自然言語テキストの文脈理解に関連した三つの解析についてまとめた専門書である。これらの解析はいずれも，形態素解析や係り受け解析などの表層に近いレベルの解析から，自然言語テキストの意味理解に向けて一步踏み込んだ処理であり，高度な自然言語処理アプリケーションを実現するにあたり重要となってくる技術であるといえる。しかし，文脈理解に関連した処理は、『意味』という漠としたものを扱う必要があることから，これらの処理が対象とする問題の定義は，どのような『意味』を捉えたいかによって異なっており，また，必然的に複雑なものとなっている。このような問題設定の多様性・複雑性は，自然言語処理の初学者が文脈解析に取り組む際のハードルとなっているように思える。

そこで本書では，文脈理解に向けて必要となる処理のうち，現時点で比較的盛んに研究されている「述語項構造解析」，「共参照・照応解析」，「談話構造解析」という三つのトピックを対象に，解析手法そのものよりも，それぞれの解析がどのような問題を解こうとしているかや，関連する言語資源がどのような考えの下で作成されているかを中心に解説する。本書では読者として，これから文脈解析に取り組もうとする研究者や大学院生に加え，文脈処理をより高次のタスクに応用したいと考えている研究者や，文脈解析を利用したシステムの開発を行いたいと考えている技術者など，文脈解析に興味をもつすべての方を想定しており，自然言語処理に関する最低限の事前知識で内容を理解できるような記述を心掛けた。本書が，多くの方々にとって，自然言語処理分野における文脈解析に向けた取組みを理解する一助となれば幸いである。

本書の構成（担当）は以下のとおりである。まず，1章「はじめに」（飯田）において文脈解析について概説した後，2章で「述語項構造」（笹野），3章で

「共参照・照応解析」(3.1.2 項, 3.1.3 項, 3.3 節が笹野, それ以外が飯田), 4 章で「談話構造解析」(飯田)についてそれぞれ解説する。また, 文脈解析というテーマからは外れるが, 付録において「自然言語処理の分野で使用される代表的な検定手法」(笹野)を紹介する。これは, 著者の 1 人(笹野)が自然言語処理の研究を進めるにあたり, 自然言語処理で使用される検定に関する情報がまとめられた書籍やサイトが見つからず苦労した経験があることから, 自然言語処理の分野で使用される代表的な検定手法についてまとめたものである。

本書を執筆するにあたり多くの方々にお世話になった[†]。東京工業大学の奥村学教授には本書のご提案および監修をいただいた。東京工業大学の高村大也准教授, 富士通研究所の横野光氏, 東北大学の松林優一郎特任助教, NTT コミュニケーション科学基礎研究所の平尾努氏, 東京工業大学の馬縹美穂氏には, 完成前の原稿を読んでいただきさまざまな改善点や助言をいただいた。東京大学の西村義樹教授, 立命館大学の岡本雅史准教授, 杏林大学の黒田航准教授には言語学分野における考え方などについてご教示いただいた。また, コロナ社様には本書の刊行のために大変なご尽力をいただいた。ここに感謝の意を表す。

平成 29 年 5 月

笹野 遼平

[†] 以下, 所属と肩書きについては, 本書編集当時のものである。

目 次

1. は じ め に

2. 述語項構造解析

2.1 述語項構造とは	5
2.2 述語項構造の定義	7
2.2.1 述 語 の 種 類	7
2.2.2 述語項構造における項	9
2.2.3 表層格に基づく関係ラベル	12
2.2.4 深層格に基づく関係ラベル	27
2.3 述語以外の項構造	41
2.3.1 日本語における名詞の項構造	43
2.3.2 英語における名詞の項構造	45
2.4 述語項構造解析	47
2.4.1 日本語述語項構造解析	47
2.4.2 英語を対象とした深層格の解析	49
2.5 述語項構造解析のツール	51
2.6 述語項構造解析の応用	53

3. 共参照・照応解析

3.1 共参照と照応	55
3.1.1 共参照	56
3.1.2 照応	62
3.1.3 共参照と照応の関係	72
3.2 共参照・照応解析	74
3.2.1 人手で作成した規則を利用した共参照・照応解析手法	75
3.2.2 統計的な共参照・照応解析手法	79
3.2.3 ゼロ照応解析	88
3.3 共参照・照応解析の評価	94
3.3.1 共参照解析の評価	94
3.3.2 照応解析の評価	106
3.4 共参照・照応解析ツール	110

4. 談話構造解析

4.1 談話とは	112
4.2 談話の構成単位	115
4.2.1 談話単位	115
4.2.2 サブトピック単位	116
4.3 談話関係分類, 談話構造	118
4.3.1 修辞構造理論	118
4.3.2 談話関係のグラフ構造: Discourse GraphBank	123
4.3.3 隣接談話関係: Penn Discourse Treebank	126
4.3.4 議論区画化	129

4.4 談話構造解析	130
4.4.1 談話単位分割	131
4.4.2 談話関係分類	132
4.5 談話構造解析ツール	136
4.6 局所的な首尾一貫性の推定	137
4.7 談話構造解析の応用	140
4.7.1 自動要約への応用	140
4.7.2 小論文の自動採点への応用	142
 付 録	144
A.1 統計的仮説検定の基礎と性質	144
A.1.1 統計的仮説検定の基礎	144
A.1.2 仮説検定の分類	146
A.1.3 統計的仮説検定の一般的性質	148
A.2 自然言語処理における検定	150
A.2.1 フィッシャーの正確検定	150
A.2.2 マクネマー検定	152
A.2.3 対応のあるデータに対する並べ替え検定	153
A.2.4 ウィルコクソンの符号順位検定	156
A.2.5 一般の並べ替え検定	158
A.2.6 ブートストラップ法	159
A.3 検定に関する一般的な注意点	162
引用・参考文献	168
索 引	182

索引

【い】		共参照解析	56	修辭構造理論	118
意味役割	7	共参照関係	55	周 辺	119
意味役割付与	49	京大コーパス	14	主 題	116
【う】		京都大学テキストコーパス	14	主題役割	7, 37
ウィルコクソンの順位和検定	159	議論区画化	129	出現形	20
ウィルコクソンの符号順位検定	156	近似検定	148	述 語	5
後ろ向き中心	76	近似的ランダム化検定	155	述語項構造	5
【え】		【け】		首尾一貫性	112
F 値	95	形態素	5	照 応	62
【か】		形態素解析	5	照応関係	55
外界照応	64	言 及	60	照応詞	62
概念照応	94	検出力	146	照応性決定	84
属性照応	67	検出力曲線	149	省略解析	88
下位範疇化	50	【こ】		信頼区間	160
係り受け解析	5	語彙概念構造	42	【せ】	
格	9	項	5, 9	正確検定	148
格交替	19	後方照応	63	ゼロ照応	66
格フレーム	28	コピュラ	8	ゼロ照応解析	88
仮説検定	144	【さ】		ゼロ代名詞	66
片側検定	147	最強力検定	146	ゼロ代名詞検出	89
含意関係認識	53	再現率	95	先行詞	62
関係名詞	42	最小の談話単位	118	選択制限	91
間接照応	69	サブトピック	115	選択選好	91
換 喩	70	【し】		前方照応	63
【き】		事態共参照	61	【そ】	
擬似自動詞	11	事態性名詞	43	層化シャッフリング	159
帰無仮説	144	実 体	60	外の関係	20
		実体照応	67, 94	【た】	
		自動採点	142	第 1 種の誤り	146
		ジャックナイフ法	159	第 2 種の過誤	146
				代名詞照応	65

AZ	129	Discourse GraphBank	125	hypothesis test	144
[B]		discourse parsing	115	[I]	
backward-looking center	76	discourse relation	114	identity-of-reference	
binomial test	146	discourse relation classifi-		anaphora	67
BLANC	94	caiton	131	identity-of-sense anaphora	
Bonferroni correction	166	discourse structure	114		67
bootstrap resampling	159	discourse structure analysis	114	implicit discourse relation	
bridging	69	discourse unit	114	126	
bridging anaphora	69	discourse unit segmentation	130	Indefinite Null Instantiation	
bridging reference	69			32	
B-CUBED	94	DNI	32	indirect anaphora	69
[C]		[E]		INI	32
case	9	EDU	118	IRA	67, 94
case frame	28	elementary discourse unit		ISA	67, 94
cataphora	63	118		[J]	
CEAF	94	endophora	63	jackknife resampling	159
center	75	entity	60	[L]	
centering theory	75	entity-grid	137	lexical unit	31
CNI	32	event coreference	61	[M]	
coherence	112	event noun	43	Mann-Whitney U Test	159
comparative anaphora	71	Excludes Relation	31	markable	58
CoNLL スコア	103	exophora	64	McNemar's test	152
Constructional Null Instantiation	32	explicit discourse relation	126	mention	60
copula	8	Extra-thematic	31	metonymy	70
Core	31	[F]		morpheme	5
coreference relation	55	FF	30	morphological analysis	5
coreference resolution	56	Fisher's exact test	151	most powerful test	146
CoreSet	33	forward-looking center	75	MUC	94
core argument	9	frame	30	multiple testing	165
cross-document coreference	73	FrameNet	30	[N]	
[D]		frameset	35	nucleus	119
Definite Null Instantiation	32	frame element	30	null hypothesis	144
dependency parsing	5	F-measure	95	[O]	
DGB	125	[G]		obligatory case	9
direct anaphora	68	governing category	50	optional argument	9
discourse	112	[H]		optional case	9
discourse element	55	Holm correction	166	Other Anaphora	71

[P]		role	35	Type II error	146
PDTB	126	roleset	35	Type I error	146
Penn Discorse Treebank	126	RST	118	t-test	146
Peripheral	31	RST Discourse Treebank		[U]	
permutation test	153	123		ungrammatical sentence	10
population	144	RST-DT	123	unsaturated noun	42
precision	94	RTE	53	[V]	
predicate	5	[S]		verbal noun	43
predicate-argument structure	5	satellites	119	verb anaphora	67
preferred center	76	semantic role	7	[W]	
pronominal anaphora	65	semantic role labeling	49	Wilcoxon rank-sum test	159
PropBank	34	significance level	144	Wilcoxon signed-rank test	156
Proposition Bank	34	significant	144	[Z]	
proto-role	29	sign test	146	zero anaphora	66
pseudo-intransitive	11	Singleton	106	zero anaphora resolution	88
[R]		statistical power	146	zero pronoun	66
randomization test	155	stratified shuffling	159	zero pronoun detection	89
recall	95	subcategorization	50	[T]	
recognizing textual entailment	53	subtopic	115	thematic role	7, 37
relational noun	42	surface case	7	topic	116
rhetorical structure theory	118	syntactic analysis	15		

—— 監修者・著者略歴 ——

奥村 学（おくむら まなぶ）
1984年 東京工業大学工学部情報工学科卒業
1989年 東京工業大学大学院博士課程修了
（情報工学専攻）、工学博士
1989年 東京工業大学助手
1992年 北陸先端科学技術大学院大学助教授
2000年 東京工業大学助教授
2007年 東京工業大学准教授
2009年 東京工業大学教授
現在に至る

笹野 遼平（ささの りょうへい）
2004年 東京大学工学部電子情報工学科卒業
2009年 東京大学大学院博士課程修了
（電子情報学専攻）博士（情報理工学）
2009年 京都大学特定研究員
2010年 東京工業大学助教
2017年 名古屋大学准教授
現在に至る

飯田 龍（いいだ りゅう）
2002年 九州工業大学情報工学部知能情報工
科卒業
2007年 奈良先端科学技術大学院大学博士後期
課程修了（情報処理学専攻）
博士（工学）
2007年 奈良先端科学技術大学院大学特任助教
2008年 東京工業大学助教
2014年 情報通信研究機構主任研究員
現在に至る

文脈解析 — 述語項構造・照応・談話構造の解析 —

Discourse Analysis

— Analyses of Predicate Argument Structure, Anaphora, and Discourse Structure —

© Ryohei Sasano, Ryu Iida 2017

2017年6月15日 初版第1刷発行

検印省略

著 者 笹 野 遼 平
飯 田 龍
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 978-4-339-02760-0 C3355 Printed in Japan

(金)



JCOPY <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、そのつど事前に、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。