形態論的制約を用いたオンライン未知語獲得

村脇 有吾†・黒橋 禎夫†

日本語の形態素解析における未知語問題を解決するために、オンライン未知語獲得という枠組みと、その具体的な実現手法を提案する。オンライン未知語獲得では、形態素解析器と協調して動作する未知語獲得器が、文が解析されるたびに未知語を検出し、その可能な解釈の候補を列挙し、最適な候補を選択する。このうち、列挙は日本語の持つ形態論的制約を利用し、選択は蓄積した複数用例の比較により行う。十分な用例の比較により曖昧性が解消されると、解析器の辞書を直接更新し、獲得された未知語が以降の解析に反映される。実験により、比較的少数の用例から高精度に未知語が獲得され、その結果形態素解析の精度が改善することが示された。

キーワード：未知語、辞書、語彙獲得、形態素解析

Online Acquisition of Japanese Unknown Morphemes using Morphological Constraints

Yugo Murawaki† and Sadao Kurohashi†

To solve the unknown morpheme problem in Japanese morphological analysis, we propose a novel framework of online unknown morpheme acquisition and its implementation. In online unknown morpheme acquisition, an unknown morpheme acquirer, which works in concert with the morphological analyzer, detects unknown morphemes from each segmented and POS-tagged sentence, enumerates its possible interpretations, and selects the best candidate. In enumeration, morphological constraints of the Japanese language are utilized, and selection is done by comparing multiple examples kept in storage. When the number of examples being compared is large enough for disambiguation, the acquirer directly updates the dictionary of the analyzer, and the acquired morpheme will be used in subsequent analysis. Experiments show that unknown morphemes are acquired from relatively small numbers of examples with high accuracy and improve the quality of morphological analysis.

Key Words: unknown morpheme, dictionary, lexicon acquisition, morphological analysis

1 はじめに

形態素解析は、文を形態素列に分割し、各形態素に品詞をタグ付けするタスクである。形態素解析は自然言語処理における基盤技術であり、構文解析や情報検索といった応用を実現するうえで高い精度の達成が不可欠となる。日本語の形態素解析では、あらかじめ定義された辞書

† 京都大学大学院情報学研究科, Graduate School of Informatics, Kyoto University
を用いる手法が高い精度を達成している（Kurohashi, Nakamura, Matsumoto, and Nagao 1994; 浅原, 松本 2002; Kudo, Yamamoto, and Matsumoto 2004）。この手法では、入力文は辞書引きにより得られた形態素のラティスに展開され、ラティス中の最適なパスが出力として選択される。しかし、辞書に基づく形態素解析には、辞書にない形態素（未知語）の解析を誤りやすいという問題がある。例えば、形態素解析器JUMAN1は、デフォルトの辞書を用いると、未知の動詞「ゲゲル」を誤って「ゲグ」と「る」に分割する。この未知語問題は、未知語を解析用の辞書に追加することで解決する。しかし、人手による辞書登録はコストがかかるため、計算機による自動化が望まれる。

人手によらない未知語問題への解決策として、2通りの手法が提案されている。ひとつ目の手法では、形態素解析における未知語モデルを改良する（Nagata 1999; 内元, 関根, 井佐原 2001; Asahara and Matsumoto 2004; 東, 浅原, 松本 2006）。日本語の形態素解析で広く用いられる未知語モデルは、字種に基づく簡単なヒューリスティクスだが、代わりに統計や機械学習に基づく未知語モデルを導入すると未知語同定の精度が向上する。二つ目の手法では、テキストから未知語を自動獲得し、形態素解析用の辞書を拡張する（Mori and Nagao 1996）。二つの手法を比べると、前者は入力文中の個々の未知語を同定しようとするのに対し、後者は同じ未知語のテキスト中での複数の使われ方を比較できるという点で異なる。複数の使われ方の比較は未知語の同定に効果的と考えられる。例えば「ようつべ」（YouTubeのスラング）という形態素を知らないまま、「ようつべって…」という文を解釈したいとする。このとき、「ようつべ」は、未知の名詞以外にも、未知の動詞「ようつべる」とも解釈できる、いずれが正しいか判断がたい。同様に、別の文「ようつべとは…」について、名詞「ようつべ」の他に、動詞「ようつぶ」の命令形とも解釈できる。しかし、両者を見比べると、2文とも名詞「ようつべ」で解釈できるところから、名詞という解釈がより自然だと推測できる。従って、本論文では後者の手法を採用する。ただし、両者は対立するものではなく、組み合わせることで、より高い解析精度が得られるようになると期待できる。

未知語獲得の従来手法はバッチ処理であり、コーパスをソートしてすべての部分文字列を調べる（Mori and Nagao 1996）。しかし、この手法は効率が悪い。なぜなら高頻度の形態素のほとんどが解析用の辞書に登録済みであり、一般に出現頻度でコーパスの90%以上を網羅している。こうした既知の形態素を改めて獲得しても無駄になる。これに対し、提案手法では、辞書に登録されていない形態素のみを獲得対象とする。

従来研究は、資源の制約から、主に小規模な新聞記事を対象に行われてきたが、近年、ウェブの出現により大規模なテキストが入手可能となっている。それに伴い、自然言語処理の様々な分野でデータの大規模化による性能向上が報告されている（Banko and Brill 2001; Brants, Popat, 1 http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html
Xu, Och, and Dean (2007). しかし、未知語獲得は、データの大規模化が単純に解消する性質の問題ではない。未知語の中には、「ブログ」のように高頻度ながら登録が漏れているものもあるが、大がいわゆるロングテールに属する低頻度の形態素である。こうした形態素の表現するテキストには偏りがあるだけでなく、データを増やすだけでは、次々と新たな未知語が出現していく。従って、とにかくデータを増やすことから未知語を獲得するよりも、個々の未知語を候補に着目し、それが獲得されるまでデータを読み込む方が自然である。そもそも、未知語の同定のために、何千、何万もの使われ方を調べる必要なく、直感的には、ほとんどの場合、10件程度を見比べればほぼ明らかではないかと思われる。

本論文では、オンライン未知語獲得という枠組みと、その具体的な実現手法を提案する。オンライン未知語獲得では、バッチ処理ではなく、逐次的に入力されるテキストから未知語を獲得する。形態素解析器自体は、通常通りテキストを文単位で解析し、形態素列を出力する。異なる点は、解析の結果で未知語獲得器が動作することである。具体的には、解析された文から未知語を抽出し、適当な時点で形態素解析器の辞書を更新する。これにより獲得された未知語が形態素解析に反映される。

オンライン未知語獲得では、獲得開始時に対象コーパスを決める必要がない。そのため、例えば、クローラが毎日新たなページを取得するという設定でも、この差分のみから未知語が獲得できる。

オンライン未知語獲得は、検出、列挙、選択のサブタスクに实现される。このうち、列挙は日本語の持つ形態論的制約を利用し、選択は蓄積した複数用例の比較による。実験により、比較的少数の用例から高精度に未知語が獲得され、その結果形態素解析の精度が改善することが示された。

本論文の構成は次の通りである。2章で未知語獲得タスクを整理し、3章でオンライン未知語獲得の枠組みを提案する。4章では、オンライン未知語獲得の実現手法のうち、列挙と選択を説明する。5章で実験結果を報告し、6章で関連研究、7章で結論を述べる。

2 未知語獲得タスク

未知語獲得とは、未知語について、テキスト中の一つ以上の用例から辞書項目を収納的に生成するタスクである。ここでは、辞書項目は辞書の項目として記述される形態素であり、テキスト中に出現したその形態素を用例とよぶ。例えば、未知語「ググる」について、「なんとなくググってみた。」「ググらずに答える。」といったテキスト中の用例から辞書項目を生成する。ただし、個々の用例の解釈には曖昧性があり、そうした曖昧性を消解することによって辞書項目が生成される。

辞書項目の生成には語幹と品詞の同定が必要となる。「ググる」の例にあるように、動詞や形
容詞は文法的役割に応じて形態変化を起こすが，この形態変化は活用という概念によって処理される．活用する形態素は語幹と語尾からなる．語幹は不変だが，語尾は活用に応じて変化する．例えば，「ゲゲって」は語幹「ゲゲ」と語尾「って」からなる．名詞は活用せず，語幹のみからなる．品詞は形態素解析の定義されたものに基づく．ただし，既存の品詞は人手での付与が前提となっている．形態，構文，意味レベルの情報が混在している．未知言語のタスクにおいて，いきなり意味レベルの情報を獲得するのは難しいため，本論文では，まず形態レベルの情報の獲得を目指す．そのために，品詞分類を整理する．以下の説明は形態素解析器 JUMAN が採用する品詞体系に基づく．品詞体系の設定方法には様々な流儀があるため一般化が難しいが，少なくとも ipadic2の品詞体系でも同様の議論が成立立つことは容易に想像できる．品詞は「品詞」，「品詞細分類」，「活用型」，「活用形」の4種類からなる．「品詞」には「名詞」，「動詞」，「形容詞」などがある．「名詞」の「品詞細分類」には，「普通名詞」や「変名詞」の他，固有名詞用の「固有名詞」，「組織名」，「地名」，「人名」などがある．しかし，固有名詞と普通名詞の識別は，形態レベルの文法的な情報のみでは困難なので，本論文では，便宜的に固有名詞も「普通名詞」とみなす．用言の「動詞」と「形容詞」には「品詞細分類」は設定されていない．代わりに活用を扱うために活用型と活用形が与えられる．活用型は活用のタイプに基づく分類であり，活用形は個々の具体的な活用形態を指す．例えば，「ゲゲる」の活用型は「子音動詞ラ行」で，「ゲゲって」の活用形は「子音連用ラ形」，「ゲゲら」は「未然形」となる．未知言語のタスクにおける品詞は，「品詞」，「品詞細分類」，「活用型」の適当な組である．簡単のために，名詞については「品詞細分類」，動詞と形容詞については「活用型」で呼ぶ．例えば，「ゲゲる」の品詞は「子音動詞ヲ行」となる．基本語彙は既に人手により辞書登録されているので，獲得対象をオープンクラスの品詞に絞り込む．つまり，「来る」などの不規則動詞や助動詞，助動詞などの付属語は獲得対象から除外される．本論文では，名詞，動詞，および形容詞を獲得対象の品詞とする．副詞もオープンクラスとみなせるが，今回は明示的な獲得対象としない．副詞と名詞の識別も，形態レベルの情報だけでは困難だからである．副詞の認識は今後の課題とする．以上をまとめると，獲得対象の品詞は表1の15種類となる．

| 表1 獲得対象の品詞 |
|------------------|
| 母音動詞，子音動詞カ行，子音動詞ガ行，子音動詞サ行，子音動詞タ行，子音動詞ハ行，子音動詞マ行，子音動詞ラ行，子音動詞ヲ行，サ変動詞，サ変動詞，イ形容詞，ナ形容詞，普通名詞，サ変名詞 |

2 http://sourceforge.jp/projects/ipadic/
形態素の単位認定基準、つまりある言葉が1形態素か否かは自明でない。例えば、「ミンクジラ」のように構成的な名詞や「宣べ伝える」のような複合動詞を1形態素とするか分割すべきか明らかでない。実際、人手で整備された既存の形態素解析用の辞書も、単位に一貫性があるとは言い難い。他の単位認定基準としては、現代日本語書き言葉均衡コーパスが人間の作業者向けに詳細な基準を設けている（小椋、小磯、富士池、原 2008）。しかし、この基準は煩雑で、しかも意味レベルの情報も利用しているため、プログラムに落とし込んで未知語の自動獲得に利用することは困難である。本論文では、厳密な単位認定にはこだわらないとする。

3 オンライン未知語獲得

3.1 システム構成

未知語獲得タスクに対して、我々はオンラインによる解析を提案する。図1にオンライン未知語獲得のシステム構成を示す。形態素解析器自体は、通常通り入力文に対して形態素列を出力する。ただし、辞書として、人手で整備した基本語彙辞書の他に、自動獲得辞書も用いる。形態素解析の裏では未知語獲得器が動く。獲得器は、形態素解析器が出力する形態素列を文ごとに受け取り、そこから未知語を抽出する。獲得器は、適当な時点で未知語を獲得し、形態素解析器の自動獲得辞書を更新する。辞書更新により未知語獲得が以降の解析に反映される。

獲得器には高い精度での未知語獲得が要求される。獲得された未知語の辞書へのフィードバックに人手が介在しないが、誤獲得が解析に悪影響を及ぼすことは避けたいからである。

獲得器は、未知語の用例を蓄積することで、それまでに解析されたテキストを獲得に利用で
きる。未解析のテキストは獲得に利用できないが、見方を変えれば、次に読むテキストをあらかじめ決める必要がないことを意味する。従って、獲得の都合に応じて対象テキストを動的に変更するという応用も可能である。

オンライン未知語獲得を実現するために、以下のサブタスクを設定する。

検出 各文の形態素解析結果から未知語の用例を検出する。

例挙 検出された各未知語用例に対して、語幹と品詞からなる辞書項目の候補を列挙する。

選択 各未知語用例に対して、最適な辞書項目の候補を選択する。選択は、過去に検出された用例を蓄積しておく、それら複数用例の比較により行う。比較される用例が増え、曖昧性が十分に解消できた時点で獲得し、形態素解析器の辞書を更新する。

未知語「ゲグル」の獲得を例にシステムの挙動を説明する。「ゲグル」は語幹「ゲゲ」と品詞「子音動詞ラ行」からなる。テキストを読み進めて、ある時点で文「なんとなくゲゲってみた。」が入ってきたとする。獲得者は、まず、この文の「ゲゲ」を手がかりに未知語用例を検出する。次に、この用例に対して考えられる辞書項目の候補を列挙する。辞書項目の候補としては、語幹「ゲゲ」と品詞「子音動詞ラ行」以外にも、同じ語幹で「子音動詞ヲ行」、語幹「ゲゲって」と品詞「子音動詞ヲ行」、語幹「なんとなくゲゲ」と品詞「子音動詞ラ行」なども考えられる。こうした複数の候補の中から正しい候補を選択する必要があるが、この1用例だけを覚えて正しい候補を判断しがたい。そこで獲得者は判断を保留し、用例を記憶に蓄えておく。さらにテキストを読み進めると、「ゲゲらずに答える。」という文が入力される。同様に検出と例挙を行ったのち、「ゲゲってみた」の用例を記憶から取り出して、「ゲゲらず」と比較する。すると、両者を共通に解釈できる辞書項目の候補は語幹「ゲゲ」と品詞「子音動詞ラ行」のみである。このように複数の用例を比較して曖昧性を解消する。比較する用例が増え、選択された候補が適当な条件を満たしたとき、その候補を獲得する。これにより、「ゲグル」が自動獲得辞書に追加される。

オンライン未知語獲得のサブタスクのうち、本論文では例挙と選択について詳述する。検出タスクについては簡単な手法を説明することとする。

3.2 未知語用例の検出

未知語検出は各文から未知語の用例を検出するタスクである。文は、形態素解析結果に基づく形態素列、または文字列として表現される。タスクの入力は、解析器が返す文の形態素列である。一方出力は、未知語用例に対応する文の部分文字列であり、その範囲を \([s_d, e_d]\) とする。ただし、\([s_d, e_d]\) が未知語用例の語幹の範囲 \([s_u, e_u]\) と厳密に一致する必要はない。辞書項目、つまり語幹と品詞の組の候補の列挙は次の例挙タスクで行う。どの程度の正確さで検出が必要かは例挙のアルゴリズムに依存する。4.3節で述べる例挙アルゴリズムは、語幹の境界候補の列挙を \(s_d\) を基点に行うので、検出範囲は \(s_u \leq s_d \leq e_u\) を満たす必要がある。
日本語において未知語用例の検出は自明なタスクではない。一番単純な検出手法として、既知語とテキストの文字列マッチングを行い、マッチしない箇所を検出するというもののが考えられる。しかし、日本語の単純な音韻体系がわずかわずか、多くの未知語に対して無関係な既知語がマッチし、検出漏れが起きる。この現象は、形態素解析器が持つ文法知識を利用することである程度抑えられる。形態素解析器は、入力文に対して、辞書引きと未知語処理により、出力すべき形態素の候補を列挙する。未知語処理により列挙される形態素候補を未定義語と呼ぶ。JUMANでは、字種に基づく簡単なヒューリスティクスが採用されている。例えば、カタカナの連続が一つの形態素候補とされる。これにより、未知語「ググる」を含む入力文「ググってみた。」に対して、未定義語「ググ」が形態素候補となり、これを含むパスが出力に選ばれる。従って、形態素解析結果中の未定義語 \( w_t \) を検出範囲 \([s_{w_t}, e_{w_t}]\) とする。ただし、\( s_{w_t} \) と \( e_{w_t} \) は、形態素 \( w_t \) の文字列表現における開始・終了位置である。

形態素解析を用いる検出手法でも検出されない未知語用例が存在する。例えば、「アブラハム」は「アブラ」（油）と「ハム」に分割され、「うざい」は「う」（卯／雨／鶴）と「ざい」（剣／在／材／罪／剣）に分割される。こうした過分割未知語の検出は今後の課題とする。

4 辞書項目の列挙と選択

4.1 列挙タスクと選択タスク

列挙は、検出された各用例に対して、文中の前後の文脈を利用して、考えられる辞書項目の候補を列挙するタスクである。辞書項目の候補は語幹と品詞からなる。ここで、語幹の同定は前方境界と後方境界の二つの同定を意味する。例えば、「なんとなくグッってみた」の場合、「なく」と「ググ」の間に前方境界が、「ググ」と「って」の間に後方境界が引かれる。そこで、辞書項目の候補を前方境界、後方境界、品詞の組で表現する。列挙される候補は、効率よく正解候補を選択するためには、なるべく数が少ないことが望ましい。

選択は、各未知語用例に対して、最適な辞書項目の候補を選択するタスクである。この際、検出済みの未知語用例を蓄積することで、複数の用例が比較できる。選択タスクの実現には、最適な候補を選択する基準と、最終的に獲得を判断するための終了条件が必要となる。

4.2 形態論的制約の利用

辞書項目の列挙において、候補絞り込みの手がかりとして形態論的制約を利用する。日本語は連語系であり、形態素は、その文法的な役割に応じて、接尾辞、助動詞、助詞などに後続される。この際、用言は後続する形態素に応じて活用形を変える。また、形態素同士の接続には品詞に応じて制約が働く。例えば、助詞「を」は、「走る」の基本連用形「走り」に後続して「走りを」という形は取り得るが、未然形「走ら」に後続して「走らを」とはならない。このよう
な連接に関する制限を形態論的制約と呼ぶ。

この形態論的制約を列挙に利用するためにサフィックスを導入する。サフィックスとは，語幹に後続し得る文字列であり，自立語の語尾（あれば）後続する付属語を連絡したものである。サフィックスの例を表2に示す。いま，ある文字列に対してあるサフィックスが後続したとする。そのとき，このサフィックスの直前が自立語の語幹の後方境界の可能性がある。

サフィックスの集合は生テキストから収集される。ここで，形態素解析が既知語について十分に高精度であることを利用する。具体的には，テキストを形態素解析し，既知語に後続するサフィックスを収集する。こうして集められたサフィックスを品詞ごとに集約する。いま，サフィックスが十分に大きなコーパスから収集されたとき，ある品詞に属する形態素の語幹に後続し得るサフィックスは，品詞に対応するサフィックス集合中のいずれかに限定される。従って，サフィックスを候補列挙に用いることで，後方境界と同時に品詞の候補が列挙できる。なおかつ，品詞候補を形態論的制約を満たすものに限定できる。ただし，一般に，サフィックスは複数の品詞に後続し得る。例えば，サフィックス「をも」は母音動詞にもサ変名詞にも後続できる。

サフィックスの収集には，Kawahara et al.の手法により編纂されたウェブコーパスを用いる (Kawahara and Kurohashi 2006)。ただし，予備実験により，この大規模コーパスでもサフィックスの異なり数が収束しないと判明した。「させられかねなかったろう」のような低頻度の長いサフィックスが存在するからである。そこで，サフィックスの最大長を5文字とし，それより長いサフィックスは先頭の5文字で統合する。実験では，約1億ページから約66万の異なるサフィックスを得た。サフィックスあたりの品詞数は平均で1.33であった。

### 4.3 サフィックスを用いた列挙手法

サフィックスを用いて辞書項目の列挙を行う。まず，列挙に利用する文中の前後の文脈，つまり前方境界と後方境界の探索範囲を文節を用いて限定する。文節については，構文解析器KNP3

| 品詞 | 基本形（例） | 語幹（例） | 語尾 | 活用形 | サフィックス（例） |
|------|-------------|------------|------|--------|------------------|
| 子音動詞 | 走る | 走 | ら | 未然形 | らず，らないで |
| 母音動詞 | あこがれる | あこがれ | お | 未然形 | ず，ないで |
| サ変名詞 | 希望 | 希望 | NIL | を | を，を |
| NIL | す | する | したら |

1 名詞は活用しないため，名詞の活用形として，直後の付属語の基本形を用いる。
が係り受け解析の前処理として文節まとめを行うので、その結果を利用して検出された未知語用例が属する文節、および最大で前後2文節を探索範囲とする。ただし、文頭、文末や句読点で探索を打ち切る。

後方境界と品詞の組の候補を図2のようにサフィックスを用いて列挙する。検出範囲の開始位置$s_d$から探索範囲の終端までの各位置で、サフィックスのマッチングを行う。サフィックスがマッチしたとき、サフィックス開始位置が後方境界の候補となり、サフィックスに対応する1個以上の品詞が候補となる。

長さの異なる複数のサフィックスがマッチした場合、以下の規則で採用するサフィックスを選択する。原則として長い候補を優先するが、サフィックスの終了位置が文節境界と一致しなければならない。ただし、サフィックスは最大5文字としているので、5文字のサフィックスがあれば無条件で採用する。

また、サフィックス以外の手がかりとして、以下を前方境界と後方境界の候補列挙に利用する。
- 文頭と文末
- 句読点や記号
- 「御」などの接頭辞
- 「首相」などの未尾要素
- KNPにより与えられる文節境界

これらの手がかりにより列挙される候補のうち、後方境界については、特殊な品詞“EOB”を与える。“EOB”はサフィックスなしに語幹単独で出現し得ることを示す。例えば、「グーグル」

図2 候補の列挙

---

3 http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html
4 ただし、ウェブコーパスの場合は、HTMLから文抽出を行うため、文頭や文末は文抽出誤りの影響を受ける可能性があり、あまり信頼できない。
などの名詞には句読点などが直接後続し得る。また、母音動詞は基本連用形（名詞化）が語幹と同形なので、語幹単独で出現し得るとみなせる。一方、「ググる」などの子音動詞の行は語幹単独では出現しない。「EOB」は選択タスクにおいて語幹単独で出現し得る品詞に展開される。

4.4 用例の蓄積

辞書項目の選択には、それまでに検出された複数の用例を利用する。具体的には、新たに入ってきた用例について、その用例と同じ辞書項目を表す可能性のある用例群を記憶から取り出して比較する。ただし、真に同じ辞書項目を表す用例のみを取り出すのは難しいので、ひとまず前方境界を共有する用例群を取り出し、後の処理で絞り込みを行う。また、獲得に至らなかった用例は記憶に追加し、獲得時には使われた用例群を削除する。

用例の効率的な管理のためにトライを利用する。各用例の格納は、前方境界の候補数だけ行う。トライのキーとして、各前方境界候補と、それよりも右で最左の後方境界候補に挟まれた文字列を用いる。例えば、図2の用例に対して、「ググ」と「何となくググ」をキーとして2箇所に格納する。用例取り出し時にはキーを使ってトライをたどり、途中のノード、およびキーの末端ノードの子孫に格納された用例群を取り出す。

4.5 辞書項目の選択

辞書項目の候補、つまり前方境界、後方境界および品詞の候補のうち、最適な候補の選択を記憶から取り出された用例群の比較により行う。図3に選択の擬似コードを示す。候補の絞り込みは前方境界、後方境界、品詞の順で行う。また、語幹については短い候補（前方境界は右、後方境界は左）から順に調べる。

用例eの各前方境界候補に対して、まず記憶から前方境界fを共有する用例群Eを取り出す(retrieveExamples)。次に、用例群の比較により、若干の後方境界候補の足切りを行う(refineRearBoundaryCandidates)。これにより、語幹の長さが0の候補や、後述の終了条件を満

```plaintext
for f in eの前方境界候補のリスト F, 右側の候補から順に do
E ← retrieveExamples(e, f)
R ← refineRearBoundaryCandidates(e, E, f)
for r in 後方境界候補のリスト R, 左側の候補から順に do
P ← refinePOSCandidates(e, E, f, r)
if P = {p} かつ (f, r, p) が終了条件を満たす then
    return (f, r, p)
end if
end for
end for
```

図3 選択の擬似コード
たさないことが明らかな候補を取り除く．残った各後方境界候補 \( r \) に対して，品詞の絞り込みを行う (refinePOSCandidates)．品詞候補が \( p \) 一つに絞り込まれ，その候補が獲得の終了条件を満たすなら，候補 \((f, r, p)\) を獲得する．

選択の方針は，単純に，多くの用例をうまく説明できる候補を選ぶというものであり，絞り込みは用例群の包含関係により行う．refinePOSCandidates では，\((f, r)\) を共有する用例群中の被覆が閾値以上の品詞候補を選ぶ．ただし，「普通名詞」，「サ変名詞」，「ナ形容詞」は区別が明確でなく，また，「非音動詞」の「基本連用形」と「普通名詞」の区別は困難なため，これらの品詞のみが候補として残った場合には「普通名詞」を採用する．

終了条件は，候補 \((f, r, p)\) を共有する用例群について，次の二つが満たされる場合とする．一つ目は前方境界の妥当性のチェックである．具体的には，句読点などの明らかな境界マークから前方境界が得られた候補の割合が閾値以上とする．例えば，未知語「新撰組」に対して，形態素解析が「新」を接頭辞と解釈するため，常に「撰組」が辞書項目の候補となる．選択アルゴリズムは短い候補を選択するので，「新撰組」よりも先に「撰組」が調べられる．しかし，「撰組」の直前に句読点等が来る用例はないので，「撰組」は獲得されない．二つ目は活用型の異なり数が閾値以上という条件である．これにより，品詞が偶発的に選択されたのではなく，実際に該当品詞として使われていることを確認する．

### 4.6 品詞分類手法の比較

品詞分類について先行研究との簡単な比較を示す．Mori et al. の後ろの「文字列」と福島・鍛治らの「後続するひらがな n-gram」，および桑江らの「最長後続ひらがな列」は，本論文のサフィックスと同様の働きをする (Mori and Nagao 1996；福島，鍛治，喜連川 2007；鍛治，福島，喜連川 2009；桑江，佐藤，藤田 2008)．

Mori et al. は前後の文字列とその頻度をベクトルで表現し，語幹候補と品詞モデルとのベクトル間の距離の近さにより品詞を判定している．しかし，同じ品詞に属する形態素が本当に似たベクトルを取るのだろうか．直観的には，品詞は大雑把な分類であり，同じ品詞に属する形態素でも振る舞いにはばらつきがありそうに思われる．そこで，ウェブコーパスを対象に簡単な実験を行った．まず，コーパスの形態素解析結果から既知語に後続するサフィックスを収集する．次に，サフィックスを各形態素ごとに集約し，形態素ごとの後続サフィックスの頻度分布を求め．同様にして，形態素が属する品詞ごとに，後続サフィックスの頻度分布を求め．そして，各形態素と品詞との間で，後続サフィックスの頻度分布の近さを求める．ただし，頻度分布の近さの尺度として Skew divergence \( s_{\alpha} \) を用いる (Lee 2001) 事例では異なり数の閾値を 3 とした．従って，獲得には最低 3 個の用例が必要となる．
ここで，q，r はサフィックスの頻度分布とし，α = 0.99 とする。

図4に、「子音動詞ヲ行」の例を示す。横軸は「子音動詞ヲ行」の各形態素の絶対頻度を表し、縦軸は各形態素の、「母音動詞」との近さと「子音動詞ヲ行」との近さと「差」を表す。低頻度区間では、二つの近さの差が小さく、近さによる品詞判定では識別が難しいと予想される形態素が目立つ。それだけでなく、高頻度区間でも差が小さい形態素が散見される。従って、出現頻度が大きくても、近さによる品詞判定が難しいと予想される場合が存在する。

福島・鍼治らは品詞識別に SVM を用い、素性として後続するひらがな n-gram を与える。素性的値に福島らは頻度、鍼治らは出現したか否かの 2 値を使う。SVM は識別器であり、品詞内の近さよりも品詞間の差異を学習すると期待される。

一方、提案手法は、サフィックスの頻度には注目せず、個々のサフィックスを品詞リストに写像する。サフィックスは形態論的制約を満たすか否かの 2 値を表現しており、候補列挙の時点で、制約を満たさない品詞は候補から除外される。このように、品詞の繋り込みが各用例に対して行われるので、単純に多くの用例を説明できる候補を選ぶだけで品詞分類が行える。

また、提案手法は一つの語幹に対応する品詞は一つという仮定を置いている。これに対し、Mori et al. と桑江らは、「楽しい」と「楽しむ」のように、一つの語幹が複数の品詞に属す可

\[ s_{\alpha}(q, r) = D_{KL}(r||\alpha q + (1 - \alpha)r), \]
\[ D_{KL}(q||r) = \sum_y q(y)(\log q(y) - \log r(y)). \]
能性を明示的にモデル化している。しかし、「楽しい」と「楽しむ」のような派生関係にある形態素の品詞の衝突は、基本語彙が登録済みのため、極めてまれと推測される。無関係な形態素同士の偶発的衝突については、提案手法はテキストを逐次的に解析するため、同一ドメインのテキストを読んでいる場合、特に起きにくいと推測される。

4.7 獲得未知語の分割可能性

獲得された未知語が実際には2個以上の形態素からなる可能性がある。未知語は比較的少数の用例から獲得するため、未知語Bが、観測された用例中でたまたまABという連続で現れていた場合、ABを1形態素として獲得してしまう。例えば、複合語「顆粒タイプ」が未知語「顆粒」よりも先に獲得されるかもしれない。

この問題に対処するために、未知語獲得時に、獲得済みの形態素が獲得形態素によって分割できるかを調べ、できる場合にはその形態素を辞書から削除する。現在のところ、分割可能性の検査には形態素解析器を用いる。これにより形態素解析器に記述された制約知識を利用する。まず、分割対象形態素の候補列挙は単純な文字列マッチングにより行う。次に、候補を一時的に辞書から取り除いた状態で、その候補の形態素解析を行い、獲得形態素によって分割されなかった場合に候補を辞書に戻す。

5 実験

5.1 実験設定

オンライン未知語獲得について、獲得される未知語の精度、および未知語獲得の形態素解析への貢献を評価する。基本語彙辞書として、形態素解析器JUMANのデフォルトの辞書を用いる。この辞書は約3万の基本語彙を収録している。表記ゆれを展開し、固有名詞を含めば、語彙数は約12万となる。

獲得対象テキストとして、ドメインが限定されたコーパスを用いる。話題を共有するテキストの方が、互いに無関係なテキストよりも未知語が集中的に出現すると期待されるからである。実験では、検索エンジン基盤TSUBAKI(Shinzato, Shibata, Kawahara, Hashimoto, and Kurohashi 2008)を用い、その検索結果をドメイン限定コーパスとみなす。各クエリに対して、システムは検索結果のページを順に読み、未知語を獲得する。獲得は千ページ目で打ち切り、同じ千ページを拡張された語彙を用いて再解析する。クエリとしては、「捕鯨問題」、「赤ちゃんボスト」、「ジャスラック」、「ツンデレ」、および「アガリクス」を使用する。

獲得された未知語は、語幹と品詞の両方が正しい場合に正解とする。ただし、2章で述べたように、語幹の単位認定は難しい。実際、Nagataと内元らは、単位認定の不一致が報告されたエラーの原因の一つとみなしている(Nagata 1999; 内元 他 2001)。単位認定の不一致を回避する
ために、正解コーパスとの単純比較ではなく、入手法による判定を採用する。未知語獲得の形態素解析への貢献の評価は次の手順で行う。獲得対象テキストを基本語彙辞書と拡張された辞書の 2 通りで形態素解析する。二つの解析結果を比較して、図 5 のように単語分割の境界が一致しない箇所を抽出する。これを “diff” ブロックとする、“diff” ブロックの正誤判定は、形態素への分割と、分割および品詞割り当ての 2 通りにより行う。ただし、形態素境界は、明らかに誤っていない場合を正誤とする。評価には、クエリごとに、再解析により解析結果が変化した文の中から無作為に抽出した 50 文を用いる。品詞の評価については、「普通名詞」と「サ変名詞」という名詞の「品詞細分類」を区別しない。また、JUMAN が未知語に与える特殊な品詞「未定義語」は名詞とみなす。

5.2 実験結果

表 3 にクエリごとの統計を示す。再解析により変化した文の割合に大きなばらつきがある (0.43–9.26%)。基本語彙辞書はこれまで新聞記事を対象に整備されてきたため、新聞記事と似ていないドメインほど未知語獲得の効果が大きい傾向がみられる。

| クエリ     | 再解析で変化した文数 | 獲得された未知語数 | 必要用例数 |
|------------|------------------|------------------|-----------|
|            | (割合)           | (精度)           |           |
| 捕鯨問題   | 2,645 / 144,509  | 187 / 190        | 5         |
|            | (1.83%)          | (98.4%)          |           |
| 赤ちゃんポスト | 523 / 120,521   | 74 / 75          | 4         |
|            | (0.43%)          | (98.5%)          |           |
| ジャスパー   | 5,768 / 271,898   | 460 / 470        | 5         |
|            | (2.12%)          | (97.9%)          |           |
| タンデレ   | 12,508 / 135,137 | 215 / 221        | 5         |
| (tundere)  | (9.26%)          | (97.3%)          |           |
| アガリクス   | 4,778 / 84,498   | 76 / 78          | 7         |
| (agaricus) | (5.65%)          | (97.4%)          |           |

1 獲得時点で利用した用例数の中央値。

また、「名詞形態指示詞」も名詞とみなす。名詞形態指示詞の形態レベルの振る舞いは名詞と同様だからである。指示詞はクローズドクラスだが、「コレ」のようなカタカナ指示詞は新聞記事には出現しないため、登録が漏れている。
獲得された未知語の精度は97.3–98.5%と高い。しかも、獲得時点で利用した用例数の中央値は4–7に過ぎない。先行研究では出現回数が10回未満の候補を信用できないとして無視していたことを考えると非常に小さな値である（Mori and Nagao 1996）。

図6に、獲得された未知語の頻度とその頻度の順位との関係を示す。ここで、頻度は、拡張された辞書を用いた再解析結果から数えたものである。順位の下位区間における急な落ち込みは、用例数の不足により獲得されていない未知語の影響と推測される。図7に、獲得の経過を示す。ここで、獲得未知語の累積出現数は、拡張された辞書を用いた再解析結果から数えたものである。終了時点での蓄積されている用例数と未知語の累積出現数の比較から、検出された未知語用例がすべて真の未知語と仮定すると、提案手法で検出される未知語のうちおよそ半分が獲得されたと推定できる。

表4に獲得された未知語の例を示す。予想される通り、獲得された未知語の大半が名詞(94.1–100%)やカタカナのみからなる形態素(67.9–79.4%)である。「タイホ」や「ぱくる」など新聞記事にはあまり見られない俗語も獲得されている。字種が混在する「ドジっ娘」や「シャ乱Q」は字種に基づく形態素解析の未知語処理では正しく解析できない。「すごい」に対する「すごい」、「解る」に対する「解る」のように、登録済みの形態素の異表記もあった。

誤り例には、「バックする」や「フラグが立つ」など明らかに構成的な表現を1形態素と認識しているものがある。ただし、これらは、さらに未知語獲得を進めて、それぞれ「バック」や「フラグ」が獲得された場合、分割可能性のチェックにより消される。他には、副詞の「やっぱ」が名詞と誤認識された。
図 7 クエリ「ジャスラック」における獲得の経過

表 4 獲得未知語の例

| クエリ       | 獲得未知語 ¹ |
|--------------|-------------|
| 捕鯨問題     | モラトリアム，グローバリズム，ツチノ，解る：音動詞行。ナイープ：ナ形容詞 |
| 赤ちゃんポスト | 円ブリオ基金，ネグレクト，助産師，婚外子，トピ主 |
| ジャスラック   | タイホ，シャ乱Q，ばくる：音動詞行，スゴイ：イ形容詞，ムカつく：子音動詞行 |
| ツンデレ     | ドジっ娘，おっさん，イケる：母音動詞，モテる，しまんぬ，ヤバイ：イ形容詞 |
| アガリクス   | コスメ，サブリ，ガラナ，αリボ酸，食効，すぐりむん |

¹ 品詞を省略した形態素は普通名詞。

表 5 に "diff" ブロックの評価結果を示す。ほとんどのブロックが拡張された語彙によって正しく解析されている（E → C および C → C）。一方、獲得による副作用は限定されている (C → E)。従って、獲得された未知語が形態素解析の精度を改善することが示された。

5.3 議論

形態素解析において、カタカナ未知語が短いカタカナ形態素によって分割されることがある。例えば、基本語彙辞書のみを用いると、未知語「アブラハム」は「アブラ」と「ハム」に過分割される。「アブラハム」は、3.2 節で述べた単純な検出手法では検出されず、従って獲得もされ
表5 “diff”ブロックの評価

| クエリ         | 分割 | 品詞 | 計  |
|----------------|------|------|-----|
|                | E → C | C → E | E → E | C → E | E → C | C → E | E → E | C → E | 計  |
| 捕鯨問題       | 19   | 36   | 0   | 3   | 19   | 36   | 0   | 3   | 58  |
| 赤ちゃんポスト  | 33   | 19   | 0   | 1   | 33   | 19   | 0   | 1   | 53  |
| ジャスラック     | 18   | 32   | 2   | 0   | 12   | 32   | 8   | 0   | 52  |
| ツンデレ        | 4    | 53   | 0   | 1   | 4    | 53   | 0   | 1   | 58  |
| アガリクス      | 30   | 29   | 0   | 0   | 30   | 29   | 0   | 0   | 59  |

(凡例　C: 正解　E: 誤り)

ない、また、未知語の獲得によって新たな過分割が発生し得る。例えば、「サー」の獲得によって、「サーバー」が「サー」と既知の「バー」によって過分割されるようになる。このような過分割の問題は、本論文が利用した形態レベルの文法的指標する舞いだけを調べても解決できない。他の方法が考えられる。例えば、「サーバー」がserverという一つの外来語だから分割できないといった知識が必要となる。

提案手法では、カタカナ「イイ」のように語尾までカタカナで表記された用言は誤って名詞と認識する。現在の形態素解析は語尾のひらがな表記を前提としている。この仮定は新間記事に対しては妥当だが、ウェブテキストに対しては無効であり、より柔軟な解析が必要になる。ただし、こうした未知語の解析は元々誤っており、獲得によって形態素解析が悪化するわけではない。

未知語問題への2通りの解決策のうち、未知語モデルによる手法は、その利点として、低頻度語の正しい同定が強調されている(Nagata 1999; Asahara and Matsumoto 2004)。しかし、ウェブの出現により、ほとんど無尽蔵のテキストが入手できるようになった現在、限られた情報のみを用いた同定は不可欠ではない。仮に、解析対象のテキストが少量で、未知語獲得を行うには用例の出現回数が足りないとしても、ウェブから解析対象テキストと関連するテキストを収集することで、用例の出現回数を増やすことができる。ここで、バッチ処理(Mori and Nagao 1996)と異なる、オンライン獲得という特徴を生かせる。すなわち、解析対象テキストから検出された用例にマークして、それらの用例が獲得に使われたか追跡することで、未知語が十分に獲得された時点で処理を停止させることができる。

最後に、残された課題を整理する。2章で整理したように、形態素に付与される様々な情報のうち、主論文はひとまず形態レベルの情報の獲得を指向した。形態レベルの手がかりでは得られない知識としては、名詞と副詞の区別の他に、固有名詞と普通名詞の区別などがある。特に名詞の細分類は固有名詞表現や省略・照応解析に役立つと期待されるので、テキストからの自動獲得を目指したい。

本論文では形態素の単位認定にこだわらなかったが、獲得された未知語の中には構成的なもの
が含まれている。参考までに，クエリ「ジャスラック」の獲得結果を調べたところ，判断に迷う
場合を含めると 10%弱 (45/460) が複合語であった。ただし，複合語の基準としては，JUMAN
4.0 から 5.0 への変更時に行った複合語の整理7を参考にした。日本語の複合名詞は，文法的な
マーカなしに構成要素が直接連結されるため，形態レベルの手がかりでは構成要素に分割でき
ない。細粒度での単位認定を実現するには，他の手がかりを利用する必要がある。

形態レベルでの未知語獲得については，検出が大きな課題として残っている。なかでもひら
がな表記の未知語は暗味性が高く，形態素解析器によってより短い既知の形態素へ分割され
ることが少なくない。予備調査として，形態素解析結果のうち，1 + 1，1 + 2，2 + 1 文字という
パターンのひらがな形態素のペアのみを対象に，未知語検出の再現率を求めたところ，本論文の
手法では 31%にとどまることが判明している。ひらがな表記の未知語は数の上では少なく，頻
度の上でも異なり数でも，未知語の大半をカタカナ名詞が占める。しかし，カタカナ名詞は形
態素解析の未知語処理では問題なく同定できるのに対し，ひらがな未知語の解析誤りは応用
に大きい悪影響を及ぼしやすい。例えば，「ようつべ」が「よつべ」，「うつべ」，「べ」に誤って分解さ
れ，「うつ」が動詞と解釈された場合，文節まとめが取れず，「よつべ」，「うつべ」，「べ」
の間に文節境界が引かれる，これに基づき見当違いな係り受け解析が行われてしまう。提案手法
の利点の一つは，既知の形態素を改めて獲得しないことによる効率の良さだが，今後はこの利
点を維持しつつ検出の再現率を上げていきたい。

6 関連研究

形態素解析における未知語の問題は，言語の類型論的特徴や文字の性質に依存する部分が少
なくない。フィン語やトルコ語は日本語と同様の豊富な言語で，自立語に複数の付属語が後続して
語を形成する。ただし，これらの言語は分から書かずするため，形態素解析は，分から書きの単
位である語を形態素に分割するタスクとなる。例えば，Morpho Challenge では，頻度つきの語
のリストから教師なしで形態素を切り出すタスクが競われている (Kurimo, Creutz, Varjokallio,
Arisoy, and Saraclar 2006; Kurimo, Creutz, and Turunen 2007)。

日本語と同様に分から書かずない言語としては，中国語やタイ語などがあるが，いずれも分
析的であり，豊富な日本語とは性質が異なる。Peng et al. は中国語を単語分割に新語検出を
組み込む (Peng, Feng, and McCallum 2004)。この手法では，テキストを一度解析した結果から
単語分割の信頼度を元に新語を検出し，それらを素性に組み込んだ状態で再解析を行う。

日本語については，未知語モデルを導入する手法がいくつか提案されている。Nagata は，字

7 「日本語形態素解析システム JUMAN version 6.0」付録 E12.1 参照。http://nlp.kuee.kyoto-u.ac.jp/nl-
resource/juman.html
種や単語長に基づく生成的な未知語モデルを単語分割に組み込む（Nagata 1999）。内元らは最大エンタロピーモデルに基づく形態素解析の中で、字種やその変移などの未知語同定に有効な情報を素性として利用する（内元他 2001）。東山らは最大エンタロピーモデルに代えて条件付き確率場を採用する（東他 2006）。Asahara et al. は文字レベルのランキングにより未知語を同定しており、学習器として Support Vector Machine を用いる（Asahara and Matsumoto 2004）。中川らは、単語分割されたテキストに対して、未知語のすべての出現を考慮して品詞を推定する手法を提案している（中川、松本 2008）。しかし、我々が想定するタスクでは、品詞推定と独立に単語分割が実現されるという仮定は現実的ではない。

テキストからの未知語の自動獲得については、Mori et al. は、語幹の前後の文字列とその頻度をベクトルで表現し、コーパス中の任意の部分文字列について、品詞のモデルとのベクトルの距離により品詞確率を判定する（Mori and Nagao 1996）。鍛治らはカタカナ用言についてテキストからの自動獲得を行っている（鍛治他 2009）。彼らは、獲得対象を「ググる」などの語幹が自明なカタカナの用言に限定し、品詞分類に特化した手法を提案する。一般的の未知語を獲得する場合には、あわせて語幹同定の問題も解く必要がある。

7 結論

本論文では、オンライン未知語獲得という枠組みと、その具体的な実現方法を提案した。実験により、未知語が高精度に獲得され、その結果形態素解析の精度が向上することが示された。形態素解析自体は成熟した技術であり、構文解析や情報検索といった応用のための前処理となっている。従って、応用処理の精度向上に提案手法を利用したいと考えている。

参考文献

浅原正幸、松本裕治（2002）。形態素解析のための拡張統計モデル。情報処理学会論文誌、43 (3)、pp. 685–695。

Asahara, M. and Matsumoto, Y. (2004). “Japanese Unknown Word Identification by Character-based Chunking.” In Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004), pp. 459–465。

東山、浅原正幸、松本裕治（2006）。条件付確率場による日本語未知語処理。情報処理学会研究報告。自然言語処理研究会報告、2006 (53)，pp. 67–74。

Banko, M. and Brill, E. (2001). “Mitigating the Paucity-of-data Problem: Exploring the Effect of Training Corpus Size on Classifier Performance for Natural Language Processing.” In Proceedings of the First International Conference on Human Language Technology Research.
Brants, T., Popat, A. C., Xu, P., Och, F. J., and Dean, J. (2007). “Large Language Models in Machine Translation.” In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007), pp. 858–867.

Kawahara, D. and Kurohashi, S. (2006). “Case Frame Compilation from the Web using High-Performance Computing.” In Proceedings of The 5th International Conference on Language Resources and Evaluation (LREC-06), pp. 1344–1347.

Kudo, T., Yamamoto, K., and Matsumoto, Y. (2004). “Applying Conditional Random Fields to Japanese Morphological Analysis.” In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), pp. 230–237.

Kurimo, M., Creutz, M., and Turunen, V. (2007). “Overview of Morpho Challenge in CLEF 2007.” In Working Notes of the CLEF 2007 Workshop, pp. 19–21.

Kurimo, M., Creutz, M., Varjokallio, M., Arisoy, E., and Saraçlar, M. (2006). “Unsupervised Segmentation of Words into Morphemes–Challenge 2005, an Introduction and Evaluation Report.” In Proceedings of the PASCAL Challenge Workshop on Unsupervised Segmentation of Words into Morphemes.

Kurohashi, S., Nakamura, T., Matsumoto, Y., and Nagao, M. (1994). “Improvements of Japanese Morphological Analyzer JUMAN.” In Proceedings of the International Workshop on Sharable Natural Language Resources, pp. 22–38.

Lee, L. (2001). “On the Effectiveness of the Skew Divergence for Statistical Language Analysis.” In Proceedings of the Eighth International Workshop of Artificial Intelligence and Statistics (AI & Statistics 2001), pp. 65–72.

Mori, S. and Nagao, M. (1996). “Word Extraction from Corpora and Its Part-of-Speech Estimation Using Distributional Analysis.” In Proceedings of the 16th Conference on Computational Linguistics, Vol. 2, pp. 1119–1122.

Nagata, M. (1999). “A Part of Speech Estimation Method for Japanese Unknown Words using a Statistical Model of Morphology and Context.” In Proceedings of the 37th Annual Meeting
of the Association for Computational Linguistics (ACL 1999), pp. 277–284.
中川哲治，松本裕治 (2008). 大域的な情報を用いた未知語の品詞推定. 情報処理学会論文誌, 49 (3), pp. 1437–1450.
小松秀樹，小磯花絵，富士池優美，原裕 (2008). 『現代日本語書き言葉均等コーパス』形態論情報規程集.
Peng, F., Feng, F., and McCallum, A. (2004). “Chinese Segmentation and New Word Detection using Conditional Random Fields.” In Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004), pp. 562–568.
Shinzato, K., Shibata, T., Kawahara, D., Hashimoto, C., and Kurohashi, S. (2008). “TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology.” In Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP 2008), pp. 189–196.
内元清貴，関根聡，佐佐原均 (2001). 最大エントロビーモデルに基づく形態素解析 未知語の問題の解決策. 自然言語処理, 8 (1), pp. 127–141.

略歴

村脇 有吾：2006 年 3 月京都大学工学部情報学科卒業．2008 年 3 月京都大学大学院情報学研究科修士課程卒業．同年 4 月，同博士後期課程入学，現在に至る．

黒橋 禎夫：1994 年京都大学大学院工学研究科電気工学第二専攻博士課程修了．
博士（工学）2006 年，京都大学大学院情報学研究科教授，現在に至る。

（2009 年 4 月 22 日 受付）
（2009 年 7 月 2 日 再受付）
（2009 年 8 月 25 日 再々受付）
（2009 年 9 月 22 日 採録）

75