

## 中間言語方式による日本語述部の英訳について\*

田 辺 利 文\*\*

吉 村 賢 治\*\*\*

首 藤 公 昭\*\*\*

### A Japanese-English Machine Translation through a Pivot Language

Toshifumi TANABE\*\*, Kenji YOSHIMURA\*\*\* and Kosho SHUDO\*\*\*

Generally, the meaning of a sentence consists of ( i ) propositional contents and ( ii ) non-propositional contents. Not only the propositional contents but also the non-propositional contents play a critical role in various NLP applications. We have presented that proper employment of Multiword Expressions(MWEs) as lexical units yields a straightforward mapping from the sentence-final, syntactical structure of Japanese sentence to its non-propositional, semantic structure (NPS).

In this paper, we introduce an experimental machine translation system, whose input is a Japanese sentence-final predicative parts and whose output is its English surface forms. The system recognizes NPS of input predicative parts as a pivot language in the translation process. However, NPS-constructor has produced a low precision. In this paper, also, we introduce a method to produce a better precision by aspect-feature constraints within the present framework.

*Key Words* : Natural Language Processing, Multiword Expression, Modality, Machine Translation, Pivot Language, Non-Propositional Contents

#### 1. はじめに

社会のグローバル化とともに、多国間での情報流通は急速に増大している。それに伴い、機械翻訳に対するニーズも非常に高まっている。中でも共通語としての英語の重要性は言うまでもない。日英機械翻訳の研究は今日まで長年に渡って行われてきたが、現時点でもなお翻訳サイトや翻訳ソフトが出力する文章の訳質は十分であるとは言いがたい。正確な訳質を得るためには、辞書の語彙を十分にした上で文の意味を正しく認識することが必要である。

日本語文の文末構造は、述語に後接して、時制、判断、否定、話し手の態度など、広義の様相情報を与える助動詞、終助詞およびそれらに相当する連語がいくつか接続した形になっており、この文末構造を的確に訳出するシステムは無いと思われる。

本論文では開発中の日本語述部の日英翻訳システムの現状について報告する。本論文の構成は以下の通りである。まず第2章で、日本語述部の構造及びNPS(Non-Propositional semantic Structure: 非命題的意味構造)について述べる。第3章では、NPSを中間言語として採用した翻訳システムの動作及び現時点での性能、問題点について言及する。第4章では、NPSの曖昧さを軽減する副詞、動詞に着目した、NPSの生成精度向上の試みについて述べ、第5章でまとめと今後の課題を論じる。

\* 平成 21 年 6 月 26 日受付

\*\* 電子情報工学科

\*\*\* 工学研究科情報・制御システム工学専攻



で整理している。木構造における葉ノード (leaf node) が意味関数に対応している。

NPS(Non-Propositional semantic Structure) は次式のような入れ子型で表すことができる。

$$M_m [M_{m-1} \dots [M_2 [M_1 [S ]]] \dots ]$$

但し、 $S$  は命題的な骨格文 (命題的内容を表わす部分)、 $M_i (1 \leq i \leq m)$  は助述表現が与える意味関数である。助述表現が文の述部にいくつも並んだ複雑な文末表現の場合でも、意味関数との対応をとることにより NPS を求めることが可能である。NPS は、構造のシンプルさと同時に対応可能な表現の多様さから工学的に重要な性質<sup>1</sup>を持っていると考えられ、さらに言語依存性も無いとされるため、言い換えや機械翻訳を行う際の中間表現として有効であると考えられる。

### 3. 構築したシステム

#### 3.1. 翻訳システムの種類

現在の翻訳システムは、統計ベース方式 (Statistical Machine Translation; SMT)、用例ベース方式 (Example-based Machine Translation; EBMT)、ルールベース方式 (Rule-based Machine Translation; RBMT) に大別される。統計ベース方式および用例ベース方式は、対訳コーパスに挙げられるような原言語と目的言語のペア (用例) を大量に蓄積し翻訳を行う方式である。いずれも高性能の出力結果を得るためには必然的に大量の用例が必要となり、実用化されているものは少数のようである<sup>2</sup>。一方ルールベース方式は、トランスファー方式と中間言語方式に大別される。トランスファー方式は、形態素解析、構文解析、意味解析からなる解析ステップ、変換ステップ、生成ステップを持つ。トランスファー方式は、市販の日英・英日翻訳システムでは多く採用されているが、翻訳システムが扱う対象の言語数を  $n$  とした場合  $n*(n-1)$  種類の変換ステップを必要とするデメリットがある。一方、中間言語方式は、翻訳システムが扱う対象の言語数を  $n$  とした場合  $2n$  種類の変換ステップで済み、翻訳システムが扱う対象の言語数が増えるほど、必要とする変換ステップ数において、中間言語方式を採用するメリットが大きくなる。ただし、中間言語方式では、中間言語の仕様は言語に依存するため、中間言語の仕様をどのように決めるかが問題である。日本語述部の英訳において、NPS に言語依存性がないと考えられること、意味関数の種類が 139 種類と多いこと、NPS が容易に生成できること、及び、NPS から英文叙述形が容易に生成できることから、NPS は中間

言語方式における中間表現として好都合である。構築したシステムは NPS を中間言語とみなした機械翻訳システムと位置づけることができる。

#### 3.2. システムの動作

構築した日本語述部の英訳システムの動作について述べる。システムへの入力日本語述部を仮定する。助述表現から意味関数への変換においては、表 1 に示すように、助述表現「のかなあ」、「ません」、「はじめる」は、それぞれ意味関数「疑問 1」、「否定 1」、「起動」が対応する。システムは、日本語述部から NPS を作成し、作成された NPS から変換規則を適用し英訳を出力する。

##### 3.2.1. NPS の作成

NPS の作成手順は、概略以下の通りである<sup>3</sup>。

- ① 入力文を拡張文節で分かち書き (形態素解析) する (首藤ら, 1979, 添島ら, 2003)。
- ② 述部内の助述表現を対応する意味関数に変換し、述語を基本形にする。
- (③ 意味関数の並びの順序を逆にする。)

例えば、「彼は戻らざるを得ないでしょう」の場合には、次のように NPS を作成する。

- ① 彼は / 戻ら / ざるを得ない / でしょう
- ② 彼は戻る / 必要性 7 / 推量 3
- ③ 推量 3 [ 必要性 7 [ “彼は戻る” ] ]

つまり、日本語においては、適切に助述表現を意味関数に変換することで NPS を得ることができる。日本語述部から NPS を生成する実験は (Shudo et al., 2004) で述べられており、再現率約 0.90、適合率約 0.38 が得られている。

##### 3.2.2. 変換規則

次に、NPS 中の各々の意味関数に対し、「変換規則」を適用し、英訳を得る。変換規則とは、例えば、意味関数が「過去時制」であるときは「訳語動詞を過去形に変換」また、「進行中」は「動詞を現在分詞形にし、その前方に be 動詞を挿入する」という規則をいい、各々の意味関数に対応した変換規則を作成した。意味関数に対する変換規則は 1 個とは限らない。例えば、意味関数が“必要性 [ X ]”である場合には、“have to X”及び“must X”の 2 種類の変換規則を対応させている。作成した変換規則は 279 個であり、意味関数は 139 種であることから

1 例えば、助述表現を「発話者の主観表現」とみなすことにより、NPS を主観情報処理に応用することも考えられる (本田ら, 2008)。

2 統計的機械翻訳を採用している google 翻訳がある (google 翻訳, 2008)。

3 述部以外に現れる非命題的意味を表わす単語 (副詞や副助詞の一部など) は、今回は取り扱わない。

表2 実験結果

意味関数の個数		1	2	3	4	全体
市販翻訳ソフトA		0.86	0.75	0.62	0.57	0.83
市販翻訳ソフトB		0.84	0.72	0.48	0.71	0.81
構築システム	適合率	0.30	0.22	0.19	0.14	0.27
	再現率	0.85	0.76	0.71	0.77	0.82

ら、意味関数1個あたりの変換規則数は平均約2.0個である。

### 3.2.3. 変換過程

変換過程は次のようになる。例えば、日本語述部「行かなければならなかった」の場合のNPSは過去[必要性[行く]]となる。まず、述語「行く」を「go」に英訳し、次に「必要性」、「過去」の変換規則を順次適用する。

過去 [必要性 [go]]  
 = 過去 [have to go]  
 = had to go

この変換過程で、日本語述部に対応した英訳「had to go」が得られる。しかし、過去[必要性[行く]]には、次のような変換も考えられる。

過去 [必要性 [go]]  
 = 過去 [must go]  
 =  $\phi$

ここで $\phi$ は変換が存在しないことを表す。助動詞 must には過去形が存在しないので変換は不可能であり、英訳は出力されない。変換過程においては意味関数に対する変換規則を全て適用・変換し、妥当な全ての英訳を出力する。

## 3.3. 予備実験

### 3.3.1. 実験手法

EDR 日本語コーパス (EDR,1996) から、述語に助述表現が1つ以上後接した文末表現を無作為に2969個抽出し、これらを学習データとみなして人手で変換規則を作成した。

次にEDR 日本語コーパス中から、述語に助述表現が1つ以上後接した文末表現であり、かつ、学習データ用

の2969文とは異なる959個を無作為に抽出し英訳を行い、市販の日英翻訳ソフトが出力する英訳と比較する。比較に際しては、NPS中の意味関数の個数ごとに適合率、再現率を用いる<sup>4</sup>。客観的に評価を行うため、英訳出力の正誤判定は英語に精通した評価者が行う。

### 3.3.2. 実験結果

実験結果を表2に示す。

### 3.3.3. 考察

英訳結果から、いずれの場合もNPS中の意味関数の個数が増えるほど性能が低下していることが読み取れる。一方、構築したシステムの再現率には大きな低下は見られず、特に意味関数を4個含むような日本語述部に対する英訳の再現率は0.77と市販翻訳ソフトの値を上回った。これは、NPSが入れ子型構造であることのメリットが表れていると考えられ、特に、ブログなどに代表されるWebテキストなど意味関数を多く含むと考えられる日本語文の英訳を行うには好都合であると考えている。表3は、直接日本語述部から英語叙述形を生成させる実験結果であるが、NPSから英語叙述形を生成させる過程での適合率は0.72、再現率は0.91であり、NPS生成を別にした場合の本モデルによる英訳自体の性能は良いと考えられる。しかし、NPSから英語叙述形を生成する過程で動詞による変換規則の適用条件は考慮していない。例えば「Vていく」が日本語述部とすると、生成されるNPSは「持続2[V]」となり、「持続2[V]」に対する変換規則は「go on V-ing」<sup>5</sup>と設定しているため、例えば「乗り換えていく」を日本語述部として入力した場合には「go on transferring」と誤訳が出力されることになる。そのため、変換規則の適用条件をより詳細に検討する必要がある。また、構築したシステムにおいて適合率が大幅に低下しているのは、主として日本語述部から生成されるNPSの曖昧さの増大が原因であると考えられる。

4 構築システムは複数の英訳を出力する。一方、市販の翻訳ソフトでは出力は1つしかないため、再現率も適合率も同一値となる。

5 これ以降、「V-ing」は「V」の分詞、「V-en」は「V」の過去分詞であることを表すものとする。

表3 (首藤ら, 1977) の日本語動詞の分類

α相：所要時間が零とみなされるか、意識されないもの(死ぬ、食べはじめる、など)
β相：ある時間接続するとみなされ、開始と終了の時間が指示できるもの(読む、食べるなど)
γ相：長時間接続されるとみなされ、開始の時点は(多くの場合)指示できるが、終了の時点 を明確に指示することは難しいもの(肥える、愛する、など)

4. NPS の曖昧さ軽減

4.1. 助述表現「ている」の意味関数

NPS を介した日本語述部の英訳に関しては、生成される NPS の曖昧さをどのように軽減するかが、性能向上のために重要となる。意味関数の曖昧さの多い助述表現として「ている」や「れる」「られる」がある。そのため、EDR 日本語コーパスからランダムに抽出した助述表現を含む 2717 文に対し、「ている」、「れる」「られる」をどの程度含むかを調査したところ、1043 文に「ている」が、567 文に「れる」「られる」が含まれていた。そこで、本論文では「ている」に限定し意味関数の曖昧さを軽減させることを目標とする。助述表現「ている」には意味関数「状況」「完了後」「長期継続中」「進行中」の 4 種が対応するが、特定の単語の存在や述部動詞の種類によって、ある程度意味関数の曖昧さ軽減ができると考えている。例えば、文中に「今」があれば「進行中」、「ずっと」があれば「長期継続中」または「状況」に、また、述語動詞の種類によっては、例えば「愛している」の「ている」は「状況」、「死んでいる」の「ている」は「完了後」と決定できる。本論文では意味関数を決定できる副詞や動詞の種類により NPS の曖昧さを軽減することを考える。ここで、それぞれの意味関数の英訳は、英語動詞を「V」とした場合、状況は「V」、完了後は「have V-en」、長期継続中は「have been V-ing」、進行中は「be V-ing」となる。

4.2. 動詞の分類

動詞の意味を考える上で、アスペクト情報(相情報)は重要であることは、(Vendler,1967), (金田一,1950), (首藤ら,1977)の研究などによって議論されている。

(Vendler,1967)では、英語動詞をアスペクト情報によって、状態動詞、到達動詞、活動動詞、達成動詞の4種に、また、(金田一,1950)では、状態、行為、変化といったアスペクトに注目し、動詞に「ている」が付くかどうか、そして「ている」が付く場合にはどのような意味になるかの観点から日本語動詞を、状態動詞、継続動詞、瞬間動詞、第4種の動詞の4種に分類している。また、(首藤ら,1977)では、動詞をアスペクトによって、α相、β相、γ相の3種に分類している。本研究では、(首藤ら,1977)の分類を元に日本語動詞リストを作成した。日本語動詞の分類を表3に示す。

表3におけるα相、β相、γ相は、それぞれ、瞬時相、一般相、継続相に対応する。「ている」が付いたときのとりうる意味関数は動詞の種類によって異なる。動詞の種類に対して、とりうる意味関数との対応を表4に示す。例えば、「死んでいる」では、「死ぬ」は瞬時相動詞であるから対応するNPSは完了後[死ぬ]となり「have died」と訳され、「食べている」では、「食べる」は一般相動詞であるから対応するNPSは進行中[食べる]、長期継続中[食べる]、完了後[食べる]となりそれぞれ「be eating」、「have been eating」、「have eaten」と訳される。また「愛している」では、「愛する」が継続相動詞であるから対応するNPSは状況[愛する]となり「love」と訳される。本研究では、386個の動詞からなる動詞リストを作成した。動詞リストにおける相の内訳は表5の通りである<sup>6</sup>。

4.3. 時を表す副詞の分類

動詞による意味関数の曖昧さ軽減に加えて副詞による意味関数の曖昧さ軽減も行うことが出来れば、生成されるNPSの精度向上が更に期待できる。本研究では、文

表4 動詞の分類

アスペクト	「ている」が付いたときのとりうる意味関数			
	状況	完了後	長期継続中	進行中
瞬時相	×	○	×	×
一般相	×	○	○	○
継続相	○	×	×	×

6 複数の相を持ちえる動詞が存在するためこれらの合計は386個を超える。

表5 動詞の相の内訳

動詞の相	数
瞬時相	101
一般相	230
継続相	100

表6 時を表す副詞の分類

種類	例	数
過去の時制的副詞	去年, 前年	44
現在の時制的副詞	今, 現在	10
未来の時制的副詞	来年, 翌年	18
時間幅副詞(1)	いつも	2
時間幅副詞(2)	今日	22
時間幅副詞(3)	よく, 時々	10
時間幅副詞(4)	まだ, 今だ	11

中に現れたときにどのような意味関数をとりうるかの観点から、時を表す副詞を先ず時制的副詞と時間幅副詞の2種に分類し、さらに、時制的副詞を過去、現在、未来の3種、時間幅副詞を4種に分類した。時間幅副詞は、大雑把には反復性の有無、及び動詞からみた時間幅の大きさをもとに分類した。表6に種類及び例、副詞の個数を示す。

時を表す副詞による意味関数の曖昧さ軽減は、基本的には動詞の相によりとりうる意味関数をさらに絞り込むアプローチによって行う。例えば「去年食べている」は

過去の時制的副詞「去年」の存在により助述表現「ている」の意味関数は「完了後」に絞り込まれる。時制的副詞と動詞の相によりとりうる意味関数の組合せを表7に示す<sup>7</sup>。同様に、時間幅副詞の場合を表8に示す<sup>8</sup>。時間幅副詞においては、「死ぬ」などの瞬時相の動詞が、意味関数「長期継続中」をとりうる可能性があるのが特徴である。瞬時相の動詞は所要時間が零であるとみなされるため、時間幅を持つ表現と共起する場合には、動作主体が複数であること、動作が反復的に行われていること、または、動詞の意味が比喩的であることなどが考えられる。本研究では、117個の副詞からなる副詞リストを作成した。

4.4. 実験

EDR 日本語コーパスから、述語に助述表現が1つ以上接した文末表現を無作為に2711個抽出し、既存のシステム、動詞リストのみを使った場合、動詞リストと副詞リストの両方を使った場合の性能をそれぞれ比較した<sup>9</sup>。評価基準として適合率、再現率、F値を用いた。実験結果を表9に示す。

また、日本語文末表現に「ている」を含む文のみの実験結果を表10に示す。

4.5 考察

提案システムの動詞のみの結果は、既存のシステムと比べて適合率が上がったことにより、F値も向上していることが分かる。「ている」を含む文のみに限定すると動詞のみの場合でF値が8ポイント上昇しており、動詞による助述表現「ている」の意味関数の曖昧さ軽減は、

表7 時制的副詞の種類及びとりうる意味関数

時制的副詞の種類別	動詞の相	例文	とりうる意味関数
過去	瞬時相	去年死んでいる	完了後
	一般相	去年食べている	完了後
	継続相	去年知っている	状況
現在	瞬時相	今, 死んでいる	完了後
	一般相	今, 食べている	進行中, 完了後, 長期継続中
	継続相	今, 知っている	状況
未来	瞬時相	来年, 死んでいる	完了後
	一般相	来年, 食べている	進行中, 完了後, 長期継続中
	継続相	来年, 知っている	状況

7 過去の時制的副詞の場合、「去年食べていた」など過去の意味関数が含まれる述部には適用しない。

8 現段階では、φは存在しない組合せであることを示している。

9 動詞リストとのマッチングは、chasenにより述部を形態素解析し、述部動詞の終止形にしたものを行った。また、副詞リストとのマッチングにおいても chasen を使い、「前田さん」のような表現とのマッチングを避けるため、単語単位でマッチングを行った。

表 8 時間幅副詞の種別及びとりうる意味関数

時間幅副詞の種別	動詞の相	例文	とりうる意味関数
時間幅副詞(1)	瞬時相	いつも死んでいる	長期継続中
	一般相	いつも走っている	長期継続中
	継続相	いつも愛している	状況
時間幅副詞(2)	瞬時相	今日死んでいる	完了後
	一般相	今日走っている	進行中, 完了後
	継続相	今日愛している	φ
時間幅副詞(3)	瞬時相	毎日死んでいる	長期継続中
	一般相	毎日走っている	長期継続中
	継続相	毎日愛している	φ
時間幅副詞(4)	瞬時相	まだ死んでいる	長期継続中
	一般相	まだ走っている	進行中, 長期継続中
	継続相	まだ愛している	状況

表 9 実験結果

	既存のシステム	提案システム	
		動詞のみ	動詞と副詞
適合率	0.54	0.61	0.61
再現率	0.94	0.89	0.88
F 値	0.69	0.73	0.72

表 10 「ている」を含む文のみの実験結果

	既存のシステム	提案システム	
		動詞のみ	動詞と副詞
適合率	0.39	0.53	0.53
再現率	0.97	0.77	0.75
F 値	0.55	0.63	0.62

ある程度妥当であると言える。しかし、再現率は 97% から 77% と大幅に下落している。これは、動詞の多義性によって、動詞の分類にゆれが生じたためであると考えられる。例えば「取る」という動詞では「食事を取っている」と「立場を取っている」のように動詞に係る名詞によって相が変化してしまうことがある。今後は動詞に係っている名詞も含めた相情報の付加が必要だと考えられる。また、動詞だけでなく副詞も組み合わせた場合、係り受けの誤りや、「彼の前に立っている」などのように、時間を意味していない副詞による誤った絞込みが行われた結果、動詞のみの場合に比べて再現率が下がったと思われる。今後の課題として、係り受け関係の導入及び動詞に係る副詞の意味を特定する機構が必要である。

### 5. おわりに

中間言語を用いた機械翻訳は、統計ベース方式や用例ベース方式のように、知識の獲得に膨大なデータを必要とせず、また、トランスファー方式のように、原言語と目標言語に依存した構文木を作成する必要がない特徴をもつ。日本語述部に限定すれば、日本語から NPS への生成、及び、NPS から英訳への生成は比較的容易であるが、生成される NPS の精度が低い問題があった。この問題に対処するために、本論文では、「ている」を含

む文末表現により生成される NPS の絞込みを動詞及び時間を表わす副詞を使って検討した。F 値が向上したことから、ある程度絞込みは適正に行われたと言える。

今後の課題として、副詞による絞込みの効果が見られなかったため、係り受け解析や意味解析を行うことによって、曖昧さを低減できる副詞を正確に特定する必要がある。また、本研究では「ている」に限定して絞込みを行ったが、曖昧さの多い表現として他に「れる」「られる」があり、「ている」の場合と同様に、動詞及び特定の単語の存在による曖昧さ軽減ができると考えている。また、意味関数の順序の制約をとり入れることも考えている。例えば、「ている」と「られる」を共に含む「食べていられる」と「食べられている」は意味が異なると思われる。現段階では、(cinque,1999)による意味関数の順序の制約をとりこむことを考えている。本論文では、中間言語として生成される NPS の精度向上に着目したが、翻訳システム全体を考えると、NPS から英訳を出力する際の精度も重要になってくる。本研究を通じて、日本語動詞と英語動詞の間に相のずれが存在することを確認した。例えば、「知る」は瞬時相動詞、「know」は継続相動詞であると考えられ、「知る」の英訳は「get to know」、「知っている」は「know」と、それぞれ訳す必要があると考えている。また、変換規則を充実させ

るため、大量の対訳コーパスによる変換規則の自動学習、ブログなどの Web テキストなどを含めた変換規則の人手の学習が必要であると考えている。実験の考察でも述べたが、変換規則の適用条件の詳細な検討も必要であり、さらに変換規則をどう記述するかも検討が必要である。また、受身や使役などの意味関数に対応する変換規則は複雑になり、変換は述部内にとどまらず日本語文全体に対して行う必要がある。この場合には、文中の格関係を把握する必要がある (首藤ら, 1979, 古賀ら, 2003, 古賀ら, 2002)。また、評価に際しては BLEU 値を用いるなど、客観的基準を用いることが必要であると考えている。これらの問題を解決した、高精度の中間言語方式による機械翻訳システムの登場は、これからの国際社会において大きく貢献できると考えている。

### 参 考 文 献

- chasen. <http://chasen-legacy.sourceforge.jp/>.  
EDR 日本語コーパス. 1996. 独立行政法人情報通信研究機構, [http://www2.nict.go.jp/r/r312/EDR/J\\_index.html](http://www2.nict.go.jp/r/r312/EDR/J_index.html)
- google 翻訳. 2008. [http://translate.google.co.jp/translate\\_t](http://translate.google.co.jp/translate_t)
- Guglielmo Cinque. 1999. *Adverbs and Functional Heads*. OXFORD UNIVERSITY PRESS
- Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi and Kenji Yoshimura. 2004. MWEs as Non-propositional Content Indicators. *The Proc. of the ACL2004 Workshop on Multiword Expressions: Integrating Processing*: pp.32-39.
- Toshifumi Tanabe, Kenji Yoshimura and Kosho Shudo. 2001. Modality Expressions in Japanese and Their Automatic Paraphrasing, *The Proc. of the NLPRS2001*: pp.507-512.
- 本田聖晃, 田辺利文, 吉村賢治, 首藤公昭. 2008. 非命題的意味解析のための日本語文末表現意味体系. 主観表現処理の最前線シンポジウム, 電子情報通信学会言語理解とコミュニケーション研究会, NLC-2007-94, pp.39-44
- 金田一春彦. 1950. 「国語動詞の一分類」『言語研究』15. (金田一春彦 (編) 『日本語動詞のアスペクト』 (むぎ書房, 1976) に再録)
- 古賀基和, 添島創, 田辺利文, 吉村賢治, 首藤公昭. 2002. 標準形変換規則によるデータスパースネスの解消, 電気関係学会九州支部第 55 回連合大会
- 古賀基和, 田辺利文, 吉村賢治, 首藤公昭. 2003. 日本語文における態の処理について - 格変換と補文の抽出 -, 福岡大学工学集報 70 号, pp.107-112
- 首藤公昭, 鶴丸弘昭, 吉田将. 1977. 日英機械翻訳のための述部処理システム, 電子通信学会論文誌, J60-D/10, pp.830-837
- 首藤公昭, 植原斗志子, 吉田将. 1979. 日本語文の標準形変換に関する考察. 昭和 54 年度電気四学会九州支部連合会大会論文集
- 首藤公昭, 植原斗志子, 吉田将. 1979. 日本語の機械処理のための文節構造モデル, 電子通信学会論文誌, J62-D/12, pp.872-879
- 添島創, 田辺利文, 吉村賢治, 首藤公昭. 2003. 日本語文分かち書きのための新しい枠組み, 福岡大学工学集報 70 号, pp.99-106
- 田辺利文, 本田聖晃, 高橋雅仁, 小山泰男, 吉村賢治, 首藤公昭. 2006. 日本語文末表現の取り扱いについて, FIT2006 第 5 回情報科学技術フォーラム.
- 田辺利文, 吉村賢治, 首藤公昭. 2001. 日本語モダリティ表現とその言い換え, 言語処理学会第 7 回年次大会. ワークショップ論文集.
- Zeno Vendler. 1967. *Linguistics in Philosophy*. Cornell University Press.