自 然 言 語 処 理 100－6
（1994．3．17）

自然言語処理研究の動向と問題点

野村 浩郷 (九工大)　林 良彦 (NTT)　丸山 宏 (日本IBM)　永井 秀利 (九工大)

自然言語処理研究会は1981年に前身である計算言語学研究会を発展させる形で発足し，着実に発展してきた．高度情報化社会において，自然言語処理は最も重要な技術の一つであり，社会の要求は今後もますます多様化・高度化の一途をたどると予想される．第100回の研究会を迎える今，これを一つの節目として，ここ3年間を中心として過去の研究会活動と研究発表の動向を振り返り，今後の研究課題を概観することにより，自然言語処理研究会のさらなる発展を期す．

Progress Report of the SIG–NLP of IPSJ

Hirosato NOMURA (Kyushu Inst. of Tech.)　Yoshihiko HAYASHI (NTT)
Hiroshi MARUYAMA (IBM Research)　Hidetoshi NAGAI (Kyushu Inst. of Tech.)

Natural language processing is one of the most important technologies in the computerized society. Special Interest Group of Natural Language Processing of IPSJ has been continuing its high activity since its establishment. In this paper we present the topics of the papers presented at its periodical conferences in this one decade, focusing on these three years. We also list a set of key words for the further elaboration in natural language research.

1 はじめに

人間が用いる言語である自然言語を計算機を用いて処理する技術である自然言語処理は，高度情報化社会において最も重要な技術の一つであり，従来から多くの研究・開発が世界中で精力的に進められ，様々な成果を上げてきた．日本語ワードプロセッサや機械翻訳システムは，従来からの研究による成果の代表的なものと言える．

計算機の能力の向上や低価格化の急速な進行は，かなり多くの計算機資源を要求する自然言語処理を身近なものへと変えてきた．それに伴い，自然言語処理に対する要求は多様化・高度化の一途をたどっている．近年のコンピュータネットワークの発達もそれに拍車をかけていると言えよう．ネットワークの発達は，計算機可読な形でその上を流れる膨大な量の自然言語情報を生んだ．その結果，そのような情報の中から必要な情報を効率良く抽出するなどの処理に対する強いニーズが生じている．

従来からの研究により，自然言語技術の基盤の一部は明確になったが，社会の要求に十分応えることができるだけの真に高度な技術を確立するためには，基礎理論，処理方式，システム化，およびシステム活用法のそれぞれにおいて解決されねばならない課題がまだまだ多く残されている．

自然言語処理研究会は，1981 年に，前身の計算言語学研究会 (和田弘主査) を発展させる形で発足した．長尾真，吉田将，田中穂積の歴代主査のご指導と会員諸兄のご尽力とにより，自然言語処理研究会は着実に発展してきた．第 100 回の研究会を向かえる今，これを一つの節目として，ここ 3 年間を中心として過去の研究会活動と研究発表の動向を振り返り，今後の研究課題を概観することにより，自然言語処理研究会のさらなる発展の一助にしたい．

本稿では，まず，自然言語処理研究会の発展状況の概要を示す．次に，自然言語処理研究会において発表された論文の内容の動向の概略を整理する．最後に，今後の自然言語処理および研究会活動の問題点に関する概要を述べる．

2 発表件数の動向

2.1 定例研究会での発表件数の動向

自然言語処理研究会では，近年，各年度 6 回の定例研究会を計画し，開催している．ここ 3 年を見た場合，当初，1991 年度は年 6 回のすべてを 1 日の開催，1992，1993 年度は年間開催数の半数である 3 回を 1 日，残りを 2 日間の開催とするように年間活動計画を立てたが，発表申し込みが非常に多く，1 日での開催予定を 2 日間に変更したり，1 件あたりの発表時間を若干短縮したりして対処する必要が多く生じた．

ここ 3 年間 (第 82 回～第 99 回) の定例研究会における発表件数を表 1 に示す．

表 1: ここ 3 年間の発表件数

開催年	91					92						93						94
開催月	3	5	7	9	11	1	3	5	7	9	11	1	3	5	7	9	11	1
回	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
件数	11	19	20	13	9	12	15	12	10	11	12	13	13	17	11	19	11	13

発表件数は比較的安定しており，1 回の研究会につき 12～13 件程度であることが多い．

当研究会では，通常，1 件あたり 40 分の発表時間として，1 日に 10 件の発表が行えるように研究会プログラムを組んでいる．ただし，申し込み件数が 11 件の場合は発表時間を 35 分に短縮して 1 日に収

めている．したがって12〜13という発表件数は，1日で開催するか2日間とするかの境目である．

今後も同程度かそれ以上の研究発表が行われると考える(自然言語処理に対するニーズの増大を鑑み，その可能性は十分に高い)と，今後，すべての定例研究会を2日間の開催として計画したり，あるいは1件あたりの発表時間を思い切って30分まで短縮して，13〜14件の発表を1日に収めるように計画したりすることが必要となるであろう．

研究発表の動向を見るために，過去10年間の分野別発表件数を表2に示す．

表2: 研究発表の分野と件数

開催年	84	85	86	87	88	89	90	91	92	93	94	合計	比率
開催数	6	6	6	6	5	6	6	6	6	6	1	60	
発表件数	37	45	41	50	36	58	63	86	72	84	13	585	
分野別内訳													
音声	0	0	0	0	0	0	1	1	0	0	0	2	0.34
形態素解析	0	2	0	4	2	1	4	4	4	6	1	28	4.79
構文解析	4	6	5	7	10	11	9	15	20	7	1	95	16.24
意味解析	6	7	7	12	1	15	14	13	13	7	1	96	16.41
談話・文脈	0	3	4	1	0	6	8	8	10	9	0	49	8.38
文(章)生成	3	0	3	4	1	2	6	8	2	0	0	29	4.96
辞書	1	0	2	0	1	3	2	4	3	8	0	24	4.10
文法	0	2	1	2	4	0	1	1	0	0	0	11	1.88
対話	1	2	1	3	1	1	0	2	3	4	3	21	3.59
機械翻訳	7	12	9	9	6	10	10	17	7	13	2	102	17.44
言語分析	3	5	1	3	6	2	0	2	0	11	2	35	5.98
情報抽出	3	2	3	3	1	4	3	3	1	4	0	27	4.62
情報検索	1	3	1	0	1	1	1	2	1	5	2	18	3.08
ツール	6	1	1	1	0	1	1	4	6	9	1	31	5.30
会議報告	2	0	3	1	2	1	3	2	2	1	0	17	2.91

この表から，過去10年間において開催回数，発表件数ともに研究会活動は一定以上の水準を保っており，この研究会が昔から盛んな研究会であったことをうかがわせる．また，発表件数は全体的に伸びてきており，研究会が着実に発展してきたことを示している．特にここ3年の発表件数は多い．1992年は，バブル経済の崩壊による景気の悪化の影響か，前年に比べて発表件数が減少しているが，翌1993年には景気の回復を待たず，1991年の水準に復帰している．この点から見ても，今後，さらなる発表件数の増加が予想される．

分野別に見ると，機械翻訳，意味解析，構文解析が上位に並んでいる．機械翻訳は，過去10年，毎年のように発表件数の上位に存在し，自然言語処理研究の重要なテーマとしての地位を保ち続けている．各年の分野別内訳を見ると，例年，合計での上位を占める機械翻訳等の3分野を中心とした2〜3の分野に発表が集中し，その年の研究動向を特徴付けてきた．しかし昨年は発表件数が分散し，例年とは異なる傾向を見せている．昨年1年のことであるため，これが今後の傾向となるかは不明であるが，これは，従来からの研究で自然言語処理の基盤技術の一部が明確化してきたことを背景に，高度情報化社会の多様で高度な要求に応えるべく，自然言語処理研究が次の段階へ進もうとしている動きとも考えることができる．

より長期的に分野別の変化を見るために，ここ3年間(第82回研究会から第99回研究会まで)とそれ以前の3年間(第66回研究会から第81回研究会まで)とを比較したものを表3に示す．

3年間での発表件数の総数が163件から241件と大幅に増加しているため，少々比較しづらいものとなっているが，目につく点としては

表 3: ここ 3 年間とそれ以前の 3 年間との比較

分野	ここ3年間 (第82〜99回)		それ以前 (第66〜81回)	
	合計	比率	合計	比率
音声	1	0.41	1	0.61
形態素解析	14	5.81	8	4.91
構文解析	42	17.43	29	17.79
意味解析	30	12.45	33	20.25
談話・文脈	27	11.20	14	8.59
文(章)生成	7	2.90	11	6.75
辞書	15	6.22	6	3.68
文法	0	0.00	6	3.68
対話	11	4.56	2	1.23
機械翻訳	38	15.77	27	16.56
言語分析	15	6.22	6	3.68
情報抽出	7	2.90	9	5.52
情報検索	9	3.73	4	2.45
ツール	20	8.30	2	1.23
会議報告	5	2.07	5	3.07
合計	241		163	

- 構文解析，機械翻訳の相変わらずの多さ
- 構成比率としての意味解析の減少
- 談話・文脈や対話の増大
- ツールの大幅な増大

などが挙げられる．意味解析が減少し，文脈処理等が増大したことは，自然言語が表す意味の追求が，文単位で表される意味への興味から，文のつながりの上で表される意味へと移り変わってきたためとも考えられる．

2.2 小規模国際会議での動向

自然言語処理研究会では，ここ 3 年間に次の 2 回の小規模国際会議を主催した．いずれも 50 件を越える発表，100 名を越える参加者を数え，大変盛況な小規模国際会議であった．

1. 自然言語処理環太平洋シンポジウム
   Natural Language Processing Pacific Rim Symposium (NLPRS)
   後　援：日本シンガポールＡＩセンター
   日　程：平成 3 年 11 月 25 日 (月)〜26 日 (火)
   場　所：National Computer Board (Singapore)

2. 自然言語処理環太平洋シンポジウム '93
   Natural Language Processing Pacific Rim Symposium '93 (NLPRS '93)
   共　催：福岡工業大学
   日　程：平成 5 年 12 月 6 日 (月)〜7 日 (火)
   場　所：福岡工業大学

表4: NLPRS における国別発表件数，参加者数

国名	件数	人数
Japan	27	31
Korea	5	9
ROC	5	5
France	4	4
China	3	3
Singapore	2	41
Hong Kong	2	2
United Kingdom	2	2
Thailand	1	2
Malaysia	1	1
Australia	0	1
合計	52	101

表5: NLPRS における分野別発表件数

分野	件数
Parsing	12
Project Report	8
Generation	7
Grammar	5
Knowledge	5
Application	4
Dialogue	4
Understanding	4
Discourse	3

表6: NLPRS '93 における国別発表件数，参加者数

国名	件数	人数
Japan	37	97
Korea	7	17
China	3	3
India	3	2
U.S.A.	2	4
France	2	2
United Kingdom	2	2
R.O.C.	1	2
Singapore	1	1
Sweden	1	1
Thailand	0	2
Indonesia	0	1
Malaysia	0	1
合計	59	135

表7: NLPRS '93 における分野別発表件数

分野	件数
Generation	6
Grammar	6
Semantics	6
Corpus/Example Based	4
Ellipsis	4
Machine Translation	4
Syntactic Analysis	4
Text Processing	4
Disambiguation	3
Tool	3
Homonyms	2
Ill-Formedness	2
Lexicon	2
Term	2
Acquisition	1
Dialogue	1
Extraction	1
Interface	1
Parsing	1
Resolution	1
Understanding	1

NLPRS における国別の発表件数および参加者を表4に，セッションタイトルに基づいて分類した分野別発表件数を表5に示す．また，NLPRS '93 における国別の発表件数および参加者を表6に，分野別発表件数を表7に示す．会議名称からして当然と言えるが，アジア各国からの参加が多く，日本を中心としたアジアにおける自然言語処理の研究動向の一側面を見ることができよう．

NLPRS では形態素解析分野の研究発表が少なかったことが目に付く．また，構文解析分野の発表数の減少が見られる．定例研究会での動向とは少し異なるが，構文解析や意味解析という従来の中心からの研究分野の広がりが生じ，少数分野にまとめるのが難しくなってきているのは定例研究会と同様の動きであろう．

2.3 COLING-92 における発表件数

表 8 ～ 9 は，1992 年に開催された COLING (International Conference on Computational Linguistics) について，同様の分布をみたものである [2].

表 8: COLING における国別発表件数，参加者数
(発表件数 5 件以上の国のみ)

国名	件数	人数
U.S.A.	46	70
Japan	35	62
France	29	134
Germany	17	57
United Kingdom	17	36
Canada	8	13
R.O.C.	8	10
Italy	7	14
Netherlands	7	19

表 9: COLING における分野別発表件数

分野	件数
applications	47
morphology, phonology, syntax	33
large-scale resources	30
computational methods ("paradigms")	23
discourse and dialogue	20
tools	19
semantics, pragmatics	18
NLP and hypermedia	11
generic questions in language industry	5
合計	206

表 8 は 5 件以上が採録された国のみについて示したものである．発表件数，参加者数ともに日本は上位に入っている．

この分類では applications の件数が多い．ただしこの中には，機械翻訳，文書作成支援，情報検索といった応用の他にコーパス・事例に関する処理や文章生成の論文も何点か含まれているので注意が必要である．

目に付く点としては，

- 自然言語処理をより実用的なものとするための，ロバスト性・効率向上のための研究が盛ん
- コーパスや用例に基づく言語知識獲得や統計的な解析手法の研究が盛ん
- 機械翻訳以外の「良い応用」を探す動きが見られるものの，まだ機械翻訳が応用の主流である

などが挙げられる．1993 年の分野別発表件数の変化 (表 2 参照) には，この COLING での動向も影響していると言えよう．

3 キーワード

ここ 3 年間の自然言語処理研究会における発表のキーワード一覧を表 10 に示す．この表では頻度 3 以上のものだけを挙げている．これは発表のタイトルから作成したものであるため，必ずしも正確なものではないが，おおよその傾向はつかむことができると考えられる．

言語に関しては「日本語」,「英語」がキーワード一覧に挙がっているが，これは納得のいくところであろう.「機械翻訳」が相変わらず多いのも，定例研究会の発表件数の動向で見てきた通りである．注目すべきは「校正支援」,「要約」,「文章構造」などの文章を扱うキーワードの増加であろう．これは自然言語処理が，文処理から文章処理へとステップアップする動きの一つであり，COLING の動向を見る際にも述べた新しい「良い応用」を探す動きであると言えよう．

表 11 に NLPRS '93 における発表のキーワード一覧を示す．これは著者自らが付けたキーワードに基づくものあり，キーワード選出の方法が表 10 の場合とは異なる．それゆえ，二つの表の違いをその

表 10: ここ 3 年間における発表のキーワード

順位	キーワード	頻度	順位	キーワード	頻度
1	日本語	29	11	曖昧性解消	5
2	機械翻訳	16	18	法律文	4
3	構文規則	10	19	キーワード抽出	3
4	日英機械翻訳	9	19	チャート	3
5	ＬＲパーザ	8	19	意味解析	3
5	校正支援	8	19	英語	3
5	対話	8	19	概念学習	3
8	辞書	7	19	結束性	3
9	かな漢字変換	6	19	質問応答	3
9	助詞	6	19	推敲	3
11	マルコフモデル	5	19	制限言語	3
11	英日機械翻訳	5	19	評価	3
11	形態素解析	5	19	文字認識	3
11	知識獲得	5	19	文章構造	3
11	文章生成	5	19	用例	3
11	要約	5			

表 11: NLPRS '93 における発表のキーワード

順位	キーワード	頻度	順位	キーワード	頻度
1	machine translation	9	13	natural language processing	3
2	dialogue	8	13	semantic description	3
3	corpus	7	13	tagging	3
3	semantics	7	13	transfer	3
5	lexicon	6	26	compound word	2
6	ellipsis	5	26	controlled linguistic model	2
6	generation	5	26	dependency	2
8	example based machine translation	4	26	dictionary	2
8	interlingua	4	26	disambiguation	2
8	Japanese	4	26	discourse	2
8	parsing	4	26	error detection	2
8	understanding	4	26	function words	2
13	case structure	3	26	information retrieval	2
13	Chinese	3	26	kana-to-kanji conversion	2
13	co-occurrence relation	3	26	mental image	2
13	document processing	3	26	speech	2
13	knowledge acquisition	3	26	subcategorization	2
13	knowledge representation	3	26	syntax	2
13	Markov model	3	26	TAG	2
13	morphology	3	26	terminology	2
13	natural language interface	3	26	thesaurus	2

まま研究会と小規模国際会議との発表動向の違いとすることはできないが，おおよその発表の傾向を見るには十分に役立つ．

ここでも相変わらず machine translation が上位となっている．dialogue や corpus，semantics がそれに続くが，corpus がこれだけ上位に来ているのは特徴的であろう．example based machine translation が高い順位にあるのも，現在の機械翻訳研究の中心がどこにあるかを示している．

4 課題と将来の動向

これまでの自然言語処理は，機械翻訳がその応用の中心となって，自然言語処理研究全体を引っ張って来たと言うことができる．現在の自然言語処理研究は，これまでの多くの研究成果を背景に，今後の研究の流れを引っ張っていけるような新たな「良い応用」を模索している段階にあると言える．「良い応用」を求める動きが強まった理由は，一つには，日本語ワードプロセッサの普及などにより，社会が自然言語処理を身近な技術として認識してきたため，また一つには，計算機あるいはコンピュータネットワーク上に計算機可読なデータが大量に蓄積されてきたためと考える．この社会的に注目された状況は，社会の高度で多彩な期待に対し，現在何ができるのか，将来何ができるようになるのかの提示を強く求めている．これに応えていくことは，今後の自然言語処理における重要な課題の一つであろう．これは機械翻訳がもはやそのような役割を果たさないということではない．機械翻訳に対する社会的期待に完全に応えるには解決すべき重要な課題はまだまだ多く含まれている．また，機械翻訳で研究されてきた用例に基づく手法を他の分野へ適用しようという動きがあるなど，「良い応用」としての地位を十分に保っている．発表件数の相変わらずの多さがそれを物語る．

キーワード等から予想される今後の動向としては，まず，意味処理の困難さの認識に基づき，できるだけ表層に近い処理による情報抽出や要約，文書(文章)処理などを目指す動きがある．これらは，現在の技術の比較的近い延長線上で実現可能な自然言語処理の応用を社会に示そうとする動きとも取れる．また，機械翻訳に限らず，様々な分野で，コーパスなどから得た用例に基づく自然言語処理を試みる動きも目に付く．コーパスなどの言語データ利用への関心の強さは，1992 年に当研究会で開催したシンポジウム [3] におけるパネル討論「言語データ　— 製作者の視点, 利用者の視点 —」への参加者の多さにも見ることができる．形態素解析や構文解析の分野では，自然言語処理をより実用的なものとするためのロバスト性・効率向上のための研究が，意味解析の分野では，単一の文が表す意味よりも，文脈，談話，対話等の文の流れの上で表される意味をとらえていこうとする動きが，ますます活発化するであろうと考える．近年，マルチメディアという言葉がもてはやされているが，これらのデータをどう利用していくかという観点からも自然言語処理が重要な役割を担っていくであろう．

5 おわりに

自然言語処理研究会の過去 10 年間の研究発表の動向を，ここ 3 年間の発表を中心として示した．また，この分析を踏まえて，今後の課題や動向に関する予測についても触れた．

自然言語処理に対する社会の関心や期待はますます強いものとなっている．まだそのすべてには応えきれていないものの，研究は着実に進展しており，自然言語処理の将来への道は必ずしも平坦ではないが希望に満ちている．会員諸氏の今後のより一層のご活躍を祈ると共に，本稿がその一助となれば幸いである．

参考文献

[1] 野村, 田中, 徳永, 内藤:自然言語処理研究の動向と課題, 情報処理学会自然言語処理研究会 83–1, 1991

[2] 林, 石崎, 古瀬, 木下:第 14 回計算言語学国際会議 (COLING–92) 報告, 情報処理学会自然言語処理研究会 93–13, 1993

[3] 自然言語処理シンポジウム, 情処シンポジウム論文集 Vol.93 No.1