社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

組込みシステムの要求仕様書に対する修正候補の定量的調査

山本　椋太† 　吉田　則裕† 　高田　広章†

† 名古屋大学大学院情報学研究科　〒 464–8603 愛知県名古屋市千種区不老町
E-mail: †{muku,yoshida,hiro}@ertl.jp

あらまし　ソフトウェア開発においては様々な文書が作成されるが，文書作成者によっては，文書中に曖昧さや誤りが生じることがあり，その文書を使用するプロセスにおける障害となることがある．そこで本研究では，組込みシステムの要求仕様書から修正候補を目視で抽出する．各修正項目を分類し，分類ごとの出現回数についてまとめた．自動抽出が容易な分類の内，一部の誤りや曖昧さについて，形態素解析ツールによる自動抽出を試みた．その結果，一部の誤りや曖昧さについて目視の場合とは異なる抽出結果となったが，同一格助詞の不正な連続出現パターンについては，目視の結果と 97%程度一致した．

キーワード　自然言語処理，要求仕様書，組込みシステム

An Empirical Study of Correction Candidates in a Requirements Specification Document for an Embedded System

Ryota YAMAMOTO†, Norihiro YOSHIDA†, and Hiroaki TAKADA†

† Graduate School of Infomatics, Nagoya University　Furo-cho, Chikusa-ku, Nagoya, Aichi, 464–8603 Japan
E-mail: †{muku,yoshida,hiro}@ertl.jp

Abstract　Software developers tend to write incorrect/ambiguous expressions in a document. Incorrect/ambiguous expressions often cause rework. In this study, we try to extract manually from a requirements specification document for an embedded system. We identify candidates of incorrect/ambiguous expressions and then categorize them. After that, we investigate the number of the candidates for each category. Finally, we implement a tool which can extract a part of candidates for each category. As a result, the experimental result shows that 97% of the detection results are matched between the automatic and manual detection approaches in the case of consecutiveness particle pattern.

Key words　Natural Language Processing, Requirements Specisification Document, Embedded System

1. はじめに

ソフトウェアの開発過程においては，さまざまな文書が作成される．たとえば，要求仕様書，システム設計書およびテスト設計書などがあげられる．しかし，人間がこれらの文書を自然言語によって作成すると誤りや曖昧さが混入する可能性があり，その状態のまま合意形成される場合がある [1], [2].

本研究では，企業において実際に使用された組込みソフトウェアの要求仕様書から修正候補を目視で抽出する．はじめに，作成元によってレビューによる修正がすでに加えられた要求仕様書を使用して対象文書から，修正候補分類を作成し分類ごとの出現回数についてまとめた．分類の例には，大別して日本語文法上の誤りや文章の誤りや曖昧さがあり，これらをさらに小分類に分けている．この文書は，およそ 150 項目からなる．上記の分類を適用して，予備実験として目視による出現回数の調査を実施した．その後，修正候補分類のうち，比較的容易に実装できるものについて実装を行い，予備実験の結果と比較した．予備実験および実装ツールにおいて，本研究では文章のみを対象としているため，図表などは参照していない．

本稿では，この分類と自動抽出の結果について考察し，技術文書に対して自動的な修正候補の抽出可能性について検討する．

2. 予備実験

2.1 分析対象と修正候補分類の方針

本章においては，組込みシステムのソフトウェアに関する要求仕様書を読み，修正候補分類を検討する．

そのために，企業において実際に使用された組込みシステムのソフトウェアについての要求仕様書（以降，D_0 と呼ぶ）を

目視により確認する．この文書は，約 150 項目からなり，大項目・中項目・小項目から構成され，大項目を中項目，小項目と徐々に詳細化する形式によって記述されている．D_0 は，作成元によってレビューによる修正がすでに加えられた比較的品質の高い要求仕様書である．

本稿においては，項目同士の意味における矛盾について確認を行っていない．すなわち，各要求仕様の項目内に閉じた上で記述の誤りや曖昧さを調査した．

2.2 修正候補分類

2.1 の文書 D_0 を目視で読み，分類を検討した．以降，分類について説明する．

(分類 1) 誤字脱字

明らかな誤字や脱字を対象とした．例 1-1 では，本来句点にするべき箇所が読点になっている．例 1-2 では，「オフにする」とするべき箇所が「オフする」となっている．

例 1-1) スイッチの状態がオフならば、LED をオフにする。

例 1-2) スイッチの状態がオフならば、LED をオフする。

(分類 2) 同一格助詞の連続

同一格助詞が不正に連続しているものを対象とした．例 2-1 では，格助詞「が」が連続しており例 2-2 では，格助詞「を」が連続しているため，これらは文法的に誤りとなっている．その結果，例 2-1 では主語の，例 2-2 では対象の係り受けが曖昧になっている．

例 2-1) スイッチが状態がオフのとき...

例 2-2) LEDをスイッチをオフにしたときにオフにする。

(分類 3) 主述の誤り

主語に対して，明らかに述語が誤っている文を対象とした．例 3-1 では，構文上間違いはないが，「LED をオフにする」とするのが正しい．

例 3-1) スイッチの状態がオフならば、 LED がオフにする。

(分類 4) そもそも曖昧さを含むような言葉

読み手の主観によって解釈が変わりうる語を含む文を対象とした．例 4-1 では，「少し」という語が含まれており，読み手によって程度が変わる．

例 4-1) スイッチの状態がオフならば、LED を少し暗くする。

(分類 5) 未定義の言葉

ある文で出現した語について，その語についての定義がない場合，その単語を未定義語だと判断する

(分類 6) 表現のゆらぎ

同じ意味だが，異なる表記をしている語を対象とした．例 6-1 では，「オフ」と「OFF」は同じ意味合いの語だが，異なる表記である．

例 6-1) スイッチの状態がオフならば、LED をOFFにする。

(分類 7) 主語が未定義

主語が欠落することで，どのオブジェクトやイベントによってあるイベントや他のオブジェクト等に対する処理が行われるのかが不明確な文を対象とした．日本語は，主語を省略することがしばしばあり，主語を省略しても読み手に意図を伝えることができる場合がある．予備実験においては，意味が理解できれば主語が省略されていても良いこととした．具体的には，D_0 が表しているシステムまたは直前の内容を受ける場合である．例 7-1 では，LED をオフに変化させるときの主語が曖昧である．

例 7-1) スイッチの状態がオフならば、 LED をオフにする。

(分類 8) 対象が未定義

目的語が欠落することで，何に対して処理を行うのかが不明確な文を対象とした．例 8-1 では，何をオフにするかが曖昧である．

例 8-1) スイッチの状態がオフならば、オフにする。

(分類 9) 条件が未定義

ある処理や遷移が発生するための条件が不明確な文を対象とした．例 9-1 では，LED がオフになるための条件が書いておらず，どのタイミングで LED をオフにするのかが曖昧である．

例 9-1) LED をオフにする。

(分類 10) 複数解釈可能な文章

1 つの文から複数の解釈が得られる場合の文を対象とした．例 10-1 では，LED が，省略されている対象が光っていると判断するのか，それとも LED が光っていることを，省略されている主語が判断するのかというように 2 通りの解釈をすることができる．

例 10-1) LED が光っていると判断する。

2.3 修正候補分類ごとの出現回数

D_0 のすべての項目から，修正候補分類ごとの出現回数を，目視によって数えた．その結果を図 1 に示す．

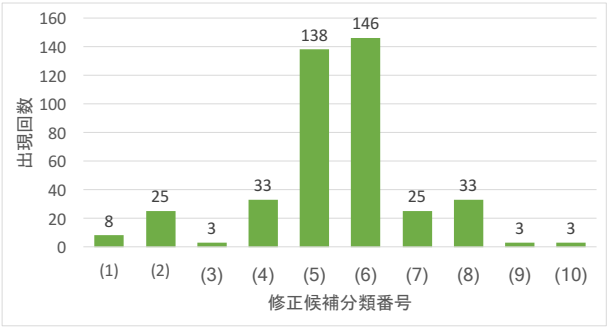

図 1 対象とした要求仕様書における修正候補の出現回数

この結果において，(分類 5) と (分類 6) の出現数が比較的多いものとなった．(分類 5) については，今回は要求仕様書の文書のみを対象としたため，図表や別紙資料などを参照しておらず，未定義語句が多かったと考えられる．未定義語としての基準は，型が示されていない語句のうち，プリミティブではないものである．(分類 6) については，特に，「オン」と「ON」，「オフ」と「OFF」の揺らぎが多く，これらが 8 割程度を占めている．また，「実行する」，「処理する」といったほぼ同一の動作を表すような語句も散見された．

次いで，(分類 2)，(分類 4)，(分類 7) および (分類 8) の出現数が多い．(分類 2) について，同一格助詞の連続は，格助詞「が」や「に」の連続が多く存在した．これらは文章のコピーアンドペーストによって増えたと考えられる．なぜなら，ほぼ同一だが条件や結果だけが異なるような文章において，同一のミスが必ず存在している．つまり，類似の文の中で最初に出現する文章の時点ですでに誤りが存在しているため，修正候補が拡

散したと考えられる．

(分類 4) について，アクタにユーザが含まれているような仕様書であるため，「ユーザによって心地よい」といったユーザの感覚に依存するような表現が存在した．また，長・中・短といった，抽象的な長さの表現が見られた．

(分類 7) について，多くの項目で主語の省略が見られたが，主語に開発対象のシステム名を補完すると解釈できるような項目は修正候補ではないこととした．たとえば「ある機能を止めたとき」というような表現において，その「止める」ことを実行する主体が，明らかに開発対象のシステムではなくとも，主語が省略されている場合があった．

(分類 8) については，「A を B に要求する」という文が多く存在し，要求を出すタイミングが明記されていない．また，「C から D を取得する」という文があり，同様にタイミングが書いていない．例えば開発対象が何らかの制御システムであり，かつ C が何らかのセンサである場合，C のサンプリングレートは制御に関わる．取得は周期ハンドラによって実行されるのか，それとも非同期に実行されるのか，この文章からは決定できない．

その他の修正候補分類については，出現回数が少ないため，2.2 において示した例の表現が見られた．

2.4 修正候補分類の重要度

2.3 において，各修正候補分類の出現回数についてまとめた．しかし，出現回数だけではなく，各修正候補分類の性質によって重要度が異なると考えている．そのため，これらの修正候補が存在することによる開発者 (読み手) に対する影響を検討する必要がある．加えて，修正候補分類ごとに，自動抽出するための難度が異なると考えられるため，このことについても検討を行う．修正候補分類に対する解釈の困難さについて，検討した結果のまとめを表 1 にまとめる．

2.4.1 開発者の視点に対する修正候補分類の影響

開発者 (読み手) の視点から検討する．(分類 1) については，重要度がさらに詳細に分類される．例 1-1 の「オンする」と「オンにする」のように，ユーザが暗黙の了解として解釈可能な誤りもあれば，「スイッチの状態がオならば」のように，遷移先状態が不明確であるため解釈不可能な誤りも存在し得る．

(分類 2) については可読性は減少するが，おおよその場合正しく解釈可能であると考えられる．しかし，主語や目的語が曖昧になることが考えられ，この誤りの数が多い場合に解釈不能になる可能性がある．

(分類 3) については，例 3-1 のような場合には，語句によっては，誤りであることに気づくのは容易であるが，どのように修正するべきか読み手では断定できないことが多く，解釈不能になる場合が多いと考えられる．さらに，語句によっては，実際に係り受けが誤りであったとしても，読み手によって気づかれない可能性がある．たとえば，主語が受動的なオブジェクトであれば，動詞が受動態になっていない限り，読み手が誤りであると理解することができる．しかし，能動的とも受動的ともとれるようなオブジェクトでは，そもそも誤っていることに読み手は気が付かない．仮に，書き手と読み手が異なる場合に，読み手が誤りの存在に気がついて修正したとしても，書き手の意図と一致しない可能性がある．

(分類 4) については，比較的大きな問題になる．「少し」などの表現は，実装時に個人の感覚によって決められるものであり，書き手の意図や顧客の要求に一致しない可能性がある．

(分類 5) については，ある未定義語の型がわからないこととなる．そのため，設計時に型を決定するためにコストがかかる．

(分類 6) については，今回頻出した「オン」と「ON」のように，明らかに同一の意味であるとわかるような語であれば，問題になるケースは少ない．

(分類 7) については，暗黙の了解にとして主語が省略されているのか，それとも書き手が主語を書き忘れたのか，その判断を読み手が行うことが問題となる．(分類 3) と類似しており，そもそも読み手が気がつかない場合も気がつく場合もあり，気がついた場合に修正したとしても書き手と読み手の意図が一致しない可能性がある．

(分類 8) については，開発者にとって大きな問題になる．目的語は，文法として省略されることが基本的には許されていない．つまり，たとえば状態遷移の記述を考えたときに，あるオブジェクトがある状態に遷移することを表現する際，主体となるオブジェクトが明記されていないことと同義である．このとき，遷移先の状態がそのオブジェクト特有のものであれば，推測できるが，そうではない場合，断定することは困難である．

(分類 9) については，開発者にとって比較的大きな問題になる．条件も目的語同様，推測することが困難である．読んでいる時点では，条件があるかどうかを意識して読まなければ，そもそも違和感を感じないような文章が多いと感じた．

(分類 10) については，比較的大きな問題になる．そもそも，複数解釈可能な文であるということに開発者が気づかない可能性がある．もし，誤った解釈を開発者がした場合，開発者は無意識に仕様を読み違える可能性がある．

2.4.2 自動抽出の可能性に対する修正候補分類の影響

自動抽出の可能性の視点から検討する．(分類 1) については，辞書に登録されている用語に関しては可能であると考えるが，過度な誤りや，辞書に登録されていない新出語が出てきた場合には対応が困難になると考えられる．

(分類 2) については，形態素解析によって抽出可能であり，容易に実装できる．

(分類 3) については，意味解析が必要になるため，抽出は困難だと考える．主語が誤っていたとしても文法上は問題にはならないため，主述の関係を意味的に解釈する必要性が生じる．

(分類 4) については，比較的専門的な用語は少ないと考えられる．たとえば，「多い」「大きい」「短い」などは，一般的に使われる言葉であるため，辞書を作成することが容易であると考えている．曖昧な用語についての辞書を作成することで，この曖昧さの検出が可能となる．

(分類 5) については，未定義語の前後で同じ語が出現しており，かつその文が定義を意味する文であるなら定義があることとする．つまり，文が何らかの条件のもと対応する処理を行う文なのか，それともある処理や語を定義する文なのかを判断することができれば，未定義語を抽出できる．定義を表す文は，

表 1　修正候補項目分類に対する解釈の困難さ

修正候補項目分類	開発者 (読み手) が正しく解釈できるかどうか	自動抽出が可能かどうか
(分類 1)	誤字脱字の程度による	誤字脱字の程度による
(分類 2)	解釈できることが多い	形態素レベルで解析可能である
(分類 3)	開発者の主観で判断する可能性がある	意味解析が必要となり，困難である
(分類 4)	開発者が決定できない場合が多い	一般的な曖昧さを含む表現は，抽出可能である
(分類 5)	正しく解釈できるが，そのためにコストがかかる	文章だけを確認するならば，可能である
(分類 6)	解釈できることが多い	機械学習または辞書によって，対応可能である
(分類 7)	開発者の主観で判断する場合がある	主語が省略可能なパターンも抽出するなら，形態素レベルで解析可能である
(分類 8)	開発者が決定できない場合が多い	目的語をとる動詞が定義されていれば解析可能である
(分類 9)	開発者が決定できない場合が多い	条件が必要かどうかを機械的に判断できないため解析は困難である
(分類 10)	開発者が気づかない可能性がある	係り受け解析によって，解析可能である

文法上の誤りがないと仮定した場合,「場合」,「際」などの副詞的名詞や接続詞など，文と文をつなぎ，それらの文に関係性を持たせる語句が存在しない文章であると考えられる．しかし，(分類 9) のような曖昧さが残っている文章に対して適用すると，誤検出となるため，先に (分類 9) に対応する必要がある．

(分類 6) については，Word2Vec などの語のベクトルを求めるような手法によって，類似する用語は表現の揺らぎであると判定できる可能性が高い．しかし，そのためには訓練用データが多く必要となるため，現実的には解決が難しい．または，予想される表現の揺らぎをあらかじめまとめておくことが望ましい．

(分類 7) については，主語が省略可能な文であったとしても主語がなければ修正候補であるとする場合，抽出は容易に可能である．もし，主語を省略可能な文を修正候補としないならば，意味解析が必要となり，困難である．

(分類 8) については，目的語を必要とする動詞が辞書によって与えられているならば解析は可能である．

(分類 9) については，無条件による処理や状態遷移などが存在しないと仮定するならば，抽出することは可能である．仮に，無条件による処理や状態遷移が存在したとしても，該当する処理や状態遷移は無条件である旨が記載されているべきであると考えられるため，この対応は妥当であると考える．また，条件部の抽出方法は文を形態素に分解し，条件を持つ主節とそれに従う従属節を見つけ出すことが必要になる．複数の節を持たない場合に，語句の型定義のであるかを区別できる必要がある．

(分類 10) については，格助詞の組み合わせによって生じる複数解釈可能な文に対しては，構文上の誤りであることから，検出することが可能である．または係り受け解析によって，複数解釈できる可能性が示唆された場合，検出されたこととできる．

3.　修正候補抽出の実装

3.1　抽出の方針

本稿においては，容易に実装可能な構文による修正候補の自動抽出方法について，検討及び実装を行った．辞書を必要とするような項目や意味解析が必要な項目に関しては実装せず，(分類 2)，(分類 7) および (分類 9) についての実装を検討した．

本稿では，句点によって分割されたものを文とし，各文に対して実装した抽出ツールを適用した．また，修正候補抽出のために，日本語形態素解析システム JUMAN(注1) を使用した．

まず，(分類 2) 同一格助詞の不正な連続パターンの抽出については下記の構文を考える．

<格助詞 A> ... <格助詞 A>

形態素解析の結果，上記のような構文がひとつの文の中で現れた場合，格助詞の不正な連続パターンとする．このとき，上記の構文であっても問題がないと判断することがある．その基準は以下のとおりである．

- 始端の格助詞の直後に動詞 (する，して等) が存在する場合
- 同一格助詞の間に読点が存在する場合
- 同一格助詞の間に副詞的名詞が存在する場合
- 二重鉤括弧『』の中に存在する格助動が構文の始端/終端に選ばれた場合

これら 4 つの基準のについて説明する．まず，始端の格助詞の直後に動詞 (する，して等) が存在する場合は，格助詞が連続しても意味を問題なく解釈することができる．いま,「A により B して C になる。」という文を考える．このとき，格助詞「に」が重複しているがその後の動詞「して」によって，文が切れていることがわかる．次に，同一格助詞の間に読点が存在する場合は，読点で節が分かれていることが多いことから，例外とした．そして，同一格助詞の間に副詞的名詞が存在する場合について，副詞的名詞である「時」,「際」および「場合」など，次の節につなげるための語句が間に存在する場合は，例外とした．二重鉤括弧『』の中に存在する格助動が構文の始端/終端に選ばれた場合について，二重鉤括弧『』の中は固有名詞であることが多いため，その中に存在する格助詞は適用対象外とするためである．

次に，(分類 7) 主語が未定義パターンの抽出については，格助詞のうち主語をとる「が」および，副助詞「は」と「も」が文中に存在しない場合，主語が存在しないこととする．このとき，各文を読点によって分割する．(分類 7) については (分類 3) との複合的な問題が存在することが考えられる．たとえば,「LED は ON にする.」という文の場合，正しくは「LED を ON

(注1)：http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

表 2　目視と実装による修正候補の抽出数

	(2)	(7)	(9)
目視	25	25	33
実装	61	282	20

表 3　修正候補項目 (2) における例外該当数

例外	出現数
特定の動詞を含む	82
二重鉤括弧『』の中	13
読点を含む	309
副詞的名詞を含む	171
格助詞の次に副助詞を含む	3

表 4　修正候補項目 (7) における例外該当数

例外	出現数
格助詞「が」を含む	251
副助詞「は」「も」を含む	96(内訳：「は」が 83,「も」が 13)

表 5　修正候補項目 (9) における例外該当数

例外	出現数
接続助詞を含む	1
副詞的名詞を含む	211
格助詞と副助詞の連続を含む	11
句点で終わっていない文	62

にする.」であるため，主語が未定義のパターンとなる．しかし，副助詞「は」を使用しており，文法上の主述関係の誤りは存在しない．

最後に，(分類 9) 条件が未定義パターンの抽出については，文中に副詞的名詞または接続助詞を持たない場合，条件が存在しないこととする．接続副詞とは「なら」などがある．また，句点で終わっていない文章を目視で確認したところ，すべて文になっていないため，除外した．

これらの結果について，目視による抽出数と実装による抽出数を表 2(注2)にまとめる．また，今回実装を行った項目について，入力文を問題ないと，ツールが判断した理由別の数について，表 3(注3)，表 4(注4)および表 5(注5)にまとめる．

3.2　自動抽出結果の考察

自動抽出の結果と目視による抽出の結果の比較について考察する．まず，(分類 2) の抽出数について，(自動抽出の結果)>(目視による抽出の結果) となっている．この理由は，目視による抽出の見落としが考えられる．目視による抽出においては，「から」や「の」の格助詞の連続を考慮しておらず，抽出数が少なくなったと考えられる．そのため，実装時における例外の定義が過剰であったとは考えていない．実装後に，再度見落としがないよう目視によって抽出作業を試みた結果，目視による抽出数が 57 となり，それらはすべて自動抽出の結果に含まれていたため，約 97%一致している．そのため，上記の考察は正しいと考えられる．それでも誤差が生じている理由は，自動抽出に接続詞の規則を含めていないためだと考えており，それを追加することで抽出数は一致した．続いて，例外該当数を見ると，すべての例外が文中に存在することがわかる．同一格助詞間に読点を含む例外が最も多かった．これらの誤りについて，今回はある程度修正された要求仕様書を用いたため，目視で確認した限りにおいては誤用はなかった．しかし，他の文を対象にしたときに誤用があると，例外として処理される数が増加する可能性があるため，例外の条件を追加する必要性があると考えている．

次に，(分類 7) の抽出数について，(自動抽出の結果)>(目視による抽出の結果) となっている．(分類 2) に比べ，自動抽出による抽出数が目視による抽出数を大きく上回っている．この理由は，目視による抽出においては，主語を省略しても良い場合を例外として数えていない．しかし，自動抽出においては，機械的に主語が存在しない文を検出対象としている．目視において主語を省略してもいい場合として，明らかに，開発対象名が主語にくると考えられるものがあげられる．実装後に，主語を省略して良いような場合であったとしても主語が存在しないような文の抽出作業を試みた結果，172 となった．主語の暗黙の了解による省略を許した場合に比べると自動抽出の抽出数に近づいたが，それでもまだ 2 割程度の誤差がある．その理由は，自動抽出において句点で区切っているため，本来文法として主語を省略して良い場合であっても抽出していることである．続いて，例外該当数を見ると，すべての例外が文中に存在することがわかる．今回は，格助詞「が」を使用していることが多いことがわかる．ここで，副助詞「も」について，自動抽出では期待していない結果となった．副助詞「も」は，様々な用途があり，必ずしも主語の直後に存在するわけではないが，主語の直後に存在することがあるため，今回は例外の対象とした．しかし，今回は「A が B の場合も C する」というように，主語の直後に副助詞「も」が存在しない場合だけであった．そのため，今後は副助詞「も」を発見した場合の制約を付与することを検討する必要がある．

そして，(分類 9) の抽出数について，今回の実装で唯一，(自動抽出の結果)<(目視による抽出の結果) となっている．当初，この結果が逆になると予想していた．なぜなら，条件を含まない文には，語句等の定義文が多く含まれると考えているためである．自動抽出の結果をみると，確かに「A は B とする」といった定義文が含まれているが，それは検出数 20 に対して 4 つの文だけであった．この検出数の理由は，例外の定義が不足していたためだと考えられる．特に，複数節存在するような長い文においては，条件が指定されていない動作の存在が目視では確認されている．当初，(分類 7) と同様に読点で文を分割することを考えたが，読点の前後で条件と処理に分割されるケースが考えられるため，今回は，読点の前後で分割しなかった．そのため，今後は日本語形態素解析システム JUMAN と連携す

(注2)：自動抽出において対象にしている文の数について，(分類 2) は 265 文を同一格助詞を始端および終端として分割した 317 箇所を対象としている．(分類 7) は，265 文をさらに読点で分割しているため，527 項目を対象としている．(分類 9) については 265 文である．
(注3)：例外は，項目ごとの重複および文の中で複数存在する場合を含む
(注4)：※例外は，項目ごとの重複および文の中で複数存在する場合を含む
(注5)：例外は，項目ごとの重複および文の中で複数存在する場合を含む

ることができる，日本語構文・格・照応解析システム KNP(注6) によって係り受け解析 [3] [4] や照応解析 [5] [6] を行い，さらに抽出能力を高めることを検討している．続いて，例外該当数を見ると，すべての例外が文中に存在することがわかる．今回，句点で終わっていない文が 3 割程度を占めており非常に多いと考え，確認をしたところ箇条書きの項目には句点を書かないように文章全体として統一されている．箇条書きの箇所は，文になっていないものばかりであり，語句だけの場合もある．一部文になっているものに関しては，定義を示すものだけが存在しているため，例外とすることは妥当であると考える．

4. 関 連 研 究

要求仕様書のような技術文書ではなく，一般的な自然文書の誤り訂正技術について数多くの研究が存在する．本研究で行った格助詞の誤り訂正についても，一般的な自然文書を対象とした研究では多くの研究が存在する．まず，今枝らや南保らのように，NTT 日本語語彙体系 [7] のような辞書を用いて訂正する手法が挙げられる [8] [9]．次に Oyama らや大木らのように SVM のような機械学習を用いた手法が挙げられる [10] [11]．Oyama ら [11] は，誤用タイプを定義し，それらに対して機械学習による自動分類を行った．76 項目の誤用タイプを定義しており，事例数が少ない誤用タイプは存在しているが，実用的な精度を得ている．本稿と比較すると注目している対象が異なるため，示されている全ての誤用タイプを利用する必要はないと考えているが，誤用をカテゴライズしたものとして非常に詳細に検討されており，技術文書に対しても検討を進める必要がある．

自然文書中の誤りの傾向を利用することで，高精度な誤りを訂正を行うことを目指した研究が行われている．Mizumoto らは，相互添削型言語学習 SNS である Lang-8 の添削履歴を用いて，誤り訂正を行う技術を提案している [12]．笠原らは，日本語学習支援を目的とした誤り訂正に関する研究の一環として，格助詞の誤り訂正技術の開発を行った [13]．この技術では，日本語学習者の誤り傾向を利用して格助詞の訂正を行っている．要求仕様書において，頻出する誤り傾向をモデル化することができれば，誤り訂正の精度向上を実現できる可能性がある．英語で記述された文書を対象としても，同様の研究が行われている．例えば，Rozovskaya らは英語の非母語話者が記述した文書の前置詞訂正を行っている．彼らの研究においても，英語の非母語話者の誤り傾向を利用している [14]．

省略された主語の推定のために利用できる技術として，ゼロ照応解析がある [5] [6]．要求仕様書における主語が欠落している文の特定や訂正に利用できる可能性があると考えられる．

5. お わ り に

本研究では，目視によって企業で実際に使用された要求仕様書から誤りや曖昧さを抽出し，それらを 10 種類に分類した．また，実際に開発者がそれらの文を読む際に，誤解を生じるかについて考察した．

次いで，これらの分類について自動抽出が可能かどうかについての検討を行った．その結果，容易に実装と思われる分類については，目視による抽出よりもより精度の高い抽出ができたものがあった．さらに，この結果から今後の自動抽出実装における改善点を考察した．

今後の課題は，日本語形態素解析システムだけではなく，係り受けや照応解析も利用し，さらに他の分類も自動抽出できるように検討をすすめ，実装済みのツールについても，さらに改良を行うことである．

謝辞 本稿における文章の誤りや曖昧さの分類について，議論をしていただきました株式会社エクスモーションの玉木 淳治氏に感謝致します．

(注6)：http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP

文　　献

[1] 来間啓伸，中島震（監修），B メソッドによる形式仕様記述，近代科学社，2007.
[2] 大森洋一，荒木啓二郎，“自然言語による仕様記述の形式モデルへの変換を利用した品質向上に向けて，” 情報処理学会論文誌プログラミング (PRO)，vol.3，no.5，pp.18–28，2010.
[3] D. Kawahara and S. Kurohashi, “A fully-lexicalized probabilistic model for japanese syntactic and case structure analysis,” Proc. the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, pp.176–183, 2006.
[4] 河原大輔，黒橋禎夫，“自動構築した大規模格フレームに基づく構文・格解析の統合的確率モデル，” 自然言語処理，vol.14，no.4，pp.67–81，2007.
[5] 笹野遼平，黒橋禎夫，“大規模格フレームを用いた識別モデルに基づく日本語ゼロ照応解析，” 情報処理学会論文誌，vol.52，no.12，pp.3328–3337，2011.
[6] R. Sasano and S. Kurohashi, “A discriminative approach to japanese zero anaphora resolution with large-scale lexicalized case frames.,” Proc. IJCNLP, pp.758–766, 2011.
[7] 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩巳，小倉健太郎，大山芳史，林良彦，日本語語彙大系，岩波書店，1997.
[8] 今枝恒治，河合敦夫，石川裕司，永田亮，桝井文人，“日本語学習者の作文における格助詞の誤り検出と訂正，” 情報処理学会研究報告コンピュータと教育 (CE)，vol.2003，no.13 (2002-CE-068)，pp.39–46，2003.
[9] 南保亮太，乙武北斗，荒木健治，“文節内の特徴を用いた日本語助詞誤りの自動検出・校正，” 情報処理学会研究報告自然言語処理 (NL)，vol.2007，no.94 (2007-NL-181)，pp.107–112，2007.
[10] 大木環美，大山浩美，北内啓，末永高志，松本裕治，“非日本語母国話者の作成するシステム開発文書を対象とした助詞の誤用判定，” 言語処理学会第 17 回年次大会，pp.1047–1050，2011.
[11] H. Oyama and Y. Matsumoto, “Automatic error detection method for Japanese case particles in Japanese language learners’ writing,” Corpus, ICT, and Language Education, pp.235–245, 2010.
[12] T. Mizumoto, M. Komachi, M. Nagata, and Y. Matsumoto, “Mining revision log of language learning sns for automated japanese error correction of second language learners.,” Proc. of IJCNLP, pp.147–155, 2011.
[13] 笠原誠司，藤野拓也，小町守，永田昌明，松本裕治，“日本語学習者の誤り傾向を反映した格助詞訂正，” 言語処理学会第 18 回年次大会，pp.14–17，2012.
[14] A. Rozovskaya and D. Roth, “Generating confusion sets for context-sensitive error correction,” Proc. of the 2010 conference on empirical methods in natural language processing, pp.961–970, 2010.