Kobe University Repository : Kernel

PDF issue: 2025-07-02

日英文型の比較に基づく英作文誤用検出システムの構築

川村，晃市

(Degree)
博士（学術）
(Date of Degree)
2019-03-25
(Date of Publication)
2020-03-01
(Resource Type)
doctoral thesis
(Report Number)
甲第7372号
(URL)
https://hdl.handle.net/20.500.14094/D1007372

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。

博　士　論　文

日英文型の比較に基づく英作文誤用検出システムの構築

平成 31 年 1 月

神戸大学大学院国際文化学研究科

川村　晃市

目次

図　目次
表　目次
第 1 章　序論 …… 1
1.1　はじめに …… 1
1.2　英作文誤用検出システムの必要性 …… 1
1.3　本論文の目的と構成 …… 2
第 2 章　本研究に関わる定義および先行研究 …… 5
2.1　誤りの定義 …… 5
2.2　global error の定義 …… 6
2.3　英作文の誤用に関する先行研究 …… 7
2.3.1　応用言語学の視点から行われた誤用研究 …… 7
2.3.2　自然言語処理の視点から行われた誤用の自動検出に関連する研究 …… 8
第 3 章　本研究で使用する言語データと構文解析器 …… 9
3.1　はじめに …… 9
3.2　研究対象とする学習者の習熟度 …… 9
3.3　使用する言語データ …… 9
3.3.1　作文データ …… 9
3.3.2　教科書データ …… 10
3.4　利用する構文解析器 …… 10
第 4 章　誤用検出における原言語の利用 …… 12
4.1　はじめに …… 12
4.2　分析の手法 …… 12
4.2.1　使用するデータ …… 12
4.2.2　分析の手順 …… 13
4.3　分析結果と考察 …… 13
4.3.1　日本語と英語における意味上のずれについて …… 13
4.3.2　日本語主語の有無と英語の誤りの関係 …… 18
4.3.3　日英両作文の主語および述語の比較 …… 18

4.4　まとめ……20
第 5 章　主語および述語の抽出における構文解析器の利用 ……22
5.1　はじめに……22
5.2　分析の手法……22
5.2.1　構文解析器による主語候補および述語候補の抽出条件 ……22
5.2.2　人手判定による主語および述語の判定方法……23
5.2.3　使用するデータ ……25
5.2.4　分析の手順……25
5.3　結果と考察……26
5.3.1　日本語作文の主語・述語判定の結果……26
5.3.2　日本語作文の主語・述語判定の考察……28
5.3.3　英作文の主語・述語判定の結果 ……29
5.3.4　英作文の主語・述語判定の考察 ……31
5.4　まとめ……32
第 6 章　日英文型に基づく誤用検出の基準および手順の提案 ……33
6.1　はじめに……33
6.2　日英文型の規定 ……33
6.2.1　中学教科書における文法項目の調査……33
6.2.2　英語文法項目に対応する日本語……37
6.3　誤用判定基準の構築 ……41
6.4　誤用検出手順の提案 ……47
6.5　まとめ……48
第 7 章　誤用検出システムの構築および評価……49
7.1　はじめに……49
7.2　システムの概要 ……49
7.2.1　インターフェイスの概要 ……50
7.2.2　処理の流れ……55
7.3　検証……58
7.4　まとめ……63
第 8 章　結論 ……64

謝辞 ……67
参考文献 ……68
要旨 ……70
本研究に関する発表……72

図　目次

図 3.1　CaboCha による構文解析結果例 ……11
図 3.2　The Stanford Parser による構文解析結果例……11
図 4.1　第 4 章で使用する作文データ ……13
図 4.2　初級レベル学習者の英作文（B） ……16
図 4.3　母語話者の英語翻訳文（D） ……17
図 4.4　母語話者の英語添削文（F） ……17
図 6.1　自動誤用検出の手順 ……48
図 7.1　システムの構成 ……49
図 7.2　作文データ入力画面 1 ……50
図 7.3　作文データ入力画面 2 ……51
図 7.4　英作文の誤り検出結果画面 1……51
図 7.5　英作文の誤り検出結果画面 2……52
図 7.6　英作文の誤り検出結果画面 3……52
図 7.7　英作文の詳細情報画面 1 ……53
図 7.8　英作文の詳細情報画面 2 ……54
図 7.9　誤りの集計結果画面 ……54
図 7.10　データ処理の流れ……55

表　目次

表 2.1　多様な誤りの定義 …… 5
表 3.1　本研究で参考にした中学教科書……10
表 4.1　学習者の書いた日本語作文とそれに対応した誤りを含む英作文 ……14
表 4.2　bigram による計測結果 ……16
表 4.3　T 検定の結果……16
表 4.4　χ^2 検定の結果 ……17
表 4.5　文構造の一致数 ……18
表 4.6　構文解析器による抽出結果……19
表 5.1　本研究で採用した文の種類（構造上の分類）　……25
表 5.2　分析用作文データ一覧……25
表 5.3　日本語作文 A における主語候補・述語候補の分類結果と各分類の件数……27
表 5.4　日本語作文 B における主語候補・述語候補の分類結果と各分類の件数……28
表 5.5　英作文 C における主語候補・述語候補の分類結果と各分類の件数……30
表 5.6　英作文 D における主語候補・述語候補の分類結果と各分類の件数……30
表 6.1　NEW HORIZON 学習項目一覧 ……34
表 6.2　初級学習者にとって重要な文法項目一覧 ……35
表 6.3　自動検出できる可能性が高い英語文法項目 ……36
表 6.4　日本語文型 ……37
表 6.5　日本語文型に対応する英語文型……39
表 6.6　文型とその誤用判定基準……42
表 7.1　検証に使用した誤用判定基準……58
表 7.2　誤りの検出結果集計表……59
表 7.3　誤りと判定された英作文の例……61

第 1 章　序論

1.1　はじめに

近年，コンピュータ技術の発達によって膨大な言語資料を容易に分析することができるようになり，これまでの手作業による研究では気づくことのできなかった情報を抽出することが可能となった．これらの技術の進化にともない，自然言語処理の分野ではコーパスなどこれまでに蓄積された言語資料を活用して言語事実が明らかにされ，形態素解析，構文解析，意味解析，文脈解析といった言語解析技術が開発されている．そして，教育工学の分野では，教育支援や学習支援を目的として，これらの言語解析技術を取り込んだスペル訂正，用例検索，誤用検出，文章校正，機械翻訳などの作文支援システムが数多く開発されている．

このようにコンピュータ技術の発達により，さまざまな作文支援システムが開発されてきたにもかかわらず，英作文誤用検出システムに関しては日本の英語教育の現場で十分に活用されているとは言い難い現状がある．原因として，従来の誤用検出システムの多くは単一言語の作文データ，後述の目標言語（英語）の作文データのみを対象としているために学習者の意図を考慮できず意味の理解が不可能な文の誤りの検出ができないこと，また，添削の際，教師が検出にもっとも労力を要する学習者の文構造上の誤りを検出できない可能性が高いことなどが考えられる．それゆえに，対応関係にある二言語，いわゆる目標言語である英語だけでなく原言語である日本語についても分析対象とすることでこれらの問題に対処でき，また教師の負担を軽減することができる英作文誤用検出システムが必要とされている．

1.2　英作文誤用検出システムの必要性

英語学習者にとって，4 技能のうちライティングは「自然と習得されるものではなく，学校などで教育を受けないかぎり習得することが困難である」ことから（Grabe and Kaplan, 1996），他の 3 技能と比較して習得することが困難な技能であると言える．このような言語習得上の特性もあり，2004 年以降，大学などの高等教育機関では学習者のライティング能力向上を目的としたライティングセンターの設置が進められているが（吉田, 2010；太田他, 2013），まだ，その数は十分ではなく，利用できる回数および時間に制約があるため，学習者への支援は限定的なものに留まる．また，ライティングセンターの多くが「自立した書き手」を育成するために，支援の重点を思考プロセスおよび内容の推敲に置いていることから（吉田，2010），文単位の文章校正が行われないケースも少なくない．このようなライティ

ングセンターの支援は正確な文法知識を有さない初級・中級学習者のライティング能力向上には適していない．これらの問題に対処するためには，初級・中級学習者が独力で学習することができる英作文添削システムの構築が必要であると考えられる．一方，教師にとっては，ライティングの文法修正フィードバックの効果に否定的な見解と肯定的な見解が存在し，統一された指導方法が確立されていないことから（Truscott, 1966; Ferris, 2004），学習者の多様な誤用に対応しなければならず，ライティングは指導するにあたり準備に負担がかかる技能であると言える．このような現状を踏まえると，英作文添削システムを構築することは，学習者には自分の学習スタイルやストラテジーに責任を持っておこなう自律学習を促し，教師には添削にかかる負担を軽減するための支援となるため，学習者の言語習得と教師の言語指導の二つの観点から重要性が高いと考えられる．

英作文添削の中でも，重要なプロセスの一つとして誤りの検出がある．学習者の誤りは第二言語習得の観点から global error と local error に大別することができるが（Burt and Kiparsky, 1974），global error は文の全体構造に影響して内容の理解を阻害するため，教師が原言語の手掛かりなしで学習者の英作文に存在する global error を特定することは難しく，特に文の構造に関わる誤りはその検出に多大な労力を要する．また，既存の英作文誤用検出システムは主として local error を対象としており，原言語を取り入れて global error の検出に焦点を当てたシステムは管見の限り存在しない．これらの理由から，教師への教育支援のために global error を自動検出できる英作文誤用検出システムの必要性は高いと考えられる．そこで本研究では，学習者の global error の中でも特に文構造の主成分に関する誤りに対応できる誤用検出システムの構築を行う．誤用の検出には，学習者の意図を考慮するために原言語（日本語）の作文データも利用する．

1.3　本論文の目的と構成

本論文の目的は，原言語を取り入れ，日英文型の比較に基づいた global error の自動検出に必要な手法および手順について提案し，英作文誤用検出システムを構築することである．

そこで，本論文の目的を達するために，1）目標言語（英語）と併せて原言語情報（日本語）を利用した誤用検出について考察し，誤用検出の可能性を示すこと（第 4 章），また，2）構文解析器（英語：The Stanford Parser，日本語：CaboCha）を利用した，誤りを含む文から主語および述語を抽出する方法を提案し，global error となりうる文の主成分である主語および述語がどの程度正確に抽出できるのか検証すること（第 5 章），そして，3）日英文型

の比較による誤用判定基準を用いた global error を自動検出する方法を提案し（第 6 章），その手法および手順に妥当性があるのか，提案に基づいた誤用検出システムを開発して検証すること（第 7 章）を行う．

まず，第 1 章「序論」においては，本研究の背景，英作文誤用検出システムの必要性，目的について述べる．そして本論文の全体構成について説明する．

第 2 章「本研究に関わる定義および先行研究」においては，まず，誤りの定義を確認し，次に，本研究の主たる検出対象である global error の定義を確認する．そして，英作文の誤用に関して，応用言語学と自然言語処理の各観点から先行研究を概観し，英作文誤用検出研究の残された課題を整理する．

第 3 章「本研究で使用する言語データと構文解析器」においては，研究対象とする学習者（日本人英語学習者）の英語習熟度，使用する言語データ（日本語作文データおよびそれに対応する英作文データ，英語添削文データ，英語翻訳文データ，そして教科書データ），利用する構文解析器（英語：The Stanford Parser，日本語：CaboCha）について概説する．

第 4 章「誤用検出における原言語の利用」では，誤用検出において原言語（日本語）を利用することの重要性について述べ，原言語を利用した学習者英語の誤用検出の可能性を探る．具体的には，英語学習者の原言語と目標言語の間に意味上のずれがあるのか，英語学習者の目標言語と英語母語話者の原言語の間に品詞構成上の差異が存在するのか，そして，原言語の主語の有無が目標言語の誤りに影響するのかを分析したうえで，原言語を用いた誤用検出の新しい手法を提案する．

第 5 章「主語および述語の抽出における構文解析器の利用」では，既存の構文解析器（英語：The Stanford Parser，日本語：CaboCha）を誤りが含まれる文に対して使用し，得られた構文解析結果（依存構造情報）から主語および述語を抽出する方法を提案する．また，提案するにあたって，主語および述語の抽出の正確性についての検証を行う．

第 6 章「日英文型に基づく誤用検出の基準および手順の提案」では，日英文型の比較による誤用判定基準を用いた新たな誤用検出方法を提案する．具体的には，初級学習者が習得すべき重要な英語文法項目を選定し，その英語文法項目に対応する日本語と合わせて日英文型を規定する．そして，日英文型比較による誤用判定を行うための基準およびその手順について提案する．

第 7 章「誤用検出システムの構築および評価」では，第 4 章から第 6 章で提案した誤用検出に必要な基準および手順に従って誤用検出システムを構築し，システムの可能性を探る

ために検証を行う．

第 8 章「結論」では，第 1 章から第 7 章までを統括し，英作文誤用検出システムがいかにして英語教育支援のための有効なツールとなり得るのか論じる．また，今後の研究課題についても述べる．

第 2 章　本研究に関わる定義および先行研究

前章では，英作文誤用検出システムのうち，global error に対応できる誤用検出システムの必要性ついて述べた．本章では，誤りを自動で検出するために必要な手法および手順の検討に先立ち，誤りの定義を確認する．また，本研究の主たる検出対象である global error についても定義を確認する．そして，英作文の誤用に関して，応用言語学と自然言語処理の各観点から先行研究を概観し，英作文誤用検出研究の課題について整理する．

2.1　誤りの定義

誤りとは何か，ということについては様々な見解が存在する．本節では，まず Coder（1967）の誤りの定義を確認し，次に他の研究者による定義を見ていく．

表 2.1　多様な誤りの定義

研究者	誤りの分類	説明
Coder（1967）	error（規則的な誤り）	言語能力上の誤り
	mistake（一過性の誤り）	言語運用上の誤り
Richards（1971）	interlingual error（言語間誤り）	学習者の母語干渉による誤り
	intralingual error（言語内誤り）	L2 規則の不適切な使用による誤り
Dulay and Burt（1973）	developmental error（発達上の誤り）	学習言語を L1 とする幼児と同じ種類の誤り
	interference error（母語干渉による誤り）	学習者の母語規則を持ち込んだ誤り
	ambiguous error（曖昧な誤り）	developmental error と interference error に跨る誤り
	unique error（特殊な誤り）	developmental error と interference error に属さない誤り
Burt and Kiparsky（1974）	global error（全体的誤り）	意味が理解できない誤り
	local error（局所的誤り）	意味が理解できる誤り

Coder（1967）は「誤り」には error と mistake があり，前者を組織的な competence（言語

能力）に関わる誤り，後者を非組織的な performance（言語運用）に関わる誤りと定義している．しかしながら，Coder（1967）の定義には曖昧性が存在しており，英作文に含まれる誤りを言語運用上の誤りか言語能力上の誤りか明確に分類することが不可能だという問題を含んでいる．このように誤りを明確に定義できないことが定義の多様性の原因ともなっている．

表 2.1 はそれぞれの研究者が提示した誤りの分類モデルである．なお，文法用語の翻訳および説明にあたっては英語教育用語辞典（白畑他，1999），および応用言語学事典を参考とした（小池他，2003）．ただし，訳語が明示されていないものに関しては，筆者が適宜訳語を充てた．

表 2.1 にある全ての定義に共通することは曖昧性を含んでいることである．そのため学習者の目標言語である英作文だけを手掛かりとして誤用を明確に分類することは不可能であり，学習者の意図を確認することが必要となる．

2.2　global error の定義

次に，検出対象である global error の定義について確認する．Burt and Kiparsky（1974），Burt（1975），Richards and Schmidt（2002），Yang and Meng（2013）は global error について以下のように定義している．

Burt and Kiparsky（1974）は誤りを global error と local error の二つに分類し，前者を意味理解が極めて困難になる，あるいは不可能になる誤り，後者を意味理解に影響しない誤りと規定している．Yang and Meng（2013）は local error は文法の誤り，global error は文章の展開，構成，文体の誤りとみなしている．また，文構造の主成分に関する誤りや文の構造に影響を与える誤りが global error であるというように（Richards and Schmidt, 1992; Burt, 1975），統語論的な視点で規定する研究者も存在する．以下，英語教育用語辞典にある global error と local error の例文を示す．

global error の例（下線部）：

- We arrived Disneyland, and I got the shopping with my friends. <u>Because it is crowed to approach the last time</u>. We took first “Pirates of the Caribbean.”
- ディズニーランドに着いて，友達と買い物をしようとしたが，混んでいたので買い物は最後にすることにして，まず最初「カリブの海賊」に乗りに行った．

local error の例：

・I go to Tokyo yesterday.

上述のように global error にはさまざまな定義が存在するが，共通しているのは，目標言語（英語）以外の情報がない場合にはその誤りを特定することが非常に困難であるということである．しかしながら，例文のように原言語（日本語）情報がある場合には誤りを特定することが可能となる．また，文構造の主成分に関する誤りを特定することも可能である．このことから，global error の検出には原言語（日本語）情報が不可欠と言える．

以上のことを鑑み，本研究では global error の一部である文構造（主語および述語）に関する誤りを主たる検出対象とする．また，目標言語自体に誤りがない場合であっても，原言語との間に文構造上の相違があるものは誤りとして検出対象とする．

2.3　英作文の誤用に関する先行研究

global error を検出する際，その言語の文型，語順，語彙などの特徴を検討することは重要である．しかしながら，原言語と目標言語の言語間距離が大きい場合，言語の構造的特徴の不理解から誤りが生じることが多い．そこで本節では，日本人英語学習者の英語基本文構造に不可欠な主語や述語に関する誤用研究について，応用言語学と自然言語処理の二つの視点から概観する．

2.3.1　応用言語学の視点から行われた誤用分析

日本人英語学習者の主語や述語について，それらの誤用を応用言語学の立場で分析した研究として，坂内・佐々木（2005），水品・麻柄（2007），工藤（2009），山内・内田（2011），Trent（2012）が挙げられる．

坂内・佐々木（2005）は日本人高校生のもつ主語名詞句と述語動詞間の一致の知識について，三人称単数のSの誤用の観点から調査した．その結果，主語と動詞との間に入れられる副詞や関係節の位置は一致形態素の使用または脱落に影響することを明らかにした．水品・麻柄（2007）は日本人中高生の英語の主語把握の誤りについて，母語である日本語の知識が干渉しているのか調査したところ，多くの中高生が母語干渉によって日本語の「～は」をそのまま英語の主語として誤用する実態を明らかにした．工藤（2009）は英語ライティング能力の異なる日本人高校生のglobal errorの特徴について調査し，ライティング能力が高い学習

者の誤用は低い学習者と比べて内容的エラーが多く出現すること，またライティング能力が低い学習者の誤用は高い学習者と比べて主述関係を含めた文全体の構造が不明確な言語的エラーが多く出現することを明らかにした．山内・内田（2011）は日本人大学生が産出する英作文の誤用を分析し，母語である日本語の統語的特徴の影響について調査した．その結果，日本語の主題／主格については「XはYだ」構造（X, Yは名詞）が借用可能になる条件が厳しいこと，また，品詞／述語項構造については「Xは（が）＋感情・感覚を表す述語」という構造が品詞や述語項構造で一致していないために借用可能性が低いこと，そして，ゼロ照応については「省略している」という意識をせず，日本語と同様に使用している可能性があることなどを明らかにした．Trent（2012）は初級レベル大学生の主語のエラーについて分析し，初級学習者は日本語と英語のどちらの言語においても正しい文主語の概念を持っていないと結論づけた．

2.3.2　自然言語処理の視点から行われた誤用の自動検出に関連する研究

主語や述語について，自然言語処理の視点から行われた誤用の自動検出に関連する研究としては永田（2013）が挙げられる．

永田（2013）は英語文における主語の人称・数と動詞の人称・数とが一致していない誤りを検出する手法について検討し，構文解析を利用する代わりに品詞解析と句解析を利用し，英語の性質に基づいた規則により主語と動詞の関係を抽出した方が正確さは高くなること，そして，文長が短い学習者の英語文に対して有効であることを確認した．

上述のように，英作文誤用検出に関して応用言語学と自然言語処理の各観点から先行研究を概観したところ，応用言語学の分野では基本文構造に不可欠な成分である主語や述語の理解は目標言語（英語）の正確性に大きく影響を及ぼしていることが明らかとなっており，その知見を教育の現場に生かすために様々な研究がなされていることがわかった．しかしながら，自然言語処理技術を用いた学習者の主語および述語に関する誤用検出の研究はまだ少なく，原言語情報を取り入れた研究は管見の限りないことなど，残された課題があることも明らかとなった．

以上のことから，本研究では誤用文の構造および学習者の意図を理解するために目標言語（英語）と併せて原言語（日本語）を取り入れ，文構造に不可欠な成分（主語および述語）に関する global error に対応できる誤用検出システムの構築を目指す．

第 3 章　本研究で使用する言語データと構文解析器

3.1　はじめに

すでに述べたように，本研究では教師の英作文指導を支援するため，global error の自動検出に必要な手法を提案し，英作文誤用検出システムの構築を行う．本章では global error の検出に必要な手法を提案するにあたり，どのような学習者を対象としてどのような作文データを使用するのか，また，どのような構文解析器を使用するのか整理しておく．

3.2　研究対象とする学習者の習熟度

本研究で対象とする学習者の英語習熟度はヨーロッパ言語共通参照枠（CEFR）の A1（Breakthrough）であり，初級レベル学習者に相当する．なお，ここでいう初級レベル学習者は擬似初心者を指す．疑似初心者とは「中学校，高等学校で 6 年間の英語教育を受けたにもかかわらず，第二言語学習者として順調な発達段階を踏んでいない学習者」のことである（中村他，2005）．

3.3　使用する言語データ

3.3.1　作文データ

本研究では，日本人大学生の 1 回生 68 名に対して日本語作文およびそれに対応する英作文の課題を課し，日本語作文・英作文は各 110 エッセイ，各 1499 文を収集した．

作文データの執筆条件として，日本語作文ではプロンプト A :「大学生にとってアルバイトをすることは重要である」，プロンプト B :「国内では全てのレストランで喫煙が完全に禁止されるべきだ」のどちらかを選択し，文字数は 400 字，制限時間は 30−40 分，辞書の使用は不可とした．また，英作文では日本語作文で選択したプロンプトを翻訳し，制限時間は 50−60 分，辞書の使用は不可とした．

日本語作文データの収集方法は，国立国語研究所の作文対訳 DB のように日本語学習者に目標言語である日本語作文を要求し，その作文を執筆者に母語で翻訳させ原言語のデータを採取するという手順を踏むと，学習者は問題に直面すると誤りを犯しそうな構造や語彙を避けるコミュニケーション・ストラテジーをとる可能性がある（Tarone, 1980）．この可能性を排除するため，本研究では日本人英語学習者にまず原言語である日本語による作文を要求し，その後，その日本語作文を目標言語である英語に翻訳するという順序で作業を課した．留意点として，日本語作文を書く際には後に英語に翻訳することを告知していない．

3.3.2　教科書データ

日英文型を規定する際には中学英語教科書を参考にした．教科書は平成 20 年 3 月に改訂され，平成 24 年度より実施された新学習指導要領に基づく教科書全 6 種類 18 冊を使用した．表 3.1 は参考にした中学英語教科書のリストである．

表 3.1　本研究で参考にした中学教科書

教科書	著者	出版社
COLUMBUS21 ENGLISH COURSE　1，2，3	東後勝明　他	光村図書出版
NEW CROWN ENGLISH SERIES 1，2，3	高橋貞夫　他	三省堂
NEW HORIZON English Course 1，2，3	笠島準一　他	東京書籍
ONE WORLD English Course 1，2，3	松本茂　他	教育出版
SUNSHINE ENGLISH COURSE 1，2，3	松畑熙一　他	開隆堂出版
TOTAL ENGLISH NEW EDITION 1，2，3	矢田裕士　他	学校図書

3.4　利用する構文解析器

本研究では CaboCha 0.69（Kudo and Matsumoto, 2003）と The Stanford Parser 3.6.0（Klein and Manning, 2003）をそれぞれ日本語と英語の構文解析器として使用する．また，一文は一句点までの文とし，抽出は一文毎に行う．

ここでは，日本語構文解析器 CaboCha および英語構文解析器 The Stanford Parser を利用して取得できる情報について，日本語作文「今日は良い天気です.」と英作文“It is fine today.”を例にして示しておく．

例文）今日は良い天気です.

```
<sentence>
<chunk id="0" link="2" rel="D" score="-1.137013" head="0" func="1">
<tok id="0" feature="名詞,副詞可能,*,*,*,*,今日,キョウ,キョー">今日</tok>
<tok id="1" feature="助詞,係助詞,*,*,*,*,は,ハ,ワ">は</tok>
</chunk>
<chunk id="1" link="2" rel="D" score="-1.137013" head="2" func="2">
<tok id="2" feature="形容詞,自立,*,*,形容詞・アウオ段,基本形,良い,ヨイ,ヨイ">良い</tok>
</chunk>
<chunk id="2" link="-1" rel="D" score="0.000000" head="3" func="4">
<tok id="3" feature="名詞,一般,*,*,*,*,天気,テンキ,テンキ">天気</tok>
<tok id="4" feature="助動詞,*,*,*,特殊・デス,基本形,です,デス,デス">です</tok>
</chunk>
</sentence>
```

図 3.1　CaboCha による構文解析結果例

例文）It is fine today.

```
((u'fine', u'JJ'), u'nsubj', (u'It', u'PRP'))
((u'fine', u'JJ'), u'cop', (u'is', u'VBZ'))
((u'fine', u'JJ'), u'nmod:tmod', (u'today', u'NN'))
```

図 3.2　The Stanford Parser による構文解析結果例

第 4 章　誤用検出における原言語の利用

4.1　はじめに

本章では，原言語を利用することの重要性を述べ，原言語を利用した学習者英語の誤用検出の可能性を探る．そのために，まず，初級英語学習者が産出する目標言語とその原言語との間に意味上のずれがあるのか，また，初級英語学習者が産出する目標言語と英語母語話者が産出する原言語との間に品詞構成上の差異が存在するのか，原言語の主語の有無が目標言語の誤りに影響するのか，実際の学習者作文データを用いて分析を行い，原言語を利用した誤用検出の妥当性を検証する．そして，初級英語学習者の英作文データと日本語作文データの主語と述語をそれぞれ比較して学習者の誤りの傾向を分析する．この分析過程において構文解析器を利用し，原言語を用いた学習者英語の誤用検出の新しい手法の提案をする．

4.2　分析の手法

分析は日本語と英語における意味上のずれ，日本語主語の有無による影響，日英両作文の主語および述語の比較による誤用検出の可能性という三つの視点から，n-gram 法による類似度分析，POS Tagger を用いた品詞構成調査，日英両作文の主語および述語の比較分析を行う．

4.2.1　使用するデータ

収集した日本語作文・英作文で各 110 エッセイ，各 1499 文のうち，初級レベル学習者が産出した英作文（B）をランダムで 50 文抽出し，それを上級レベル学習者と英語母語話者がそれぞれ添削した英語添削文（E），（F）そして，初級レベル学習者が産出した英作文（B）に対応する日本語作文（A）とそれを上級レベル学習者と英語母語話者がそれぞれ翻訳した英語翻訳文（C），（D）各 50 文使用する．なお，ここでいう上級レベル学習者は非英語母語話者であり，英語圏の大学院を修了した者を指す．また，英語母語話者は日本語能力検定試験において N1 認定を受け，日本の大学への正規留学経験を持つ者を指す．分析に使用する言語データの一覧は以下に示す（図 4.1）．

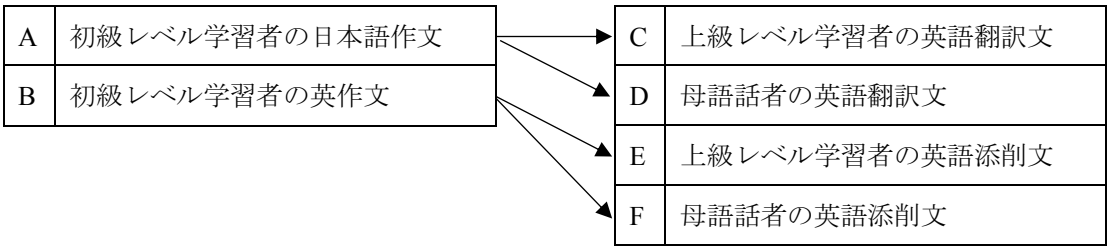

図 4.1　第 4 章で使用する作文データ

4.2.2　分析の手順

日本語と英語における意味上のずれにおいては，実例を挙げながら global error に焦点を当て，意味上の比較を行う．そして 50 文全てに対して BC 間の類似度と BE 間の類似度，そして BD 間の類似度と BF 間の類似度を比較し，意味上のずれがあるか判断する．類似度の比較は bigram によって行う．さらに 50 文の文章間類似度の平均値を算出し，差があるかどうかの判定を T 検定により行う．また，形態素の観点から，B，D と F をそれぞれ，名詞（N），動詞（V），形容詞（Adj.），副詞（Adv.），その他（Mis.）に分類してその構成比を調べ，品詞構成上の違いがあるか判断する．その際，差があるかどうかの判定は χ^2 検定により行う．

日本語主語の有無による影響においては，A の文構造（主語・述語）と B の文構造（主語・述語）のが一致しているか判定し，B の主語の有無に応じて分類する．

日英両作文の主語および述語の比較による誤用検出の可能性においては，構文解析器を利用し，A と B のそれぞれ 100 文から主語および述語を自動で抽出し，それぞれが一致しているか調べる．その際，主語または述語を複数抽出，もしくは抽出できなかった文を除いたうえで，主語と述語が一致，主語のみ一致，述語のみが一致，主語と述語が不一致の 4 つに分類し，誤用傾向を調べる．そして，それぞれのカテゴリーにおいて，B の誤用箇所（主語・述語・なし）を下位分類し，その特徴を調べる．構文解析器による主語候補および述語候補の抽出条件，主語および述語の判定基準については第 5 章で詳しく述べる．

4.3　結果と考察

4.3.1　日本語と英語における意味上のずれについて

A. 日英両作文の比較からみる意味上のずれの実例

表 4.1 は学習者の書いた日本語作文とそれに対応した誤りを含む英作文 10 文を列挙したものである．

表 4.1　学習者の書いた日本語作文とそれに対応した誤りを含む英作文

No.	学習者の日本語作文	学習者の英作文
1	分煙と言っても、少し壁で囲われているだけだから子供達だって入れてしまうし、喫煙席のそばになれば壁の意味がないほどの**レストランもある**。	But, a little wall hold therefore, in the children, **smoking room** near **is** no wall meaning.
2	よくテレビで、アルバイトのしすぎで、学校に行けなくなってしまう人や、精神的に追いやられて、うつ病になってしまう**人もいます**。	**I** well **watch** on TV, the student can't go to university, they are damaged mental and they are a sick.
3	**２つ目は**、早死にしてしまう**ことです**。	Secondly, **you may die** more fast.
4	**これは**、集団生活でしか得られない**力である**。	**It's get** only in many people live world.
5	**レストランは**様々な人が利用する場所**であるからです**。	**Restaurants use** place many people.
6	レストランなどの飲食する場所でない場合は分煙などをすれば良いと思うが、食べる時ににおいなどでせっかくの気分が悪くなったり、味にも影響があると**私は考える**。	There **is good idea** that separate smoking space, but smoking is feeling bad for not smoker in eating place.
7	これらの理由から**私は**大学生がアルバイトをすることは大切だと**思います**。	These **are reason** that I think important to does part time job.
8	**私は**大学生がアルバイトをすることは、大切だと**思う**。	**It is important** that college student do part time job.
9	大学生にとってアルバイトをすることは重要であると**私は思う**。	A **part time job is important** for college students.
10	だから喫煙はレストランにおいて禁止されるべきだと**思いました**。	Therefore **I think** that the smoking should be banned at a restaurant.

注：日・英両作文共に全て原文である．太字箇所は主語と述語である（英作文においては推定）．

まず，上表右列の学習者の英作文について，左列の原言語（日本語）を手掛かりとして利用せず，文構造の主成分部分（主語および述語部分）の誤り，いわゆる global error を特定できれば抽出する．次に，原言語（日本語）を手掛かりとして利用し，文構造の主成分部分（主語および述語部分）の誤りを特定し抽出する．そして，それらを比較する．

i）原言語（日本語）を手掛かりとして利用しなかった場合

まず，10 文中 3 文（No. 4, 6, 7）に誤りがあると推測できる．そのうち No.4 のみ global error と特定できる．No.6 と No.7 は主語と動詞の不一致である local error と判定することができる．残りの文に関しては主成分部分に誤りがあるかどうか特定できない．

ii）原言語（日本語）を手掛かりとして利用した場合

10 文全てに誤りがあると判定できる．そのうち No.10 を除く全ての文が global error と特定できる．No.10 は動詞の時制誤りの local error と判定することができる．

これらを比較すると，原言語（日本語）を手掛かりとして利用することによって，学習者の英作文に global error があるのかどうか判定することができる．それゆえに，誤用検出における原言語の利用は重要性が高いと言える．

B. bigram による計測結果

表 4.2 は BC，BE，BD と BF 間の bigram 類似度の平均値（Ave.）および標準偏差（SD）の値をまとめたもの，表 4.3 は T 検定の結果をまとめたものである．

T 検定の結果，上級レベル学習者，母語話者に関わらず，原言語を参考にした英作文の添削文と原言語を参考にしない日本語作文の翻訳文には差があり，意味上のずれが存在することが明らかとなった．

表 4.2　bigram による計測結果

No.	BC	BD	BE	BF
1	0.17	0.23	0.63	0.41
2	0.38	0.55	0.47	0.36
3	0.39	0.40	0.46	0.57
4	0.33	0.28	0.64	0.61
5	0.27	0.53	0.73	0.54
・	・	・	・	・
50	0.16	0.30	0.40	0.20
Ave.	0.34	0.32	0.55	0.50
SD	0.13	0.11	0.16	0.15

表 4.3　T 検定の結果

BC と BE	$p = 0.00$ 5%を超えないので有意に差がある
BD と BF	$p = 0.00$ 5%を超えないので有意に差がある

C. 4 大品詞構成比

図 4.2，4.3，4.4 は，それぞれ B，D，F を形態素解析して得られた品詞構成比である．また，表 4.4 は χ^2 検定の結果をまとめたものである．

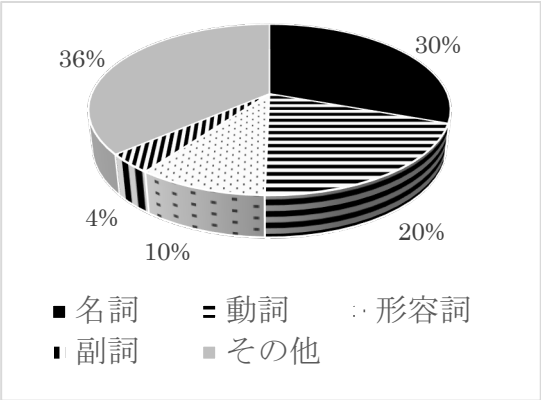

図 4.2　初級レベル学習者の英作文（B）

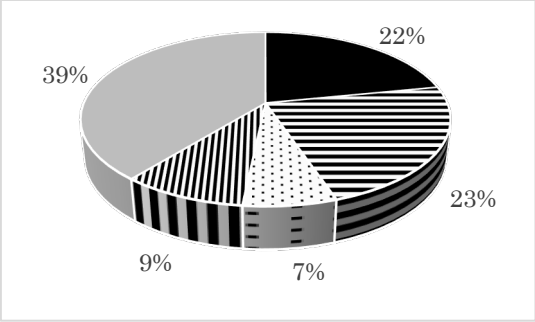

図 4.3　母語話者の英語翻訳文（D）

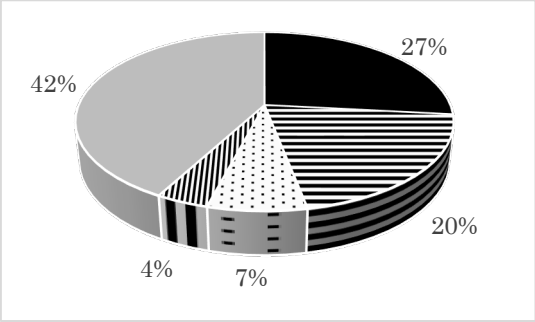

図 4.4　母語話者の英語添削文（F）

表 4.4　χ^2 検定の結果

	BD	BF
N	有意に差がある（χ^2=12.30, p=.001）	有意に差がない（χ^2=1.85, p=.174）
V	有意に差がない（χ^2=1.64, p=.201）	有意に差がない（χ^2=0.01, p=.918）
Adj.	有意に差がない（χ^2=3.32, p=.068）	有意に差がない（χ^2=2.29, p=.131）
Adv.	有意に差がある（χ^2=15.05, p=.000）	有意に差がない（χ^2=0.00, p=1.000）

χ^2 検定の結果，BD 間では名詞および副詞において差があるものの，BF 間では 4 大品詞すべてにおいて差がないことが明らかになった．この結果が意味することは，原言語情報がない場合，母語話者の添削文は学習者の英作文と近くなり，学習者が意図した表現から乖離している可能性がある，つまり意味上のずれが存在するということである．

以上のことから，原言語情報のない既存の学習者コーパスの英作文は学習者が意図した表現と異なる可能性が高い．学習者の言語習得状況を正確に把握し，global error を正確に検出するためには学習者の目標言語と併せて原言語を収集して分析する必要があると言える．

4.3.2　日本語主語の有無と英語の誤りの関係

表 4.5 は学習者の日本語作文（A）とそれに対応する英作文（B）の主語および述語の一致・不一致を日本語作文（A）の主語の存在別に分類したものである．

表 4.5　文構造の一致数

A の主語の有無	A と B の文構造の一致	一致数
あり	一致	7
あり	不一致	19
なし	一致	15
なし	不一致	9

上表より，学習者は日本語で主語が存在する文章を英語で表現することを苦にしていることがわかる．これは，原言語の主語の存在が目標言語の誤りに影響を及ぼしていることを示しているが，詳細については 4.3.3 で検証する．

これまでの結果より，学習者の言語習得状況を正確に把握し，global error を正確に検出するためには，学習者の目標言語と併せて原言語を収集する必要があるということが判明し，原言語情報を利用した誤用検出の妥当性は高いことが示された．

4.3.3　日英両作文の主語および述語の比較

比較は，学習者の日本語作文の主語および述語がその日本語作文に対応する英作文の主語および述語と一致するかどうか，確認することによって行う．ここでは「一致」は意味的に一致しているかを示すものであり，正用・誤用を示すものではない．つまり，ここでは正用・誤用に関わらず，全ての英作文を原言語と目標言語の主語と述語の一致に応じて分類して考察する．ただし，原言語と目標言語の主語と述語が不一致であるが正用文であるもの，いわゆる意訳文は初級学習者の誤りの傾向を探るという趣旨から外れるため，分析の射程外として省く．

表 4.6 は，学習者の日本語作文（A）とそれに対応する英作文（B）の主語および述語の意味的一致・不一致を，I 主語と述語が一致，II 主語のみ一致，III 述語のみが一致，IV 主語と述語が不一致の 4 つに分類し，英作文の誤りの箇所についてまとめたものである．留意点として，英作文の誤りの部分は主語と述語で重複して出現する場合がある．

表 4.6　構文解析器による抽出結果

<table>
<tr><th colspan="2" rowspan="2">構文解析器による抽出</th><th colspan="3">日本語作文と英作文の一致</th><th colspan="3" rowspan="2">英作文の正用・誤用数</th></tr>
<tr><th>主語</th><th>述語</th><th></th></tr>
<tr><td rowspan="12">68</td><td rowspan="3">I</td><td rowspan="3">○</td><td rowspan="3">○</td><td rowspan="3">19</td><td rowspan="2">誤用</td><td>主語</td><td>0</td></tr>
<tr><td>述語</td><td>4</td></tr>
<tr><td colspan="2">正用</td><td>15</td></tr>
<tr><td rowspan="3">II</td><td rowspan="3">○</td><td rowspan="3">×</td><td rowspan="3">14</td><td rowspan="2">誤用</td><td>主語</td><td>1</td></tr>
<tr><td>述語</td><td>9</td></tr>
<tr><td colspan="2">正用</td><td>3</td></tr>
<tr><td rowspan="3">III</td><td rowspan="3">×</td><td rowspan="3">○</td><td rowspan="3">4</td><td rowspan="2">誤用</td><td>主語</td><td>2</td></tr>
<tr><td>述語</td><td>3</td></tr>
<tr><td colspan="2">正用</td><td>0</td></tr>
<tr><td rowspan="3">IV</td><td rowspan="3">×</td><td rowspan="3">×</td><td rowspan="3">31</td><td rowspan="2">誤用</td><td>主語</td><td>26</td></tr>
<tr><td>述語</td><td>25</td></tr>
<tr><td colspan="2">正用</td><td>1</td></tr>
</table>

まず，主語と述語の両方が一致した場合の英作文では，19 文中 15 文が正用文であり，4 文が誤り文である. 正用文の特徴は,「私は～する」という単純な文構造である第一文型（SV）と第三文型（SVO）の文が多い．一方，誤り文の特徴としては述語部分で誤用がみられ，内容的誤り（意味誤り）ではなく，すべて形態（3 単現の S の欠落，時制の誤り）や用法（be 動詞・一般動詞が同時に使用，不定詞と助動詞を同時に使用）の言語的誤り（文法誤り）である．ここから，構文解析器を使用した誤用検出では述語部分の文法誤りに注意を払う必要があるということが明らかとなった．

次に，主語のみが一致した場合の英作文では，14 文中 3 文が正用文，10 文が誤り文であり，1 文が分析対象外の文（意訳文）である．正用文の特徴は 3 文中 2 文が「～である」，「～です」という状況を説明する be 動詞を使用した基礎的な表現である．一方，誤り文の特徴としては語彙知識の不足による誤りが半分を占める．ここでいう語彙知識の不足とは，そもそもその語彙自体が難しいものや，典型的な日本語独自の表現であるが英語で表現すると難しい語彙を未習得である状態を指す．例えば「～を阻害してしまう」「～で食欲をわ

かせます」などの語彙である．さらに「～しなければならない」，「～はずだ」といった助動詞に関する誤りも多い．唯一の主語に関する誤用は形態（3 単現の S の欠落）の誤りであった．ここから，構文解析器を利用した誤用検出では述語部分の語彙誤りに注意を払う必要があるということが明らかとなった．

また，述語のみが一致した場合の英作文では，4 文全てが誤り文である．誤り文の特徴としては語彙知識の不足による誤用が多く，基本構造の不理解による誤りも存在する．ここでの語彙知識の不足による誤りも，典型的な日本語独自の表現であるが英語で表現すると難しい語彙の未習得が原因として起こるものである．例えば「～大変です」「～苦手です」といった語彙である．ここからわかることとして，構文解析器を利用した誤用検出では主語が一致していない時，述語の一致・不一致に関わらず，英語文を全体的に確認する必要があるということである．

最後に，主語と述語がどちらも一致しない場合の英作文では 31 文中 1 文が正用文，29 文が誤り文であり，1 文が分析対象外の文（意訳文）である．また，正用文の 1 文は構文解析器で主語と述語を正確に抽出できなかった文である．誤り文の特徴としては主語が無生物主語であるものが 23 文，人称代名詞以外の主語は 28 文となっており，主語をうまく英語で表現できていない文が多い．語彙知識の不足による主語・述語の誤りも多く見られるが，there 構文と形式主語 it のように特殊な構文に関する誤りも存在する．ここからわかることとして，構文解析器を利用した誤用検出では主語および述語が一致していないものはほぼ誤りであると断定でき，特に主語がなにか特定する必要があるということである．

4.4　まとめ

本章では，日本人初級英語学習者を対象として，日本語と英語における意味上のずれ，日本語主語の有無による影響，日英両作文の主語および述語の比較による誤用検出の可能性という三つの視点から，原言語情報を利用した学習者英語の誤用検出の可能性を探った．以下，それぞれに得られた結果をまとめる．

日本語と英語における意味上のずれに関しては，日本人初級英語学習者が産出する目標言語（英語）とその原言語（日本語）との間には意味上のずれがあり，学習者が産出する目標言語と英語母語話者が産出する原言語との間には品詞構成上の差異も存在することが明らかとなった．

日本語主語の有無による影響については，原言語（日本語）の存在が目標言語（英語）の

誤りに影響を及ぼしていることが明らかとなった．したがって，学習者の言語習得状況を正確に把握し，global error を正確に検出するためには，学習者の目標言語と併せて原言語を収集する必要があるということが判明し，原言語情報を利用した誤用検出の妥当性は高いことが示された．

日英両作文の主語および述語の比較による誤用検出の可能性については，構文解析器を利用した学習者英語の誤用分析によって，主語と述語が一致する場合は述語部分の文法誤り，主語のみが一致する場合は述語部分の語彙誤り，述語のみが一致する場合は英語文全体，主語と述語がどちらも一致しない場合は主語の特定，特に無生物主語・人称代名詞以外の主語に問題があることが明らかとなった．

これらの結果より，構文解析器を利用し，原言語情報を用いた学習者英語の誤用検出の新しい手法の可能性が示された．次章では，原言語を利用した学習者英語の誤用検出システムを構築するために，構文解析器を利用して原言語および目標言語から基本文成分（主語および述語）を自動検出することが可能か検証を行う．

第 5 章　主語および述語の抽出における構文解析器の利用

5.1　はじめに

本章では，初級英語学習者の日本語作文データを原言語，それに対応する誤りを含む英作文データを目標言語として，それぞれの作文データに対し，既存の日本語および英語構文解析器（日本語：CaboCha，英語：The Stanford Parser）を利用して構文の解析結果（依存構造情報）を取得し，そこから独自の条件に従って選定した主語候補および述語候補が，人手によって抽出した主語および述語と比較してどの程度一致するのか，そして選定の正確性は文構造部分の誤用検出を行ううえで実用レベルにあるのか検証する．これにより，構文解析器を利用した新たな英作文の誤用検出手法の確立を目指す．

5.2　分析の手法

分析は以下の二点に沿って行う．

Q1　日本語作文に対する構文解析器の解析結果から選定した主語候補および述語候補と人手によって抽出した主語および述語と比較すると，どの程度一致しているのか

Q2　英作文に対する構文解析器の解析結果から選定した主語候補および述語候補と人手によって抽出した主語および述語と比較すると，どの程度一致しているのか

簡潔に記述するため，以降，人による判定を人手判定，構文解析器による抽出を自動抽出と言うこととする．

5.2.1　構文解析器による主語候補および述語候補の抽出条件

日本語および英語に関するそれぞれの主語候補および述語候補の抽出条件は，構文解析器を利用して取得した解析結果から，下記にある独自の条件に従って主語候補と述語候補を選定する．

日本語作文に対しては，構文解析の結果から得た最終文節を述語候補とし，その述語候補と依存関係にある格助詞の「が」，または係助詞の「は」，「も」を含む文節を主語候補としている．手順としては，構文解析器の結果から述語候補とリンクするすべての文節を対象として，これらから「は」，「が」，「も」を含むもののみを抽出する．英作文に対しては，依存構造の解析結果から“nsubj”によって関連付けした二つの単語を抽出すると同時に，これら

の単語に関連する copula（cop）などの部分も併せて抽出する．これにより，be 動詞や will などの助動詞，否定の表現も併せて抽出される．また，“nsubj”によって関連付けした二つの単語の組み合わせが複数抽出された場合，最初の組み合わせを主語候補と述語候補として選択する．

5.2.2　人手判定による主語および述語の判定方法

A. 主語および述語の判定基準

日本語作文中の主語および述語と英作文中の主語および述語を人手判定する際には，それぞれの定義を以下のように定める．

日本語の主語に関しては，主語とは何か，ということについて様々な学説が存在し，主語の定義には多くの問題が存在する．そこで本研究では学校文法に従い，格助詞の「が」，係助詞「は」（複合助詞「には」を含む）と並立助詞「も」を伴った文節を主語と定義する．ただし，一文節では意味が不明瞭である場合，意味をなす最小限の複数の文節の集合を主語とする．例えば，「周りの人の方が」などが該当する．また，「～することは」や「～するのが」などの動詞を名詞化する表現についても，動詞を含む複数の文節の集合を主語とする．「優先すべきことは」などが該当する．ここで助詞を三種類に限定した理由は，「が」，「は」，「も」は主格の役割をする助詞のうち高頻度で使用されている助詞だからである．留意点として，格助詞「が」の用法には主体と対象があるが，対象として使用される場合（ガ格名詞句）には主語とは認めないこととする．また，主語が省略されている場合，無主語が存在すると考える．よって本章で扱う日本語の主語は「が」，「は」，「も」を伴った一文節または複数の文節，および無主語の四種類となる．また構造的な観点から主語をみると，節が一組ある文，および節が二組以上ありそれらの節が主従の関係で構成された文においては，主語は一つまたは存在しない．節が二組以上あり，それらが対等の資格で並列されて構成された文においては，主語が複数存在する場合もある．

日本語の述語に関しては，文末に置かれる動作や状態を表す用言（動詞，形容詞）＋補助用言＋コピュラ，および名詞＋コピュラを述語とする．形容動詞＋補助用言＋コピュラも述語と考える説もあるが，本研究では時枝（1950）の文法に従って形容動詞を品詞として扱わず，体言＋助動詞（「である」，「だ」，「なり」）とする．このことによりコピュラは全て助動詞となる．また，主語と同様に述語も意味をなす最小限の複数の文節の集合体とする．例えば，「かじっているわけにはいきません」，「悪いものです」などが該当する．

英語の主語および述語に関しては，状態や動作を表す語および心的態度を表す語を述語とし，その述語に対応する動作主（名詞，代名詞，名詞句および名詞節）を主語とする．

B. 主語および述語の判定方法

日本語の主語および述語の判定と英語の主語および述語の判定は，前述の判定基準に沿って行う．

人手判定による主語および述語の抽出にあたっては，学習者の誤用による影響や文構造上の理由により主語や述語が省略されている場合や，複数存在する場合があるため，以下の抽出条件を定めておく．

i) 主語と述語は一組だけ抽出する．
本来，二つ以上の節が対等の資格で並列されて構成された文，いわゆる重文である場合，節の数だけ主語と述語が存在する可能性があるが，ここでは他の文の種類と同様に主語と述語を一組だけ抽出する．

ii) 節が一組あるタイプ I，節が二組以上あるタイプ II のいずれの場合においても，
・主語が省略されている場合，述語のみ選択する．（日本語）
・誤用により主語および述語の候補が複数ある場合，文意を考慮して最も適切と考えられる組み合わせを選択する．（英語）

なお，文の種類は単文，複文，そして重文の三種類に分類する考え方（橋本，1939）が一般的であるが，本研究では，文の種類をタイプ I とタイプ II の二種類に分類して分析を行う（表 5.1）．なお，ここでいう節とは，「主語・述語の関係を含んで文の一部となる連文節」（山口ほか，2001）を指し，文節とは異なる．また，「明日，東京に行く．」のように主語が省略されたものも節として扱う．

表 5.1　本研究で採用した文の種類（構造上の分類）

橋本（1939）による文の種類			本研究での文の種類	
種類	条件		種類	条件
単文	主語・述語の関係が，一回しか成立しない文	→	タイプ I	節が一組ある文
複文	主語・述語の関係が成り立っている文の中に，更に別の主語・述語の関係が認められる部分（従属節）を含む文	→	タイプ II	節が二組以上ある文
重文	二つ以上の節が対等の資格で並列されて構成された文	→		

5.2.3　使用するデータ

今回の分析では，データとして収集した日本語作文・英作文は各 110 エッセイ，各 1499 文のうち，主語が存在する日本語作文 100 文を任意に取り出し，それに対応する英作文 100 文を取り出した．同様に主語が存在しない日本語作文 100 文を任意で取り出し，それに対応する英作文 100 文を取り出し，それぞれ分析に使用した．作文データの一覧は表 5.2 に示した通りである．

表 5.2　分析用作文データ一覧

A	日本人英語学習者の主語がある日本語作文	原言語	100 文
B	日本人英語学習者の主語がない日本語作文	原言語	100 文
C	日本人英語学習者の主語がある日本語作文に対応する英作文	目標言語	100 文
D	日本人英語学習者の主語がない日本語作文に対応する英作文	目標言語	100 文

5.2.4　分析の手順

日本語作文に対する構文解析器の解析結果の正確さを調べるために，まず日本語作文 A および日本語作文 B から人手判定により，それぞれ主語および述語を抽出する．次に日本語構文解析器 CaboCha を利用し，前述の抽出条件に従って選定した主語候補および述語候補を抽出する．人手判定の結果と自動抽出の結果の比較はここで設定した文の種類（I およ

び II）ごとに行い，それぞれどの程度一致しているか確認する．

英作文に対しては解析結果の正確性を調べるために，まず英作文 C および英作文 D から人手判定により主語および述語を抽出する．次に英語構文解析器 The Stanford Parser を利用し，前述の抽出条件に従って選定した主語候補および述語候補を抽出する．人手判定の結果と自動抽出の結果の比較はここで設定した文の種類（I および II）ごとに行い，それぞれどの程度一致しているか確認する．

5.3　結果と考察

5.3.1　日本語作文の主語・述語判定の結果

表 5.3，5.4 は，それぞれ主語がある日本語作文 100 文 A と主語がない日本語作文 100 文 B について，主語候補と述語候補の抽出結果をまとめたものである．

表 5.3，5.4 では構文解析器 CaboCha を使用し，自動抽出によって得た主語候補および述語候補と人手判定による主語および述語を比較して集計した文の数を示している．ここでいう一致とは，自動抽出の候補と人手判定で抽出した主語および述語が全く同じ場合の完全一致と一部分だけ（文字単位ではなく文節単位で判定）が同じ場合の部分一致を指す．そして，一致する部分がない場合は不一致とする．一文に人手判定による主語および述語は一つだけであるが，構文解析器による自動抽出では主語候補に限って複数抽出される場合があるため，主語のある日本語作文は抽出された主語候補が，ない，一つ，そして複数の三分類（0，1，複数），主語のない日本語作文は抽出された主語候補が，ない，と一つ以上の二分類（0，1 以上）としている．なお，傍線（—）は該当する項目がないことを示す．

表 5.3 より，主語がある日本語作文 A では節が一組あるタイプ I と節が二組以上あるタイプ II の比率がおおよそ 5 対 5 であることがわかる．主語候補と述語候補の抽出結果については個別に見ていく．まず主語候補について，構文解析器による自動抽出と人手判定による結果でもっとも厳しい基準である完全一致の割合は，タイプ I, タイプ II それぞれ 63.3%（31 / 49），62.7%（32 / 51），合わせて 64.0%であった．次に述語候補について，同様の基準で自動抽出と人手判定による結果が完全一致した割合をみると，それぞれ 87.8%（43 / 49），84.3（43 / 51），合わせて 86.0%であった．また主語候補と述語候補の抽出結果について併せてみると，タイプ I の完全一致の割合は 59.2%（29 / 49），タイプ II の完全一致の割合は 54.9%（28 / 51）であった．

表 5.3　日本語作文 A における主語候補・述語候補の分類結果と各分類の件数

		主語候補				述語候補				主語候補・述語候補	
文の種類		抽出タイプ		マッチタイプ		抽出タイプ		マッチタイプ		マッチタイプ	
I	49	1	38	完全一致	31	1	49	完全一致	43	完全一致	29
				部分一致	4			部分一致	6		
				不一致	3			不一致	0		
		複数	10	完全一致	10	—	—	—	—		
				部分一致	0			—	—		
				不一致	0			—	—		
		0	1	—	—	0	0	—	—		
II	51	1	37	完全一致	32	1	51	完全一致	43	完全一致	28
				部分一致	3			部分一致	8		
				不一致	2			不一致	0		
		複数	8	完全一致	8	—	—	—	—		
				部分一致	0			—	—		
				不一致	0			—	—		
		0	6	—	—	0	0	—	—		

次に，表 5.4 より，主語がない日本語作文 B では，節が一組あるタイプ I と節が二組以上あるタイプ II の比率がおおよそ 1 対 9 であることがわかる．主語候補と述語候補の抽出結果について個別に見ていく．まず主語候補について，構文解析器による自動抽出と人手判定による結果は，それぞれ 46.2%（6 / 13），79.3%（69 / 87），合わせて 75.0%であった．次に述語候補について，もっとも厳しい基準である完全に一致した割合をみると，61.5%（8 / 13），86.2%（75 / 87），合わせて 83.0%であった．また主語候補と述語候補の抽出結果について併せてみると，タイプ I の完全一致の割合は 46.2%（6 / 13），75.9%（66 / 87）であった．

表 5.4　日本語作文 B における主語候補・述語候補の分類結果と各分類の件数

		主語候補		述語候補				主語候補・述語候補	
文の種類		抽出タイプ		抽出タイプ		マッチタイプ		マッチタイプ	
I	13	1 以上	7	1	13	完全一致	8	完全一致	6
						部分一致	5		
						不一致	0		
		0	6	0	0	—	—		
II	87	1 以上	18	1	87	完全一致	75	完全一致	66
						部分一致	12		
						不一致	0		
		0	69	0	0	—	—		

5.3.2　日本語作文の主語・述語判定の考察

以上の結果から，構文解析器の抽出の正確さは，主語のある日本語作文では文の種類によってほとんど影響を受けないこと，そして主語のない日本語作文では文の種類，特に節が一組みである文において正確さが低下することが明らかとなった．

主語がある日本語作文 A では，自動抽出された述語候補は 86.0%の正確さであったのに対し，主語候補は 54.9%の正確さにとどまった．主な原因は，主語の候補を複数抽出している文が 18 文（A では 10 文，B では 8 文）も存在していることによる．この 18 文は全て，複数の主語候補の中に人手判定と同じ主語が包含されており，構文解析器の主語抽出結果に何らかの条件を付加することで主語の候補を減らし，正確さを高めることができると考えられる．具体的には，主語候補の中に「最近では」，「レストランには」，「(～の）時は」などのいわゆる副詞のように使用されている語が含まれており，それらを抽出しないような条件を付与することが考えられる．また，節が二組以上あり対等の資格で並列されて構成された文，いわゆる重文において，主語候補の中からどのようにして正しい複数の主語を抽出するのか，ということも今後の改善すべき課題と言える．

主語がない日本語作文 B においては，述語候補は 83.0%，主語候補は 75.3%の正確さで自動抽出できることが明らかとなった．改善すべき点としては，誤って主語として候補を抽出

した 25 文中 10 文において，主語と述語の判定基準で示した「格助詞「が」が対象として使用される場合（ガ格名詞句）には主語とは認めないこととする.」という条件に当てはまるものを主語候補として抽出しており，それらを抽出しないように条件を付加することでさらに正確さを高めることが可能と考えられる.

以上のことより，本研究で収集したような日本語作文に対して，構文解析器を直接利用し，基本文構造である主語や述語を抽出することの妥当性は高いと考えられる.

5.3.3　英作文の主語・述語判定の結果

表 5.5，5.6 は，それぞれ主語がある日本語作文に対応する英作文 100 文 C と主語がない日本語作文に対応する英作文 100 文 D について，主語候補と述語候補の抽出結果をまとめたものである.

表 5.5，5.6 は構文解析器 The Stanford Parser を使用し，自動抽出によって得た主語候補および述語候補と人手判定による主語および述語を比較して集計した文の数を示している．日本語の場合と同様に，ここでいう一致とは，自動抽出の候補と人手判定で抽出した主語および述語が全く同じ場合の完全一致と一部分だけ（文字単位ではなく単語単位で判定）が同じ場合の部分一致を指す．また，一致する部分がない場合は不一致とする．英作文の構文解析器による自動抽出では複数の語が抽出されないように基準を定めたため，英作文は抽出された候補が，ないと一つの二分類（0，1）としている．傍線（—）は該当する項目がないことを示す.

表 5.5 より，主語がある日本語作文に対応する英作文 C では節が一組あるタイプ I と節が二組以上あるタイプ II の比率がおおよそ 7 対 3 であることがわかる．主語候補と述語候補の抽出結果について個別に見ていく．まず主語候補について，構文解析器による自動抽出と人手判定による結果でもっとも厳しい基準である完全一致の割合は，それぞれ 81.7%（58 / 71），86.2%（25 / 29），合わせて 83.0%であった．次に述語候補について，同様の基準で自動抽出と人手判定による結果が完全一致した割合をみると，それぞれ 77.5%（55 / 71），79.3%（23 / 29），合わせて 78.0%であった．また主語候補と述語候補の抽出結果について合わせてみると，タイプ I の完全一致の割合は 71.8%（51 / 71），タイプ II の完全一致の割合は 75.9%（22 / 29）であった.

表 5.5　英作文 C における主語候補・述語候補の分類結果と各分類の件数

		主語候補				述語候補				主語候補・述語候補	
文の種類		抽出タイプ		マッチタイプ		抽出タイプ		マッチタイプ		マッチタイプ	
I	71	1	63	完全一致	58	1	65	完全一致	55	完全一致	51
				部分一致	2			部分一致	5		
				不一致	3			不一致	5		
		0	8	—	—	0	6	—	—		
II	29	1	29	完全一致	25	1	29	完全一致	23	完全一致	22
				部分一致	0			部分一致	1		
				不一致	4			不一致	5		
		0	0	—	—	0	0	—	—		

表 5.6　英作文 D における主語候補・述語候補の分類結果と各分類の件数

		主語候補				述語候補				主語候補・述語候補	
文の種類		抽出タイプ		マッチタイプ		抽出タイプ		マッチタイプ		マッチタイプ	
I	40	1	37	完全一致	36	1	40	完全一致	29	完全一致	28
				部分一致	0			部分一致	3		
				不一致	1			不一致	8		
		0	3	—	—	0	0	—	—		
II	60	1	57	完全一致	54	1	60	完全一致	48	完全一致	48
				部分一致	0			部分一致	3		
				不一致	3			不一致	9		
		0	3	—	—	0	0	—	—		

次に，表 5.6 より，主語がない日本語作文に対応する英作文 D では，節が一組あるタイプ I と節が二組以上あるタイプ II の比率がおおよそ 4 対 6 であることがわかる．まず主語候補の抽出結果について個別に見ていく．まず構文解析器による自動抽出と人手判定による結果をもっとも厳しい基準である完全に一致した割合をみると，それぞれ 90.0%（36 / 40），90.0%（54 / 60），合わせて 90.0%であった．次に述語候補について，同様の基準で自動抽出と人手判定による結果が完全一致した割合をみると，それぞれ 72.5%（29 / 40），80.0%（48 / 60），合わせて 77.0%であった．また主語候補と述語候補の抽出結果について合わせてみると，タイプ I の完全一致の割合は 70.0%（28 / 40），タイプ II の完全一致の割合は 80.0%（48 / 60）であった．

5.3.4　英作文の主語・述語判定の考察

以上の結果より，主語がある日本語作文に対応する英作文 C では，自動抽出された主語候補は 83.0%，述語候補は 78.0%の正確さであることが判明した．主語候補と比較して述語候補の正確さが低い原因としては，学習者の述語部分の誤用の多さが考えられる．

誤りの例（原文のまま）：

- Restaurants is came to many people.
- That do not coming guest at smoker.
- I had be given five thousand moneys once a month by high school student.
- My father smoking cigarette.
- There are don't like human in the future.
- My grandfather is die of fast because he is smoker.

また，複数の動詞を同時に使用，また不適切な形態で使用などの理由により構文解析器では正常に述語候補を抽出できていない．このことは，述語を抽出できないことから述語部分に誤用があると推定できるが，誤用の種類まで推定することは不可能であり，この点に関しては今後の課題と言える．

最後に，主語がない日本語作文に対する英作文 D では，自動抽出された主語候補は 90.0%の正確さ，述語候補は 77.0%の正確さであり，主語がある日本語作文に対応する英作文 C と同様，高い正確性を示していることが明らかとなった．

以上のことより，本研究で収集したような日本語作文に対応する誤りを含む英作文に対しても構文解析器を直接利用し，基本文構造である主語や述語をする抽出することの妥当性は高いと考えられる．

5.4　まとめ

本章では，初級英語学習者の産出した日本語と英語の各作文から人手によって抽出した主語および述語と，日本語および英語構文解析器（日本語：CaboCha，英語：The Stanford Parser）を利用して自動抽出した主語および述語を比較し，その正確さについて分析を行った．

学習者の日本語作文に関しては，主語がある日本語作文から人手によって抽出した主語および述語と，構文解析器を利用して自動抽出した主語候補および述語候補を比較したところ，主語は 54.9%，述語は 86.0%の正確さで抽出できること，主語がない日本語作文では主語は 75.3%，述語は 83.0%の正確さで抽出できることが明らかとなった．

また，学習者の英作文に関しては，主語がある日本語作文に対応する英作文から人手によって抽出した主語および述語と，構文解析器を利用して自動抽出した主語候補および述語候補を比較したところ，主語は 83.0%，述語は 78.0%の正確さで抽出できること，主語がない日本語作文では主語は 90.0%，述語は 77.0%の正確さで抽出できることが明らかとなった．

以上のことから，既存の構文解析器を利用して学習者作文から主語および述語を抽出することが可能であること，学習者の誤りを含む文に対しても構文解析器が使用可能なこと，そして，構文解析器の解析結果に条件を付与することでその主語および述語抽出の正確さを高めることが可能であることなどが確認できた．

次章では，誤りの判定基準を原言語および目標言語から自動抽出した主語および述語からなる日英文型の比較により設定することは妥当性があるのか確認する．

第 6 章　日英文型に基づく誤用検出の基準および手順の提案

6.1　はじめに

本章では，原言語情報を利用した新たな英作文の誤用検出手法の確立を目指し，1）初級学習者が習得すべき重要な英語文法項目を検討して，その英語文法項目に対応する日本語と併せて日英文型を規定する．そして，2）日英文型比較による自動誤用判定を行うための基準およびその手順について提案する．すでに述べたように，本研究で対象とする学習者は初級学習者のうち擬似初心者（中学校および高等学校において 6 年間の英語教育を受けたにもかかわらず，第二言語学習者として順調な発達段階を踏んでいない学習者）にあたる学習者である．それゆえに，これら学習者に対してシステムを実践的に使用するために，重要な英語文法項目を中学英語教科書から抽出して文型を規定する．

なお，ここでいう文型とは学校文法における動詞型の分類に基づいた語順のルール（5 文型）を指すものではなく，目標言語である英語と原言語である日本語の文構造上の対応関係を指す．

6.2　日英文型の規定

6.2.1　中学教科書における文法項目の調査

英語学習者が習得すべき重要な文法項目について考える際，まずもって考慮すべきは教科書である．本研究は初級英語学習者を対象としていることから，まず，中学教科書 6 種類（COLUMBUS 21, NEW CROWN, NEW HORIZON, ONE WORLD, SUNSHINE, TOTAL ENGLISH）3 学年分の計 18 冊（各教科書の詳細は第 3 章を参照）で扱われている学習項目について調査し，次に，主語および述語に関わる学習項目からシステムに組み込む英語文法項目を検討する．

まず，中学英語教科書で扱われている学習項目について調査する．表 6.1 は平成 30 年度，もっとも採択率の高かった教科書 NEW HORIZON において，各単元で中心的に扱われている学習項目の一覧である．他の 5 種類の教科書と比較したところ，記載されている学習項目はほぼ同様であるものの，学習指導要領には生徒が学習すべき文法項目についての具体的な指示がないことから，教科書間で扱われている学習項目にわずかな差異が存在する．例えば，COLUMBUS 21 では第二学年で付加疑問文，第三学年で受動態の未来形が学習すべき学習項目として扱われているが，NEW HORIZON ではそれらの取り扱いがない．

表 6.1　NEW HORIZON 学習項目一覧

学年	学習項目
1	人称代名詞（単数，主格），be 動詞（肯定文，疑問文，否定文），指示代名詞，人称代名詞（単数，所有格），一般動詞（肯定文，疑問文，否定文），疑問詞（what），形容詞，可算名詞，疑問詞（how），提案（let's），命令文（肯定），3 単現の S（肯定文，疑問文，否定文），人称代名詞（複数），疑問詞（who），形式主語，疑問詞（which），接続詞（or），疑問詞（where），疑問詞（whose），所有代名詞，人称代名詞（単数，目的格），現在進行形（肯定文，疑問文，否定文），命令文（否定），助動詞（can），助動詞（肯定文，疑問文，否定文），疑問詞（when），過去形（一般動詞；肯定文，疑問文，否定文）
2	過去形（be 動詞；肯定文，疑問文，否定文），過去進行形，自動詞（look），未来形（be going to；肯定文，疑問文，否定文），第 4 文型（show, call），丁寧表現（could you ~?, please）不定詞（副詞的用法，名詞的用法，形容詞的用法），助動詞（could），助動詞（have to），助動詞（will），助動詞（must），助動詞（may），if 節，思考動詞（think, know），接続詞（when），接続詞（because），there is 構文（肯定文，疑問文，否定文），助動詞（should），動名詞，比較級，最上級，原級，助動詞（shall）
3	受動態（現在形，過去形），第 5 文型（make），現在完了（継続，経験，完了；肯定文，疑問文，否定文），形式主語（it to 構文），分詞（現在，過去），関係代名詞（what），関係代名詞（限定用法）

注 : 学習項目は出現順に記載しているが，複数の単元にわたって出現した項目についてはまとめて記載している．3 単現の S とは主語が三人称の単数形，かつ動詞が現在形である場合に動詞に付与される形態素“-s”を指す．

このような状況を踏まえ，本研究では初級学習者にとって重要な文法項目を集約するにあたり，一部または全ての英語教科書に記載がない文法項目であっても，初級学習者が習得すべきだと判断できる項目に関しては文法項目のリストに含めることとする．一方，一部または全ての教科書で記載がある文法項目であっても，重要でないと判断できる項目に関しては文法項目リストに含めないこととする．表 6.2 は初級学習者にとって重要な文法項目を文法的特徴に基づいて分類したものである．

表 6.2　初級学習者にとって重要な文法項目一覧

<table>
<tr><td rowspan="4">広義の文法カテゴリー</td><td rowspan="2">品詞カテゴリー</td><td colspan="2">動詞（be 動詞，一般動詞），助動詞（中核的法補助詞），名詞（加算名詞，不可算名詞），代名詞（人称代名詞，指示代名詞，所有代名詞），形容詞（形容詞，冠詞，数詞），副詞，前置詞，接続詞</td></tr>
<tr><td colspan="2">準動詞（不定詞，動名詞，分詞），関係詞（関係代名詞），疑問詞</td></tr>
<tr><td rowspan="2">狭義の文法カテゴリー</td><td>文法機能カテゴリー</td><td>格（人称），時制（現在，過去，未来），進行形，完了形（現在），**法（可能，意志，義務，蓋然性）**，態（能動態，受動態），否定，命令</td></tr>
<tr><td>その他</td><td>3 人称単数現在形の S，比較，疑問，丁寧表現，形式主語，there is 構文</td></tr>
</table>

注：中核的法補助詞は can, could, may, might, shall, should, will, would, must を指す（Quirk et al., 1985）．厳密には，準動詞，関係詞，疑問詞は品詞ではないが，品詞に準ずるものであることから，ここでは品詞として扱う．太字項目は追加された文法項目を意味する．

ここでは初級学習者にとって重要な文法項目を三つのカテゴリーに分類しているが，文法項目の分類，いわゆる文法カテゴリーについては様々な考え方が存在し，その定義には問題が含まれている．この点について，松村（1971）は以下のように指摘している．

> 一般的には一つの言語内における，個別的でなくかなりの範囲に共通に表われる事柄を指すのに用いられる．しかし，この規定はあまりに漠然としており，かつ，言語の本質的事実に迫ることができないものである．（p.749）

そこで本研究では，品詞カテゴリーを単語の意味的・文法的な分類，文法機能カテゴリーを文の成分が他の成分に対して持つ関係による分類，その他を品詞カテゴリーおよび文法機能カテゴリー以外の文法的な働きのあるものと定めて文法項目の分類を行っている．なお，ここでいう文法カテゴリーとは文法機能カテゴリーとその他を併せた狭義の文法カテゴリーを指すものとする．

次に，中学英語教科書の調査により明らかとなった初級学習者にとって重要な文法項目の中から，本研究で扱う英語文法項目を抽出する．抽出にあたっては以下の 4 条件に基づいて行う．

1. 品詞カテゴリーの項目は，主語または述語になりうる品詞項目であること
2. 文法機能カテゴリーの項目は，主語または述語に文法的な働きをもつ文法項目であること
3. 文法機能カテゴリーの項目は，該当する文法項目が日本語の文法にも存在すること
4. その他のカテゴリーの項目は，global error の原因となりうる文法項目であること

表 6.3 は上述の条件を満たす文法項目をまとめたもので，自動的に検出できる可能性が高い文法項目である．

表 6.3　自動検出できる可能性が高い英語文法項目

	主語に関わる項目	述語に関わる項目
品詞カテゴリー	A.名詞，B.代名詞（1.人称代名詞，2.指示代名詞）	E.動詞（1.be 動詞，2.一般動詞），F.助動詞（法補助詞）
	C.不定詞（名詞的用法），D.動名詞	—
文法カテゴリー	a.人称（主格）	c.極性（1.肯定，2.否定），e.時制（1.現在，2.過去），g.法（1.可能，2.意志，3.義務，5.許可，6.禁止）
	b.形式主語	h.there is 構文

本研究では，主語および述語を基本とした日英文型を規定するために，表 6.3 にある六つの品詞カテゴリー（名詞，代名詞（人称代名詞，指示代名詞），不定詞（名詞的用法），動名詞，動詞（be 動詞，一般動詞），助動詞（法補助詞））と六つの文法カテゴリー（人称（主格），形式主語，極性（肯定，否定），時制（現在，過去），法（可能，意志，義務，許可，禁止），there is 構文）を組み合わせて文型を規定する．なお，本研究ではアカデミックライティングを対象として誤用検出システムを構築するため，アカデミックライティングにおいて基本的に使用されることがない文法カテゴリーの項目（命令，疑問，丁寧表現，話法）については，上述の抽出のための 4 条件に適合・不適合にかかわらず除外している．

6.2.2　英語文法項目に対応する日本語

前節では自動誤用検出システムが対象とする英語文法項目を決定した．本節ではその英語文法項目（6 品詞 6 文法項目）について，まず日本語文型を規定し（表 6.4），次にその日本語文型に対応する英語文型を規定する（表 6.5）．なお，日本語文型を規定するにあたり，特に述語ベースの文型においては，対応する語の全ての活用形の中からもっとも妥当であると考えられる活用形のみ記載している．

表 6.4　日本語文型

No.	品詞カ	文法カ	主語ベースの文型
1	A	—	名詞 +（は / が / も）+ 述語
2	B1	a	（私 / あなた / 彼 / 彼女 / それ / 私たち / あなたたち / 彼ら / 彼女ら / それら）+（は / が / も）+ 述語
3	B2	—	（これ / あれ / それ / これら / あれら / それら)+(は / が / も）+ 述語
4	C，D	b	（すること / こと / するの）+（は / が / も）+ 述語
No.	品詞カ	文法カ	述語ベースの文型
5	E1	c1, e1, h	主語 +（ある / いる）
		c2, e1, h	主語 +（ない / いない）
		c1, e2, h	主語 +（あった / いた）
		c2, e2, h	主語 +（なかった / いなかった）
6	E1	c1, e1	主語 + 名詞 +（です / である / だ）
		c2, e1	主語 + 名詞 +（でない）
		c1, e2	主語 + 名詞 +（でした / であった / だった）
		c2, e2	主語 + 名詞 +（でなかった）
7	E1	c1, e1	主語 + 形容詞 +（です / φ）
		c2, e1	主語 + 形容詞 +（ない / ではない）
		c1, e2	主語 + 形容詞 + た
		c2, e2	主語 + 形容詞 +（なかった / ではなかった）
8	E2	c1, e1	主語 +（思う / 考える）

		c2, e1	主語 +（思わない / 考えない）
		c1, e2	主語 +（思った / 考えた）
		c2, e2	主語 +（思わなかった / 考えなかった）
9	E2	c1, e1	主語 +（知る / わかる）
		c2, e1	主語 +（知らない / わからない）
		c1, e2	主語 +（知った / わかった）
		c2, e2	主語 +（知らなかった / わからなかった）
10	E2	—	主語 + 一般動詞（存在動詞，思考動詞，認識動詞を除く）
11	F	g1, c1, e1	主語 +（名詞 + できる / 動詞 + ことができる / 動詞 + られる）
		g1, c2, e1	主語 +（名詞 + できない / 動詞 + ことができない / 動詞 + られない）
		g1, c1, e2	主語 +（名詞 + できた / 動詞 + ことができた / 動詞 + られた）
		g1, c2, e2	主語 +（名詞 + できなかった / 動詞 + ことができなかった / 動詞 + られなかった）
12	F	g2, c1, e1	主語 + 動詞 + つもり
		g2, c2, e1	主語 + 動詞 + つもりはない
		g2, c1, e2	主語 + 動詞 + つもりだった
		g2, c2, e2	主語 + 動詞 + つもりはなかった
13	F	g3, c1, e1	主語 +（名詞 + しなければならない / 動詞 + なければならない / 名詞 + すべき / 動詞 + べき）
		g6, c2, e1	主語 +（名詞 + してはいけない / 動詞 + てはいけない / 名詞 + すべきではない / 動詞 + べきではない）
		g3, c1, e2	主語 +（名詞 + しなければならなかった / 動詞 + なければならなかった / 名詞 + すべきだった / 動詞 + べきだった）
		g6, c2, e2	主語 +（名詞 + してはいけなかった / 動詞 + はいけなかった / 名詞 + すべきではなかった / 動詞 + べきではなかった）

14	F	g5, c1, e1	主語 ＋（名詞 ＋ してもよい / 動詞 ＋ てもよい）
		g6, c2, e1	主語 ＋（名詞 ＋ してはいけない / 動詞 ＋ てはいけない / 名詞 ＋ すべきではない / 動詞 ＋ べきではない）
		g5, c1, e2	主語 ＋（名詞 ＋ してもよかった / 動詞 ＋ てもよかった）
		g6, c2, e2	主語 ＋（名詞 ＋ してはいけなかった / 動詞 ＋ てはいけなかった / 名詞 ＋ すべきではなかった / 動詞 ＋ べきではなかった）

注：品詞カは品詞カテゴリー，文法カは文法カテゴリーの略語である．

表 6.5　日本語文型に対応する英語文型

No.	品詞カ	文法カ	主語ベースの文型
1	A	—	Noun + Verb
2	B1	a	（I / you / he / she / it / we / they）+ Verb
3	B2	—	（this / that / it / these / those / they）+ Verb
4	C，D	b	（it / to Verb / Verb-ing）+ Verb
No.	品詞カ	文法カ	述語ベースの文型
5	E1	c1, e1, h	Subject +（am / is / are / be / have / has / exist / exists）
		c2, e1, h	Subject +（am not / is not / are not / be not / do not have / does not have / do not exist / does not exist）
		c1, e2, h	Subject +（was / were / had / existed）
		c2, e2, h	Subject +（was not / were not / did not have / did not exist）
6	E1	c1, e1	Subject +（am / is / are / be）
		c2, e1	Subject +（am not / is not / are not / be not）
		c1, e2	Subject +（was / were）
		c2, e2	Subject +（was not / were not）
7	E1	c1, e1	Subject +（am / is / are / be）
		c2, e1	Subject +（am not / is not / are not / be not）
		c1, e2	Subject +（was / were）
		c2, e2	Subject +（was not / were not）

8	E2	c1, e1	Subject +（think / believe / consider / guess / suppose / assume）
		c2, e1	Subject +（do not think / does not think / do not believe / does not believe / do not consider / does not consider /do not guess / does not guess / do not suppose / does not suppose / do not assume / does not assume）
		c1, e2	Subject +（thought / believed / considered / guessed / supposed / assumed）
		c2, e2	Subject +（did not think / did not believe / did not consider / did not guess / did not suppose / did not assume）
9	E2	c1, e1	Subject +（know / get to know / understand / find / notice / realize / recognize）
		c2, e1	Subject +（do not know / does not know / do not get to know / does not get to know / do not understand / does not understand / do not find / does not find / do not notice / does not notice / do not realize / does not realize / do not recognize / does not recognize）
		c1, e2	Subject +（know / get to know / understand / find / notice / realize / recognize）
		c2, e2	Subject +（did not know / did not get to know / did not understand / did not find / did not notice / did not realize / did not recognize）
10	E2	—	Subject + Verb（except for exist Verb, thinking Verb, cognitive Verb）
11	F	g1, c1, e1	Subject +（can Verb / be able to Verb / am able to Verb / is able to Verb / are able to Verb）
		g1, c2, e1	Subject +（cannot Verb / can not Verb / not be able to Verb / am not able to Verb / is not able to Verb / are not able to Verb）
		g1, c1, e2	Subject +（could Verb / was able to Verb / were able to Verb）
		g1, c2, e2	Subject +（could not Verb / was not able to Verb / were not able to Verb）
12	F	g2, c1, e1	Subject +（will Verb / be going to Verb / am going to Verb / is going to Verb / are going to Verb / plan to Verb / plans to Verb）

		g2, c2, e1	Subject +（will not Verb / am not going to Verb / is not going to Verb / are not going to Verb / do not plan to Verb / does not plan to Verb）
		g2, c1, e2	Subject +（would Verb / was going to Verb / were going to Verb / planned to Verb）
		g2, c2, e2	Subject +（would not Verb / was not going to Verb / were not going to Verb / did not plan to Verb）
13	F	g3, c1, e1	Subject +（ought to Verb / must Verb / should Verb / have to Verb / has to Verb）
		g6, c2, e1	Subject +（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb / is forbidden / are forbidden）
		g3, c1, e2	Subject +（ought to have P.P. / had to Verb / should have P.P.）
		g6, c2, e2	Subject +（ought not to have P.P. / should not have P.P. / was not allowed to Verb / were not allowed to Verb / was forbidden / were forbidden）
14	F	g5, c1, e1	Subject +（can Verb / may Verb / might Verb）
		g6, c2, e1	Subject +（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb / is forbidden / are forbidden）
		g5, c1, e2	Subject +（might have P.P. / could Verb）
		g6, c2, e2	Subject +（ought not to have P.P. / should not have P.P. / was not allowed to Verb / were not allowed to Verb / was forbidden / were forbidden）

注：P.P.は動詞の過去分詞形を指す.

誤用を自動で検出するために，6 品詞 6 文法項目を組み合わせて，日本語文型およびそれに対応する英語文型をそれぞれ 14 規定した．次節ではこれらの日英文型を用いてどのように誤りを判定するのか，その判定基準について提案を行う.

6.3　誤用判定基準の構築

本節では自動で誤用を検出するためのルール，いわゆる誤用判定基準の構築を行う．誤用

判定基準は，日本語文型と英語文型のそれぞれの主語および述語の比較により，独自に構築を行う．表 6.6 は文型とその誤用判定基準をまとめたものである．

表 6.6　文型とその誤用判定基準

No.	主語ベースの文型
1	主語が名詞でなければ誤用
2	主語が（I / you / he / she / it / we / they）のいずれかでなければ誤用
3	主語が（this / that / it / these / those / they）のいずれかでなければ誤用
4	主語が（it / to Verb / Verb-ing）のいずれかでなければ誤用
No.	述語ベースの文型
5	述語が（am / is / are / be / have / has / exist / exists）のいずれかでなければ誤用
	述語が（am not / is not / are not / be not / do not have / does not have / do not exist / does not exist）のいずれかでなければ誤用
	述語が（was / were / had / existed）のいずれかでなければ誤用
	述語が（was not / were not / did not have / did not exist）のいずれかでなければ誤用
6	述語が（am / is / are / be）のいずれかでなければ誤用
	述語が（am not / is not / are not / be not）のいずれかでなければ誤用
	述語が（was / were）のいずれかでなければ誤用
	述語が（was not / were not）のいずれかでなければ誤用
7	述語が（am / is / are / be）のいずれかでなければ誤用
	述語が（am not / is not / are not / be not）のいずれかでなければ誤用
	述語が（was / were）のいずれかでなければ誤用
	述語が（was not / were not）のいずれかでなければ誤用
8	述語が（think / believe / consider / guess / suppose / assume）のいずれかでなければ誤用
	述語が（do not think / does not think / do not believe / does not believe / do not consider / does not consider /do not guess / does not guess / do not suppose / does not suppose / do not assume / does not assume）のいずれかでなければ誤用
	述語が（thought / believed / considered / guessed / supposed / assumed）のいずれかでなければ誤用

	述語が（did not think / did not believe / did not consider / did not guess / did not suppose / did not assume）のいずれかでなければ誤用
9	述語が（know / get to know / understand / find / notice / realize / recognize）のいずれかでなければ誤用
	述語が（do not know / does not know / do not get to know / does not get to know / do not understand / does not understand / do not find / does not find / do not notice / does not notice / do not realize / does not realize / do not recognize / does not recognize）のいずれかでなければ誤用
	述語が（know / get to know / understand / find / notice / realize / recognize）のいずれかでなければ誤用
	述語が（did not know / did not get to know / did not understand / did not find / did not notice / did not realize / did not recognize）のいずれかでなければ誤用
10	述語が存在動詞，思考動詞，認識動詞以外の一般動詞でなければ誤用
11	述語が（can Verb / be able to Verb / am able to Verb / is able to Verb / are able to Verb）のいずれかでなければ誤用
	述語が（cannot Verb / can not Verb / not be able to Verb / am not able to Verb / is not able to Verb / are not able to Verb）のいずれかでなければ誤用
	述語が（could Verb / was able to Verb / were able to Verb）のいずれかでなければ誤用
	述語が（could not Verb / was not able to Verb / were not able to Verb）のいずれかでなければ誤用
12	述語が（will Verb / be going to Verb / am going to Verb / is going to Verb / are going to Verb / plan to Verb / plans to Verb）のいずれかでなければ誤用
	述語が（will not Verb / am not going to Verb / is not going to Verb / are not going to Verb / do not plan to Verb / does not plan to Verb）のいずれかでなければ誤用
	述語が（would Verb / was going to Verb / were going to Verb / planned to Verb）のいずれかでなければ誤用
	述語が（would not Verb / was not going to Verb / were not going to Verb / did not plan to Verb）のいずれかでなければ誤用
13	述語が（ought to Verb / must Verb / should Verb / have to Verb / has to Verb）のいずれかで

	なければ誤用
	述語が（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb / is forbidden / are forbidden）のいずれかでなければ誤用
	述語が（ought to have P.P. / had to Verb / should have P.P.）のいずれかでなければ誤用
	述語が（ought not to have P.P. / should not have P.P. / was not allowed to Verb / were not allowed to Verb / was forbidden / were forbidden）のいずれかでなければ誤用
14	述語が（can Verb / may Verb / might Verb）のいずれかでなければ誤用
	述語が（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb / is forbidden / are forbidden）のいずれかでなければ誤用
	述語が（might have P.P. / could Verb）のいずれかでなければ誤用
	述語が（ought not to have P.P. / should not have P.P. / was not allowed to Verb / were not allowed to Verb / was forbidden / were forbidden）のいずれかでなければ誤用

日本語文型と英語文型をどのように比較させれば誤用と判定できるのか，1 から 14 までの各誤用判定基準を以下に示す．

主語ベースの誤用判定基準

文型 1：日本語文が「名詞 +（は / が / も）+ 述語」という文型である場合，英語文の主語も名詞でなければ誤用であると判断する．ただし，正確な誤用判定には，品詞の一致だけではなく意味も一致しているか検討する必要があるため，検出の精度を高めるためには追加基準が必要な文型である．

文型 2：日本語文が「（私 / あなた / 彼 / 彼女 / それ / 私たち / あなたたち / 彼ら / 彼女ら / それら）+（は / が / も）+ 述語」という文型である場合，英語の主語が人称代名詞の主格（I / you / he / she / it / we / they）のいずれかでなければ誤用であると判断する．

文型 3：日本語文が「（これ / あれ / それ / これら / あれら / それら）+（は / が / も）+ 述語」という文型である場合，英語の主語が指示代名詞（this / that / it / these / those / they）のいずれかでなければ誤用であると判断する．

文型 4：日本語文が「（すること / こと / するの）+（は / が / も）+ 述語」という文型である場合，英語の主語が形式主語（it），不定詞の名詞的用法（to Verb），動名詞（Verb-

ing）のいずれかでなければ誤用であると判断する．

述語ベースの誤用判定基準

文型 5：日本語文が「主語 +（ある / いる）」という文型である場合，英語の述語が存在を意味する be 動詞（am / is / are / be），または，一般動詞（have / has / exist / exists）のいずれかでなければ誤用であると判断する．

また，日本語文が否定文や過去文である場合，その組み合わせにより，さらに 3 パターン（現在形・否定形，過去形・肯定形，過去形・否定形）の誤用判定基準が必要となる．英語の述語は，それぞれ（am not / is not / are not / be not / do not have / does not have / do not exist / does not exist），（was / were / had / existed），（was not / were not / did not have / did not exist）のいずれかでなければ誤用であると判断する．

文型 6 および文型 7：日本語文が「主語 + 名詞 +（です / である / だ）」，および「主語 + 形容詞 +（です / φ）」という文型である場合，英語の述語が状態を意味する be 動詞（am / is / are / be）のいずれかでなければ誤用であると判断する．現在形・否定形，過去形・肯定形，過去形・否定形の 3 パターンにおいては，英語の述語は，それぞれ（am not / is not / are not / be not），（was / were），（was not / were not）のいずれかでなければ誤用であると判断する．

文型 8：日本語文が「主語 +（思う / 考える）」という文型である場合，英語の述語が思考を意味する一般動詞（think / believe / consider / guess / suppose / assume）のいずれかでなければ誤用であると判断する．現在形・否定形，過去形・肯定形，過去形・否定形の 3 パターンにおいては，英語の述語は，それぞれ（do not think / does not think / do not believe / does not believe / do not consider / does not consider /do not guess / does not guess / do not suppose / does not suppose / do not assume / does not assume），（thought / believed / considered / guessed / supposed / assumed），（did not think / did not believe / did not consider / did not guess / did not suppose / did not assume）のいずれかでなければ誤用であると判断する．

文型 9：日本語文が「主語 +（知る / わかる）」という文型である場合，英語の述語が認識を意味する一般動詞（know / get to know / understand / find / notice / realize / recognize）のいずれかでなければ誤用であると判断する．現在形・否定形，過去形・肯定形，過去形・否定形の 3 パターンにおいては，英語の述語は，それぞれ（do not know / does not know / do not get to know / does not get to know / do not understand / does not

understand / do not find / does not find / do not notice / does not notice / do not realize / does not realize / do not recognize / does not recognize)，（know / get to know / understand / find / notice / realize / recognize)，（did not know / did not get to know / did not understand / did not find / did not notice / did not realize / did not recognize）のいずれかでなければ誤用であると判断する．

文型 10：日本語文が「主語 ＋ 一般動詞（存在動詞，思考動詞，認識動詞を除く)」という文型である場合，英語の述語も動詞でなければ誤用であると判断する．ただし，正確な誤用判定には，品詞の一致だけではなく意味も一致しているか検討する必要があるため，検出の精度を高めるためには追加基準が必要な文型である．

文型 11：日本語文が「主語 ＋（名詞 ＋ できる / 動詞 ＋ ことができる / 動詞 ＋ られる)」という文型である場合，英語の述語が可能を意味する助動詞（can Verb)，または（be able to Verb / am able to Verb / is able to Verb / are able to Verb）のいずれかでなければ誤用であると判断する．現在形・否定形，過去形・肯定形，過去形・否定形の 3 パターンにおいては，英語の述語は，それぞれ（cannot Verb / can not Verb / not be able to Verb / am not able to Verb / is not able to Verb / are not able to Verb)，（could Verb / was able to Verb / were able to Verb)，（could not Verb / was not able to Verb / were not able to Verb）のいずれかでなければ誤用であると判断する．

文型 12：日本語文が「主語 ＋ 動詞 ＋ つもり」という文型である場合，英語の述語が意志を意味する助動詞（will Verb)，または（be going to Verb / am going to Verb / is going to Verb / are going to Verb / plan to Verb / plans to Verb）のいずれかでなければ誤用であると判断する．現在形・否定形，過去形・肯定形，過去形・否定形の 3 パターンにおいては，英語の述語は，それぞれ（will not Verb / am not going to Verb / is not going to Verb / are not going to Verb / do not plan to Verb / does not plan to Verb)，（would Verb / was going to Verb / were going to Verb / planned to Verb)，（would not Verb / was not going to Verb / were not going to Verb / did not plan to Verb）のいずれかでなければ誤用であると判断する．

文型 13：日本語文が「主語 ＋（名詞 ＋ しなければならない / 動詞 ＋ なければならない / 名詞 ＋ すべき / 動詞 ＋ べき)」という文型である場合，英語の述語が義務を意味する助動詞（ought to Verb / must Verb / should Verb)，または（have to Verb / has to Verb）のいずれかでなければ誤用であると判断する．

また，この日本語文が否定文である場合，「主語 +（名詞 + してはいけない / 動詞 + てはいけない / 名詞 + すべきではない / 動詞 + べきではない）」となり，英語の述語が禁止を意味する助動詞（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb），または（is forbidden / are forbidden）のいずれかでなければ誤用であると判断する．

また，上記の二つの文型における過去形のパターンにおいては，英語の述語は，それぞれ（ought to have P.P. / had to Verb / should have P.P.），または（ought not to have P.P. / should not have P.P. / was not allowed to Verb / were not allowed to Verb / was forbidden / were forbidden）のいずれかでなければ誤用であると判断する．

文型 14 : 日本語文が「主語 +（名詞 + してもよい / 動詞 + てもよい）」という文型である場合，英語の述語が許可を意味する助動詞（can Verb / may Verb / might Verb）のいずれかでなければ誤用であると判断する．

また，文型 13 と同様にこの日本語文が否定文である場合，「主語 +（名詞 + してはいけない / 動詞 + てはいけない / 名詞 + すべきではない / 動詞 + べきではない）」となり，英語の述語が禁止を意味する助動詞（can not Verb / cannot Verb / may not Verb / must not Verb / ought not to Verb / should not Verb），または（is forbidden / are forbidden）のいずれかでなければ誤用であると判断する．

以上，14 の日英文型を元にした誤用判定基準を自動で英作文の誤用を検出するためのルールとして利用することを提案する．

6.4 誤用検出手順の提案

6.2 では，初級学習者が習得すべき重要な英語文法項目を決定し，その英語文法項目を含む日英文型を規定した．また，6.3 では日英文型の比較による自動誤用判定を行うための基準ついて提案を行った．本節ではこの誤用判定基準を組み込んだシステムを構築するために，日英文型比較による自動誤用判定を行うための手順を提案する．

本システムでは，英作文にある構造上の誤用を容易に自動検出するために，日本語作文とそれに対応する英作文のそれぞれから，構文解析器を利用して抽出した主語と述語に基づいて日英文型を規定し，文型の比較によって誤用の判定を行う．これら一連の手順を以下に示す（図 6.1）．

1. 日本語作文とそれに対応する英作文をそれぞれ日本語構文解析器 CaboCha, および英語構文解析器 The Stanford Parser にかける.

⇩

2. 構文解析結果（依存構造情報）を得る.

⇩

3. 解析結果から, 独自ルールにしたがって主語と述語の組み合わせをそれぞれ自動で抽出する.

⇩

4. 日本語文型に基づき，主語と述語の組み合わせに自動で分類する.

⇩

5. 誤用判定基準に従って，すでに定めた文型と抽出した文型を比較する.

⇩

6. ERROR, POSSIBLE, UNKNOWN のいずれかの誤用判定結果を得る.

注 : ERROR は完全な誤り，POSSIBLE は誤りではないが正しいとは限らない．UNKNOWN は判断が不可能なことを指す.

図 6.1　自動誤用検出の手順

6.5　まとめ

本章では，原言語を利用した新たな英作文の誤用検出手法の確立のため，まず，初級学習者にとって重要な文法項目を文法的特徴に基づいて分類したうえ，本研究で扱う英語文法項目を決定した. 併せて, 英語文法項目（6 品詞 6 文法項目）について日本語文型を規定し，それに対応する英語文型を 14 種類規定した．そして，日英文型比較による自動誤用判定を行うための誤用判定基準および自動誤用判定の手順について提案を行った. 次章では，日英文型比較による誤用判定基準を組み込んだシステムの構築およびその評価を行う.

第 7 章　誤用検出システムの構築および評価

7.1　はじめに

前章では，誤用検出システムの中核をなす誤用判定の基準，および手順について提案を行った．本章では，誤用検出に必要な基準および手順に従ってシステムを構築し，その有効性を示すために検証を行う．

7.2　システムの概要

本システムの目的は教師の英作文添削支援であるため，教師が実際に教育現場で使用することを想定している．それゆえに，教師がすでに学習者の作文データ（日本語作文とそれに対応する英作文）の有し，それらの作文データをシステムに入力できる段階にあることを前提としている．

本システムでは，教師が英語学習者の日本語作文およびそれに対応する英作文を入力すると，規定した誤用判定基準にしたがって英作文に含まれる誤りを検出し，その検出結果および集計結果を得ることができる仕様となっている．また，各英作文の詳細情報も表示されるように設計されている．

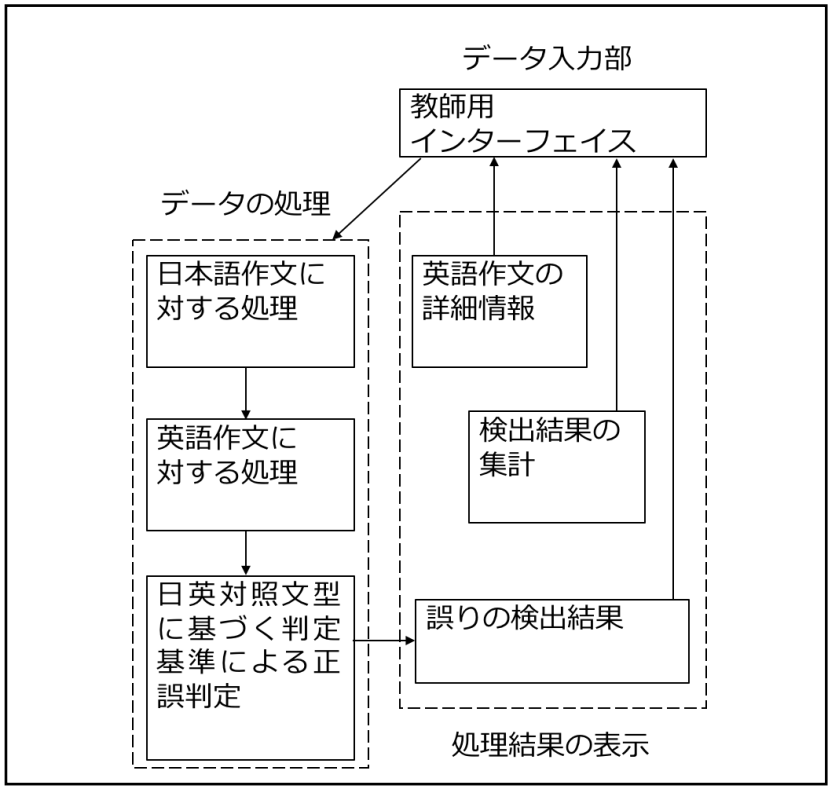

図 7.1　システムの構成

システムの構成は，データを入力する部分，そのデータを処理する部分，そして，処理した結果を表示する部分からなっている（図 7.1）．データを入力する部分，および，処理した結果を表示する部分は PHP で構築されており，web ブラウザ用にデザインされている．また，データを処理する部分については Python 3 で構築されている．

7.2.1　インターフェイスの概要

本システムのインターフェイスは，データを入力する部分，および，処理した結果を表示する部分である．

まず，データを入力する部分について説明を行う．図 7.2，7.3 は作文データ入力画面である．既に述べたとおり，本システムでは教師が学習者の作文データ（日本語作文とそれに対応する英作文）を直接入力する必要がある．図 7.2，7.3 のように，作文データ入力画面では左側部分に日本語作文，右側に日本語に対応する英作文を直接入力するように設計されている．本システムにおける誤用判定は，日本語作文と英作文のそれぞれから抽出した主語および述語の比較によって行われるため，対応関係にある作文のどちらか一方でも欠けてしまうと正常に誤用判定を行うことができない．そこで，本システムでは教師の作文データ入力ミスを防ぐために日本語作文および英作文の入力文数を表示し（図 7.3），さらに日本語作文および英作文が同数でなければデータの処理に進むことができないように設定している．

図 7.2　作文データ入力画面 1

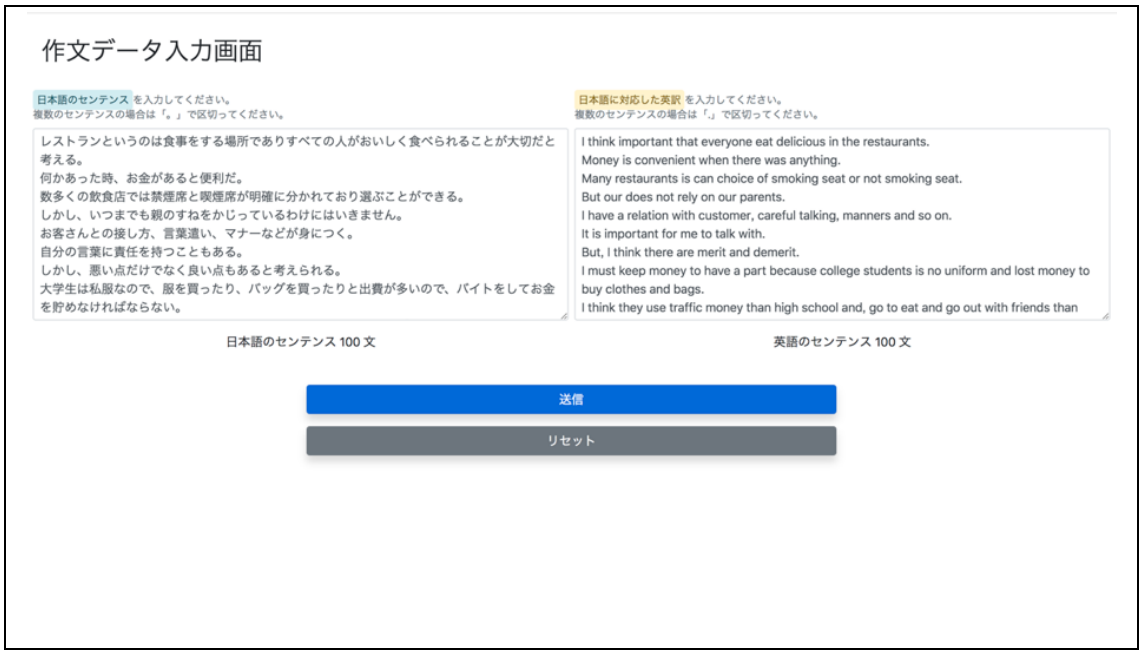

図 7.3　作文データ入力画面 2

次に，処理した結果を表示する部分について説明を行う．結果を表示する部分は，すべての英作文の誤り検出結果を表示する画面，個別の英作文の詳細情報を表示する画面，そして誤りの検出結果を集計した画面からなっている．

図 7.4，7.5，7.6 は英作文の誤り検出結果画面である．

英作文エラーの検出結果

青文字は主語部分のエラー、赤文字は述語部分のエラーを示しています。
各文の詳細情報は英文をクリックしてください。

	日本語	英語
1	レストランというのは食事をする場所でありすべての人がおいしく食べられることが大切だと考える。	I think important that everyone eat delicious in the restaurants.
2	何かあった時、お金があると便利だ。	Money is convenient when there was anything.
3	数多くの飲食店では禁煙席と喫煙席が明確に分かれており選ぶことができる。	Many restaurants is can choice of smoking seat or not smoking seat.
4	しかし、いつまでも親のすねをかじっているわけにはいきません。	But our does not rely on our parents.
5	お客さんとの接し方、言葉遣い、マナーなどが身につく。	I have a relation with customer, careful talking, manners and so on.
6	自分の言葉に責任を持つこともある。	It is important for me to talk with.
7	しかし、悪い点だけでなく良い点もあると考えられる。	But, I think there are merit and demerit.
8	大学生は私服なので、服を買ったり、バッグを買ったりと出費が多いので、バイトをしてお金を貯めなければならない。	I must keep money to have a part because college students is no uniform and lost money to buy clothes and bags.
9	大学生は交通費が高校の時より多くかかったり、友達とご飯や遊びに行くことも多くなると思います。	I think they use traffic money than high school and, go to eat and go out with friends than then.
10	しかし、少しずつ慣れていき、お客様から「ありがとう」と言われたとき、とてもうれしくなりました。	But as soon as adjust, when they say "thank you, " I am very happy.

図 7.4　英作文の誤り検出結果画面 1

31	日本語	英語
32	大学生は、社会に出るため、独立するための猶予期間と言える。	College student is moratorium.
33	このように、アルバイトをすることで、自分を成長させることができるのです。	I grow big person in have a part time job.
34	このような人は少なくないと思うのです。	I think such people.
35	だが、アルバイトは、仕事場の雰囲気を直接感じとることができる。	A part time job can human relations in job.
36	このことを気づき、知るというとても良い機会だと思います。	This is good study.
37	また、大学に通うには、多くの学費を払わなくてはなりません。	It is use many money to go to college.
38	そうしたら、みんな気持ちよく食事をとれると思います。	I think all people can eat well.
39	タバコを吸いたい人のために喫煙ルームを設けるのは一つの手段だと思いました。	I think that it is one means to make the smoking room for the person who wants to smoke.
40	何の苦労も知らず、買ったり食べたり、遊んでいると思います。	I think they don't know tried, buy, eat and game.
41	レストランに入っても、必ず禁煙のスペースを選びます。	I enter the restaurant, I select not smoke space.
42	禁煙者にとって、ドアで区切られるのは、閉鎖的で落ち着いて食事を楽しめなくなるのでは、と感じる。	Smoking people think not enjoy and not relax eat, because door half place is close.
43	多くのレストランで分煙という看板をよく目にします。	I see a sign about separation of smokers and non smokers.
44	その為アルバイトをしてお金をかせがなくてはならない。	I work part time and I save money.
45	必ず少しは出ているだろう。	It run a little certainly.

図 7.5　英作文の誤り検出結果画面 2

	日本語	英語
89	及ぼすということもあるので、食べ物を食べる場でタバコを吸うのはいけないと思います。	a body, but also it may be even the person who doesn't smoke gives harm, because of the smoke.
90	みんなが楽しく快適に来客、食事ができるレストランが魅力的だと思います。	All people is enjoy and comfortable come eat restaurants beautiful.
91	そこで、いろいろなことを知り、学べると思う。	In having a part time job, we know and learn a lot of thing.
92	そのため洋服を買うのにお金が必要になる。	I need money to buy clothes.
93	ですから全ての国のレストランを禁煙にしてしまうのは無理があると思います。	So I think to be not able to stop smoking at the restaurant of all countries.
94	その中でも接客業のアルバイトをするべきだと思います。	I think that they should have a service trade.
95	それに妊婦さんや小さい子どもに対してもタバコは害のある悪いものだと思います。	And I think that cigarette hurt pregnant and children.
96	学費を払ったり、好きなことをするにはアルバイトでお金をかせぐしかありません。	A part time job can pay college money and play to like things.
97	どんなことでも、しっかり、経験しておくことが大切だと思います。	I think that is important to experience anything.
98	社会性が培われたり、お金の大切さを学ぶことができます。	They can get power of society and they can study treasure of money.
99	それに、お金をかけてまで自分の体を傷つける必要はないと思います。	And, I think not need a wound to my health for money.
100	しかし、分煙にはするべきだと思います。	But I think separation of smoking areas.

集計結果

図 7.6　英作文の誤り検出結果画面 3

英作文の誤り検出結果画面では，入力した全文について対応関係にある日本語作文および英作文を一行で表示している（図 7.4）．英作文は本システムにより主語部分に誤りがあると判定された場合には主語部分が赤字で表示され，述語部分に誤りがあると判定された場合には述語部分が青字で表示されるように設計されている．ただし，黒字で表示された英作

文であっても正用文とは限らない．それは該当する文型が存在せず，判定できない文も含まれているためである．

また，本システムは誤りの検出結果を表示するとともに，英作文の詳細情報も確認できる機能を備えている．詳細情報を確認したい英作文にカーソルを当てると，その英作文があるセルが赤色に変わりクリックすることができる（図 7.5）．そこでクリックすると画面が英作文の詳細情報画面に切り替わる（図 7.7）．英作文の詳細情報画面で表示される情報は，システムが独自ルールに基づいて判断した 1）日本語作文の主語および述語，2）英作文の主語および述語，3）英作文の主語および述語の文型，そして 4）英作文の主語および述語の誤用判定結果の 8 つである．これらのデータの処理に関しては 7.2.2 で詳しく説明する．

図 7.7　英作文の詳細情報画面 1

補足機能として，文法タイプ，いわゆる英作文の主語および述語の文型は使用しているコードがそのまま表示されるため，そのコードにカーソルをあてることで文型情報（品詞カテゴリーおよび文法カテゴリー）が表示されるようにしている（図 7.8）．

英作文の誤り検出結果画面の末尾には，誤りの集計結果画面に切り替わるボタンが用意されており（図 7.6），そこをクリックすることで誤りの検出結果集計画面に切り替わる．図 7.9 は誤りの検出結果集計画面である．

詳細情報

39	日本語	タバコを吸いたい人のために喫煙ルームを設けるのは一つの手段だと思いました。
	英語	I think that it is one means to make the smoking room for the person who wants to smoke.

文法タイプの詳細はカーソルを当ててください。

		日本語	英語	文法タイプ	判定結果
39	主語		I	unknown	unknown
	述語	思いました	think	F1_aff_past	Error

品詞カテゴリー：一般動詞（思考）
文法カテゴリー：肯定・過去

図 7.8　英作文の詳細情報画面 2

集計結果

		判別できた文型の数	ERRORの数
述語ベース	A	2	1
	B	1	1
	C	42	0
	F1	46	18
	F2		
	G	1	1
	H	5	1
合計/総数		97/100	22/100
主語ベース	D	7	0
	E		
	I		
	J		
合計/総数		7/100	0/100

図 7.9　誤りの集計結果画面

誤りの検出結果集計画面では，入力した英作文のうち規定した文型に当てはまる英作文の数，および誤りと判定された英作文の数を集計して表示している．

7.2.2　処理の流れ

図 7.10 は英作文の誤用判定の流れをチャートで示したものである．

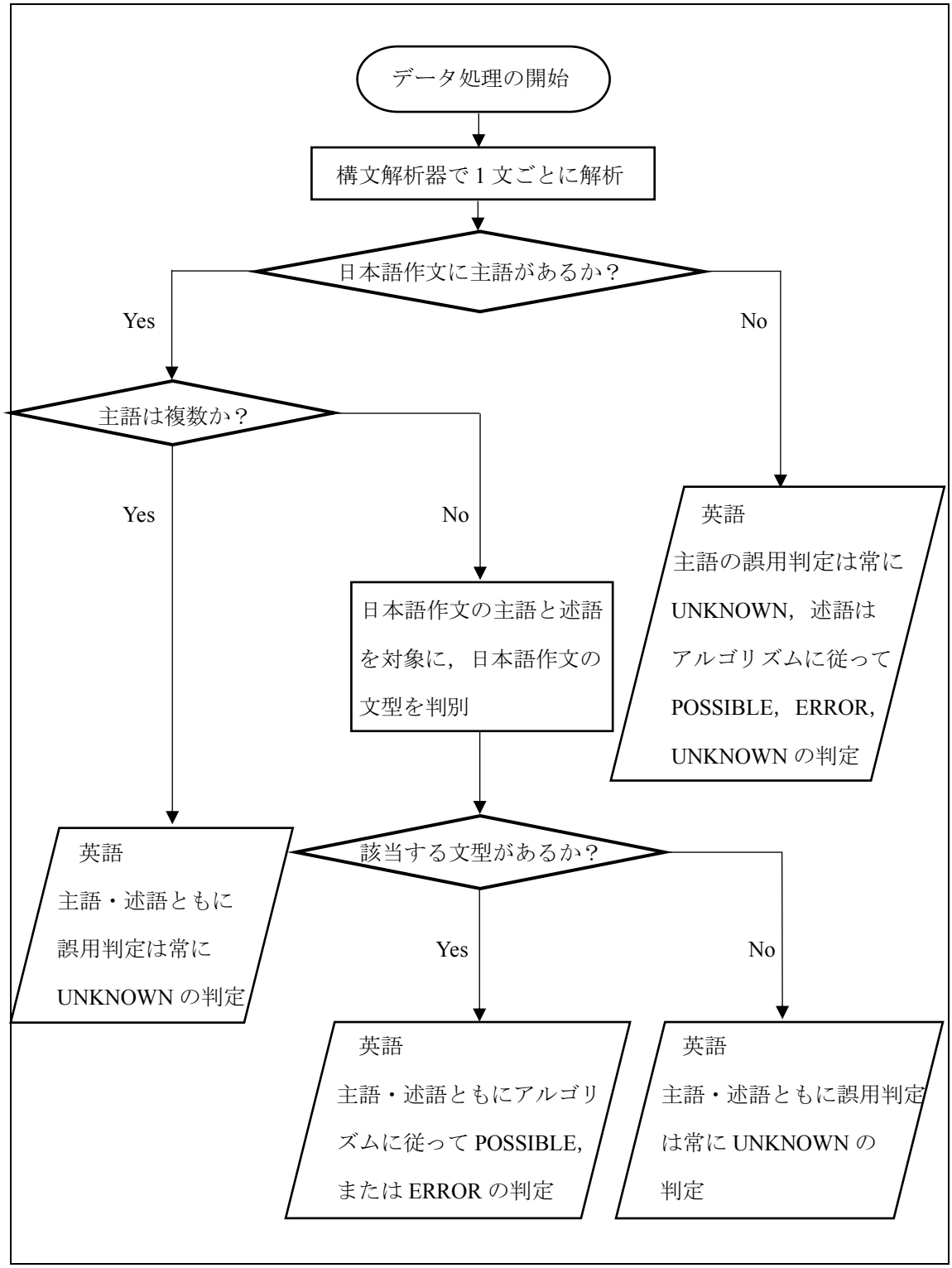

図 7.10　データ処理の流れ

データを処理するアルゴリズムは以下の例文を用いて説明する.

日本語文：「私は大学生でした。」

英作文：“I am a college student.”

Step 1 日本語作文および英作文をそれぞれ CaboCha および The Stanford Parser で構文解析し，依存構造情報を得る.

日本語作文「私は大学生でした。」を構文解析して得られる情報は以下の通りである.


```
<sentence>
<chunk id="0" link="1" rel="D" score="0.000000" head="0" func="1">
<tok id="0" feature="名詞,代名詞,一般,*,*,*,私,ワタシ,ワタシ">私</tok>
<tok id="1" feature="助詞,係助詞,*,*,*,*,は,ハ,ワ">は</tok>
</chunk>
<chunk id="1" link="-1" rel="D" score="0.000000" head="2" func="4">
<tok id="2" feature="名詞,一般,*,*,*,*,大学生,ダイガクセイ,ダイガクセイ">大学生</tok>
<tok id="3" feature="助動詞,*,*,*,特殊・デス,連用形,です,デシ,デシ">でし</tok>
<tok id="4" feature="助動詞,*,*,*,特殊・タ,基本形,た,タ,タ">た</tok>
</chunk>
</sentence>
```


英作文“I am a college student.”を構文解析して得られる情報は以下の通りである.

```
(('student', 'NN'), 'nsubj', ('I', 'PRP'))
(('student', 'NN'), 'cop', ('am', 'VBP'))
(('student', 'NN'), 'det', ('a', 'DT'))
(('student', 'NN'), 'compound', ('college', 'NN'))
```

Step 2 それぞれの依存構造情報から独自のルールによって主語および述語を抽出する.

日本語作文

- 最終文節（chunk id がもっとも大きい数字）を述語として抽出
  ここでは chunk id = 1 の「学生でした」述語として抽出
- 主語候補として，抽出された述語と依存関係にある文節（述語の chunk id と同じ数の chunk link を持つ文節）をすべて抽出し，係助詞「は」，格助詞「が」，「も」を持つ文

節を主語として抽出

ここでは link=1 の文節「私は」を主語として抽出

英作文

・nsubj（名詞句主語）タグが付与されている文節を主語として抽出

ここでは“I”を主語として抽出

・主語から依存関係にある文節までの中にある語のうち，動詞，助動詞，コピュラを述語として抽出（コピュラが述語の候補になった場合，その直後にくる文節を述語に含める.）

ここでは“I”から“student”の間にある“am a college student”を述語として抽出

Step 3 あらかじめ規定した文型情報と誤用判定基準（第 6 章参照）および構文解析器による依存構造情報と独自ルールにより抽出した主語・述語情報を合わせて，誤用判定結果（POSSIBLE，ERROR，UNKNOWN のいずれか）を含む 14 の情報を得る.

日本語作文から 1. 文章番号，2. 文型（述語ベース），3. 文型（主語ベース），4. 検出主語候補の数，5. 述語，6. 述語の品詞，7. 主語，英作文から 8. 主語の誤用判定結果，9. 述語の誤用判定結果，10. 主語，11. 主語の品詞，12. 述語，13. 述語の品詞，14. 主語から述語までの文字列

Step 4 誤用判定は抽出された日本語の主語の数（1，複数，0）に応じて異なる処理を行い，異なる分析結果をえる.

・日本語の主語が一つ検出された場合

1 から 14 までの情報が全て得られ，8. 主語の誤用判定結果と 9. 述語の誤用判定結果も POSSIBLE，ERROR，UNKNOWN のいずれかになる.

・日本語の主語が複数検出された場合

2. 文型（述語ベース）と 3. 文型（主語ベース）を判定せず，8. 主語の誤用判定結果と 9. 述語の誤用判定結果の両方を UNKNOWN とする.

・日本語の主語が検出されなかった場合

3. 文型（主語ベース）を判定せず，8. 主語の誤用判定結果を UNKNOWN とし，9. 述語の誤用判定結果は POSSIBLE，ERROR，UNKNOWN のいずれかになる.

ここでは，1. 1，2. B_aff_past，3. E，4. 1，5. 大学生でした，6. 名詞-一般-*，7. 私は，

8. POSSIBLE，9. ERROR，10. I，11. PRP，12. student，13. NN，14. I am a college student という分析結果を得ることができる．

7.3　検証

英作文誤用自動検システムの有効性を検証する目的で，学習者の日本語作文 1001 文と英作文 1001 文を入力して誤用検出を行った．誤用検出の検証に先立ち，前処理として英作文 1499 文から日本語作文と 1 対 1 の対応をしていない文，直接話法を含む文，そして命令文は除外している．

表 7.1　検証に使用した誤用判定基準

主語ベースの文型		
文型 1	人称代名詞 （1 人称：主格）	JPN：私 +（は / が / も）+ 述語 ENG：(I)
述語ベースの文型		
文型 2	Be 動詞 （存在）	JPN：主語 +（ある / いる） ENG：(am / is / are / be / have / has / exist / exists)
文型 3	Be 動詞 （状態：名詞型）	JPN：主語 + 名詞 +（です / である / だ） ENG：(am / is / are / be)
文型 4	Be 動詞 （状態：形容詞型）	JPN：主語 + 形容詞 +（です / φ） ENG：(am / is / are / be)
文型 5	一般動詞 （思考）	JPN：主語 +（思う / 考える） ENG：(think / believe / consider / guess / suppose / assume)
文型 6	一般動詞 （認識）	JPN：主語 +（知る / わかる） ENG：(know / get to know / understand / find / notice / realize / recognize)
文型 7	助動詞 （可能）	JPN：主語 +（名詞 + できる / 動詞 + ことができる / 動詞 + られる） ENG：(can Verb / be able to Verb / am able to Verb / is able to Verb / are able to Verb)

注：JPN は日本語，ENG は英語を指す．今回の検証では，文型 1 は一人称に限定している．

表 7.2　誤りの検出結果集計表

文の種類		文数	誤り数	実誤り数
複数の主語		138	9	—
単一の主語		497	—	—
文型 2	Be 動詞（存在）現在・肯定	53	28	14
文型 3	Be 動詞（状態：名詞型）現在・肯定	59	17	17
文型 5	一般動詞（思考）現在・肯定	60	13	10
文型 2	Be 動詞（存在）現在・否定	7	7	2
文型 7	助動詞（可能）現在・肯定	13	6	6
文型 4	Be 動詞（状態：形容詞型）現在・肯定	10	5	4
文型 5	一般動詞（思考）過去・肯定	4	4	4
文型 3	Be 動詞（状態：名詞型）過去・肯定	2	2	2
文型 2	Be 動詞（存在）過去・肯定	1	1	0
文型 3	Be 動詞（状態：名詞型）現在・否定	1	1	1
文型 6	一般動詞（認識）過去・肯定	3	1	1
文型 4	Be 動詞（状態：形容詞型）現在・否定	1	1	0
文型 1	人称代名詞（１人称・主格）	97	13	10
無主語		366	—	—
文型 5	一般動詞（思考）現在・肯定	208	106	104
文型 3	Be 動詞（状態：名詞型）現在・肯定	17	10	8
文型 5	一般動詞（思考）過去・肯定	10	9	9
文型 3	Be 動詞（状態：名詞型）過去・肯定	2	2	2
文型 7	助動詞（可能）現在・肯定	1	1	1
文型 7	助動詞（可能）過去・肯定	1	1	1
文型 5	一般動詞（思考）現在・否定	1	1	1
文型 3	Be 動詞（状態：名詞型）現在・否定	2	1	1
文型 2	Be 動詞（存在）現在・否定	1	1	1
文型 2	Be 動詞（存在）現在・肯定	1	0	1

注：(—) は集計していない項目

検証は，第 6 章で規定した 14 種類の誤用判定基準（表 6.6）のうち，7 種類の誤用判定基準（表 7.1）を取り上げて行った．表 7.1 では，述語ベースの文型の現在・否定、過去・肯定、過去・否定の判定基準については省略されているが，検証は 4 パターン全てに対して行っている．

誤用検出の結果，システムは正しく動作することを確認した．表 7.2 は誤りの検出結果を集計したものである．「単一の主語」と「無主語」の日本語文の文型とそれぞれの文型に該当する文の数は「文の種類」欄と「文数」欄に表示している．これらの文型に対して自動判定した誤りのある英作文の数は「誤りの数」欄に，「無主語」と「単一の主語」について降順に並べ替えて示している．そして，手動で判定した実際に誤りのある英作文の数は「実誤りの数」欄に表示している．

表 7.2 より，明らかになったことは二点ある．一点目は初級学習者が好んで使用する日本語のタイプ（複数の主語，単一の主語，無主語），二点目は初級学習者がよく誤りをおかす文型である．

まず，学習者の日本語作文 1001 文は，主語が複数抽出された文（複数の主語）が 138 文，主語が一つのみ抽出された文（単一の主語）が 497 文，主語が抽出されなかった文（無主語）が 366 文となることから，主語が一つのみ抽出された文（単一の主語）の選好性が高いこと，そして，「複数の主語」の場合は英語文の主語の有無のみを判定しているため，138 文のうち主語がない英作文は 9 文であることが「誤りの数」欄から読み取れる．

次に，日本語作文の主語が一つである文に対応する英作文では，文型 2 の Be 動詞（存在）現在・肯定タイプ，文型 3 の Be 動詞（状態：名詞型）現在・肯定タイプ，文型 5 の一般動詞（思考）現在・肯定タイプと文型 1 の人称代名詞（1 人称・主格）タイプの順に初級学習者の誤りが多く見られる．文型 1 は唯一の主語ベースの文型であるが，「私」を使った 97 文のうち，“I”がない文が 13 文もあることがわかる．一方，日本語作文に主語がない文に対応する英作文では，文型 5 の一般動詞（思考）現在・肯定タイプ，文型 3 の Be 動詞（状態：名詞型）現在・肯定タイプ，文型 5 の一般動詞（思考）過去・肯定タイプの順で初級学習者の誤りが多く見られる．もっとも多くの誤りを検出した文型 5 であることから「主語 +（思う / 考える）」という日本語の表現を正しく英語に訳すことができないことを意味する．

ここで誤りとして自動検出された文の具体例を表 7.3 に示す．表 7.3 で示す情報はインターフェイスで表示される詳細情報画面（図 7.7，7.8）で確認できる情報の一部である．

表 7.3　誤りと判定された英作文の例

No.	文型		日本語主語	日本語述語	英語主語述語フレーズ	G.E	R.J
単一の主語							
1	2	Be 動詞（存在）現在・肯定	レストランも	ある	wall hold	○	○
2	2	Be 動詞（存在）現在・肯定	人も	います	I well watch	○	○
3	2	Be 動詞（存在）現在・肯定	理由は	あります	reasons are	△	×
4	3	Be 動詞（状態 : 名詞型）現在・肯定	2 つ目は	ことです	you may die	○	○
5	3	Be 動詞（状態 : 名詞型）現在・肯定	これは	力である	It’s get	○	○
6	3	Be 動詞（状態 : 名詞型）現在・肯定	レストランは	場所であるからです	Restaurants use	○	○
7	5	一般動詞（思考）現在・肯定	私は	考える	*	△	○
8	5	一般動詞（思考）現在・肯定	私は	思います	These are reason	○	○
9	5	一般動詞（思考）現在・肯定	私は	思います	I thinking	×	○
10	1	人称代名詞（１人称・主格）	私は	思う	It is important	○	○
11	1	人称代名詞（１人称・主格）	私は	考える	They smoke	○	○
12	1	人称代名詞（１人称・主格）	私は	思う	job is important	○	○
13	2	Be 動詞（存在）	ものは	ない	Nothing is as	△	×

		現在・否定			good as		
15	4	Be 動詞（状態：形容詞型）現在・肯定	睡眠時間も	少ない	time is even less	×	×
無主語							
16	5	一般動詞（思考）現在・肯定	—	思う	Student should not be give	○	○
17	5	一般動詞（思考）現在・肯定	—	思います	Smoking is harmful	○	○
18	5	一般動詞（思考）現在・肯定	—	思います	Everyone want	○	○
19	3	Be 動詞（状態：名詞型）現在・肯定	—	ことです	Study do not neglect	○	○
20	3	Be 動詞（状態：名詞型）現在・肯定	—	体質だ	I don't like	○	○
21	3	Be 動詞（状態：名詞型）現在・肯定	—	被害です	accident around people who don't smoking	○	○
22	5	一般動詞（思考）過去・肯定	—	思いました	I think	×	○

注：(—) は抽出された主語がないこと，*は検出不可能であったことを示す.

表 7.3 における「英語主語述語フレーズ」欄は，主語から依存関係にある語までの動詞，助動詞，コピュラを含むフレーズである（前節参照）.「G.E」欄では global error かどうかの判定結果，そして「R.J」欄ではシステムの検出結果が正しいかどうかの判定結果を表す.

上表において，「G.E」と「R.J」の両方に丸印が入っている文（No. 1, 2, 4, 5, 6, 8, 10, 11, 12, 16, 17, 18, 19, 20, 21）はシステムにより global error が検出され，その誤用判定結果も正しいということを示している．このことから原言語と目標言語の主語と述語を比較することで容易に文構造に関する global error を検出することが可能であると言える．次に「G.E」にバツ印，「R.J」に丸印が入っている文（No. 9, 22）は global error ではないが，その誤用判定結

果も正しいということを示している．また,「G.E」に∆印,「R.J」に丸印が入っている文（No. 7）は構文解析器で依存構造情報が検出できなかったため，自動的に文構造上に誤りがある誤用文と判断している．そのため誤りの種類は特定していない．最後に「G.E」に∆印,「R.J」にバツ印が入っている文（No. 3, 13）と，「G.E」と「R.J」の両方にバツ印が入っている文（No.15）はシステムによって正しく誤りとして抽出できなかった文である．there is 構文，否定語が主語になる文,「少ない」など，文節の一部分「ない」に対して構文解析器が否定語として処理してしまう場合など，これらの項目に関しては現在のシステムでは対応できないため，残された課題として今後の研究において解決を図りたい．

以上のことより，本システムの利用によって教師は学習者の日英両作文から抽出した主語と述語を比較することで容易に文構造に関する global error を検出することが可能であることが明らかとなり，学生の言語習得状況を容易に把握することができると考えられる．また，自由英作文の内容の多様さから，英作文添削は教師にとって大変負担を強いる作業であるが，本システムを使用することにより，教師による添削時間の短縮が可能となることから，その有効性は高いと判断できる．

7.4　まとめ

本章では，英作文添削における誤用自動検出に必要な基準および手順に従って誤用検出システムを構築し，システムの有効性を示すために検証を行った．誤用検出の検証に際しては，第 6 章で規定した 14 文型のうち 7 文型を使用した．その結果，本システムを使用することで学習者の文構造に関する global error を容易に検出することができ，教師の誤用検出に関わる負担を大きく軽減することが可能だということが明らかとなり，システムの有効性を示すことができた．

第 8 章　結論

本論文では，原言語を取り入れ，日英文型の比較に基づいた global error の自動検出に必要な手法および手順について提案し，英作文誤用検出システムの構築を行った．

まず，第 1 章「序論」においては，本研究の背景，英作文誤用検出システムの必要性，目的について述べた後，本論文の全体構成について説明した．

第 2 章「本研究に関わる定義および先行研究」においては，まず，誤りの定義を確認し，次に，本研究の主たる検出対象である global error の定義を確認した．そして，英作文の誤用に関して，応用言語学と自然言語処理の各観点から先行研究を概観し，英作文誤用検出研究の残された課題を整理した．

第 3 章「本研究で使用する言語データと構文解析器」においては，研究対象とする学習者（日本人英語学習者）の英語習熟度，使用する言語データ（日本語作文データおよびそれに対応する英作文データ，英語添削文データ，英語翻訳文データ，そして教科書データ），利用する構文解析器（英語：The Stanford Parser，日本語：CaboCha）について概説した．

第 4 章「グローバルエラー検出における原言語情報の利用」では，目標言語（英語）と併せて原言語（日本語）を利用した誤用検出について考察し，原言語を利用した学習者英語の誤用検出の可能性を探った．その結果，目標言語（英語）と原言語（日本語）を併せて添削した場合と目標言語（英語）のみで添削した場合では，誤りの箇所にずれがあることが明らかとなった．また，英語学習者の原言語と目標言語の間には意味上のずれが存在すること，英語学習者の目標言語と英語母語話者の原言語との間には品詞構成上の差異が存在すること，そして，原言語（日本語）の主語の有無は目標言語（英語）の誤りに影響を及ぼしていることも明らかとなった．

これらの結果を踏まえると，学習者の言語習得状況を正確に把握し，global error を正確に検出するためには，目標言語と併せて原言語も分析対象にする必要があるということが判明し，原言語を利用した誤用検出の妥当性は高いことが示された．

第 5 章「主語および述語の抽出における構文解析器の利用」では，構文解析器（英語 : The Stanford Parser，日本語 : CaboCha）を利用した，誤りを含む文から主語および述語を抽出する方法を提案し，主語および述語の抽出の正確性について検証を行った．その結果，主語がある日本語作文に関しては主語が 54.9%，述語は 86.0%の正確さで抽出できること，主語がない日本語作文では主語が 75.3%，述語は 83.0%の正確さで抽出できることが明らかとなっ

た．また，英作文に関しては，主語がある日本語作文に対応する英作文では主語が 83.0%，述語は 78.0%の正確さで抽出できること，主語がない日本語作文に対応する英作文では主語が 90.0%，述語は 77.0%の正確さで抽出できることが明らかとなった．

これらの結果を踏まえると，既存の構文解析器を利用して学習者作文から主語および述語を抽出することが可能であること，学習者の誤りを含む文に対しても構文解析器が使用可能なことなどが明らかとなった．

第 6 章「日英文型に基づく誤用検出の基準および手順の提案」では，原言語情報を利用した新たな英作文の誤用検出手法の確立を目指し，1）初級学習者が習得すべき重要な英語文法項目を検討し，その英語文法項目に対応する日本語と併せて日英文型を規定した．そして，2）日英文型比較による自動誤用判定を行うための基準およびその手順について提案した．

具体的には，原言語を利用した新たな英作文の誤用検出手法の確立のため，まず，初級学習者にとって重要な文法項目を文法的特徴に基づいて分類したうえ，本研究で扱う英語文法項目を決定した．併せて，英語文法項目（6 品詞 6 文法項目）について日本語文型を規定し，それに対応する英語文型を 14 種類規定した．そして，日英文型比較による自動誤用判定を行うための誤用判定基準および自動誤用判定の手順について提案を行った．

第 7 章「誤用検出システムの構築および評価」では，英作文添削における誤用自動検出に必要な基準および手順に従ってシステムを構築し，システムの有効性を示すために検証を行った．誤用検出の検証に際しては，第 6 章で規定した 14 文型のうち 7 文型を使用した．その結果，本システムを使用することで学習者の文構造に関する global error を容易に検出することができ，教師の誤用検出に関わる負担を大きく軽減することが可能だということが明らかとなり，システムの有効性を示すことができた．

本研究では，教師の英作文指導を支援するために，global error の自動検出に必要な手法を提案（global error を検出するために目標言語（英語）と併せて原言語（日本語）を分析対象とすること，学習者の誤りを含む英作文に構文解析器を利用し，得られた依存構造情報から独自のルールに基づいて目標言語および原言語の主語および述語をそれぞれ特定すること，目標言語と原言語から抽出した主語と述語に基づいて定めた日英文型を誤用判定基準として用いること）して，英作文誤用検出システムの構築を行った．そして，教師が本システムを使用することで，学習者の文構造に関する global error を比較的容易に検出することが可能なこと，そして教師の誤用検出に関わる負担を軽減することが可能なことがわかり，その有効性が示された．

その一方で課題も残った．現在システムへのデータ入力は手動で行っているため，実際の教育現場で利用できるシステムにするために自動で入力できるシステムにすること，また，今回開発したシステムでは対処できてない文法項目が残っていることから，すでに提案した文型の実装を含めて文型を拡張すること，そして，現在のシステムでは対応できない（there is 構文，否定語が主語になる文，「少ない」など，文節の一部分「ない」に対して構文解析器が否定語として処理してしまう場合など）項目に関しては精査することである．以上の三点については残された課題として捉え，今後の研究において取り組みたい．

謝辞

本論文を執筆するにあたり，指導教員の康敏教授から格別なるご指導とご高配を賜りました．そもそも文系出身である私が教育工学の分野で博士論文をまとめることができたのは，自主性を尊重してくださり研究者としてのやりがいと面白さを示してくださったからに他なりません．ここに心から感謝の意を表します.

大月一弘教授には，ご多忙中にも関わらず私の拙い研究活動に辛抱強く付き合ってくださり多くのご教示を賜りました．心から感謝いたします．情報コミュニケーションコースの森下淳也教授，村尾元教授，清光英成准教授，西田健志准教授には，コロキアムや授業を通して貴重なご助言をいただきました．深く感謝いたします．外国語コンテンツ論コースの柏木治美教授，言語コミュニケーションコースの朴秀娟講師，人文学研究科の Richard HARISON 教授には，第二言語習得，日本語，英語の観点から専門的なご助言をいただきました．本当に有難うございます．また，より良い研究をするため互いに切磋琢磨することができた康研究室のゼミ生である邵帥さん，清光研究室のゼミ生である執行健人さんにも感謝いたします.

最後に，これまで研究のために支援してくれた家族に心から感謝します.

参考文献

坂内昌徳・佐々木裕美（2005）.「第二言語における主語と動詞の「一致」の知識：日本人英語学習者のデータから」.『研究紀要』, 45, 101-109.

Burt, M. K. (1975). Error analysis in the adult EFL classroom. *TESOL Quarterly*, *9*(1), 53-63.

Burt, M. K., & Kiparsky, C. (1974). Global and local mistakes. *New frontiers in second language learning*, 71-80.

CaboCha / 南瓜 https://taku910.github.io/CaboCha/（参照日 2016.09.29）

Corder, S. P. (1967). The significance of learner's errors. *IRAL-International Review of Applied Linguistics in Language Teaching*, *5*(1-4), 161-170.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax?. *Language learning*, 23(2), 245-258.

Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime…?). *Journal of second language writing*, *13*(1), 49-62.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistics perspective.* Harlow, UK: Pearson Education.

橋本進吉（1939）.『改制新文典別記』. 東京：冨山房.

Klein, D. and Manning, C. D. (2003). Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, 423-430.

小池生夫（編）（2003）.『応用言語学事典』. 東京：研究社.

Kudo, T., & Matsumoto, Y. (2003). Fast methods for Kernel-based text analysis, *ACL 2003 in Sapporo*, Japan.

工藤洋路（2009）.「英語ライティング能力のレベルが異なる学習者の Global Error の特徴に関する研究」.『ARCLE REVIEW』, 3, 110-121.

松村明（1971）.『日本文法大辞典』. 東京：明治書院.

水品江里子・麻柄啓一（2007）.「英文の主語把握の誤りとその修正」.『教育心理学研究』, 55(4), 573-583.

永田亮（2013）.「構文解析を必要としない主語動詞一致誤り検出手法」.『電子情報通信学会論文誌. D, 情報・システム』, 96(5), 1346-1355.

中村朋子・中尾有香・畑野喜信（2005）.「疑似初心者の自由英作文に見る英語運用能力：中学校 3 年生・大学 1 年生上位者との比較」『中国地区英語教育学会研究紀要』, 35, 19-27.

太田裕子・佐渡島紗織（2013）.「「自立した書き手」を育成するライティング・センターのチューター研修とチューターの意識：早稲田大学における実践事例と PAC 分析」.『Waseda Global Forum』, 9, 237-277.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London, UK: Longman.

Richards, J. (1971). A non-contrastive approach to error analysis. *English Language Teaching*, *25*, 204-219.

Richards, J. C., & Schmidt, R. W. (2002). Longman dictionary of language teaching and applied linguistics. (3rd ed.). Harlow, UK: Pearson Education Limited.

白畑知彦・冨田祐一・村野井仁・若林茂則(2006)『英語教育用語辞典』. 東京:大修館書店.

Tarone, E. (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language learning*, *30*, 417-431.

時枝誠記（1950）.『日本文法口語篇』. 東京：岩波書店.

Trent, N. (2012). The challenge of English sentence subjects (shugo) to Japanese Leaners. *The 2012 Pan-SIG Proceedings*, 187-195.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, *46*(2), 327-369.

山口明穂・秋本守英（2001）.『日本語文法大辞典』. 東京：明治書院.

山内真理・内田充美（2011）.「日本人英語学習者の中間言語にみられる L1 の痕跡」.『千葉商大紀要』, 49(1), 43-56.

吉田弘子（2010）.「大学ライティングセンターに関する考察：その役割と目的」.『大阪経大論集』, 61, 99-109.

Yang, Y. F., & Meng, W. T. (2013). The effects of online feedback training on students' text revision. *Language Learning & Technology*, *17*(2), 220-238.

要旨

本論文では，原言語を取り入れ，日英文型の比較に基づいた global error の自動検出に必要な手法および手順について提案し，英作文誤用検出システムの構築を行う．

学習者の誤りは第二言語習得の観点から local error と global error に大別することができるが，global error は文の全体構造に影響して内容の理解を阻害するため，教師が学習者の英作文に存在する global error を特定することは困難であり，多大な労力を要する．また，従来の英作文誤り自動検出システムは主として local error を対象としており，学習者の意図を考慮する必要がある global error に焦点を当てたシステムはきわめて限定的であり，教育現場での使用は現実的ではない．それゆえに，英語教育および教育工学の観点から，教師の教育支援のための global error の自動検出は重要性が高いと考えられる．

そこで，本論文では global error を射程とした自動検出に必要な手法および手順を独自に提案し，誤り自動検出システムの開発を行った．

提案の詳細を以下に示す．まず，目標言語（英語）と併せて原言語情報（日本語）を利用した誤用検出について考察し，global error 検出の可能性を示した（第 4 章）．次に，構文解析器（英語：The Stanford Parser，日本語：CaboCha）を利用して，誤りを含む文から主語および述語（動詞）を抽出する方法を提案し，主語および述語がどの程度正確に抽出できるのか検証した（第 5 章）．最後に，日英文型の比較による誤用判定基準を用いた global error を自動検出する方法を提案し（第 6 章），その提案に基づいた英作文誤用検出システムを構築した（第 7 章）．

まず，第 1 章「序論」においては，本研究の背景，英作文誤用検出システム構築の必要性，目的について述べた後，本論文の全体構成について説明した．

第 2 章「先行研究」においては，まず，誤りの定義を確認し，次に，本研究の主たる検出対象である global error の定義を確認した．そして，英作文誤用検出に関して，応用言語学と自然言語処理の各観点から先行研究を概観し，英作文誤用検出研究の残された課題を整理した．

第 3 章「リサーチデザイン」においては，調査対象とする学習者（初級日本人英語学習者），使用する言語データ（日本語作文データおよびそれに対応する英作文データ，英語添削文データ，英語翻訳文データ，そして教科書データ），利用する構文解析器（英語：The Stanford Parser，日本語：CaboCha）について概説した．

第 4 章「グローバルエラー検出における原言語情報の利用」では，目標言語（英語）と併せて原言語（日本語）を利用した誤用検出について考察し，原言語を利用した学習者英語の誤用検出の可能性を探った. 分析の結果, 学習者の言語習得状況を正確に把握し, global error を正確に検出するためには，学習者の目標言語と併せて原言語も分析対象とする必要があることが明らかとなった.

第 5 章「主語および述語（動詞）の抽出における構文解析器の利用」では，構文解析器（英語：The Stanford Parser，日本語：CaboCha）を利用した，誤りを含む文から主語および述語（動詞）を抽出する方法を提案し，主語および述語の抽出の正確性について検証を行った. その結果，学習者の誤りを含む文に対しても構文解析器が使用可能なこと，また，選定方法に条件を加えることで正確性がより高くなる可能性があることなどが判明し，構文解析器を利用することの有効性が確認された.

第 6 章「日英文型に基づく誤用検出の基準および手順の提案」では，日英文型の比較による誤用判定基準を用いた global error を自動検出する方法を提案した．具体的には，初級学習者が習得すべき重要文法項目を含む英語文型を選定し，それに対応する日本語文型を規定した. そして，日英文型比較による誤用判定を行うための基準およびその手順について提案を行った.

第 7 章「誤用検出システムの構築および評価」では，日英文型に基づく誤用検出の基準および手順に従って誤用検出システムを構築し，システムの可能性を探るために検証を行った．その結果，本システムを使用することで文構造に関する global error を容易に検出することができ，教師の誤用検出に関わる負担を大きく軽減することが可能だということが明らかとなり，システムの有効性を示すことができた.

第 8 章「結論」では，第 1 章から第 7 章までを統括し，英作文誤用検出システムがいかにして英語教育支援のための有効なツールとなり得るのか論じた. また，今後の研究課題についても述べた.

本研究に関する発表

学術論文

査読あり

Kawamura, K., Kashiwagi, H. & Kang, M. (2018). Developing an approach toward automatic error detection in learners' English writing based on the source language. *International Journal on Advances in Intelligent Systems*, vol.11 nr 3&4, pp.202-211.

Kawamura, K., Kashiwagi, H. & Kang, M. (2017). An approach toward automatic error detection in learners' English writing based on the source language. *Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, eLmL 2018*, pp. 62-65. (Best Paper Award)

川村晃市（2017）.「構文解析器を利用した主語・述語抽出の検討」.『国際文化学』（神戸大学），31, pp. 51-64.

Kawamura, K., Kashiwagi, H. & Kang, M. (2017). An Approach of Semi-Automatic Correction Feedback for Learners' English Writing Using the Source Language. *Proceedings of Symposium on Language and Sustainability in Asia*, pp. 17-22.

査読なし

川村晃市・柏木治美・康敏 (2019).「教師の英作文添削を支援する誤り検出システム」.『日本教育工学会研究報告集 JSET19-1』（掲載予定）

川村晃市・柏木治美・康敏 (2017).「原言語情報を利用した学習者英語の誤用分析の可能性」.『日本教育工学会研究報告集 JSET17-2』，pp. 225-232.

口頭発表

2017/09 「原言語情報を利用した学習者英作文の分析手法の提案」日本教育工学会第 33 回全国大会（島根大学）

2016/09 「学習者英作文支援システムの構築における原言語活用の検討」日本教育工学会第 32 回全国大会（大阪大学）