

意味カテゴリ情報に基づく帰納的学習を用いた 英語冠詞・前置詞誤り校正システムの試作と印象評価

乙武 北斗*

*福岡大学工学部

荒木 健治**

**北海道大学大学院情報科学研究科

吉村 賢治*

ototake@fukuoka-u.ac.jp

概要

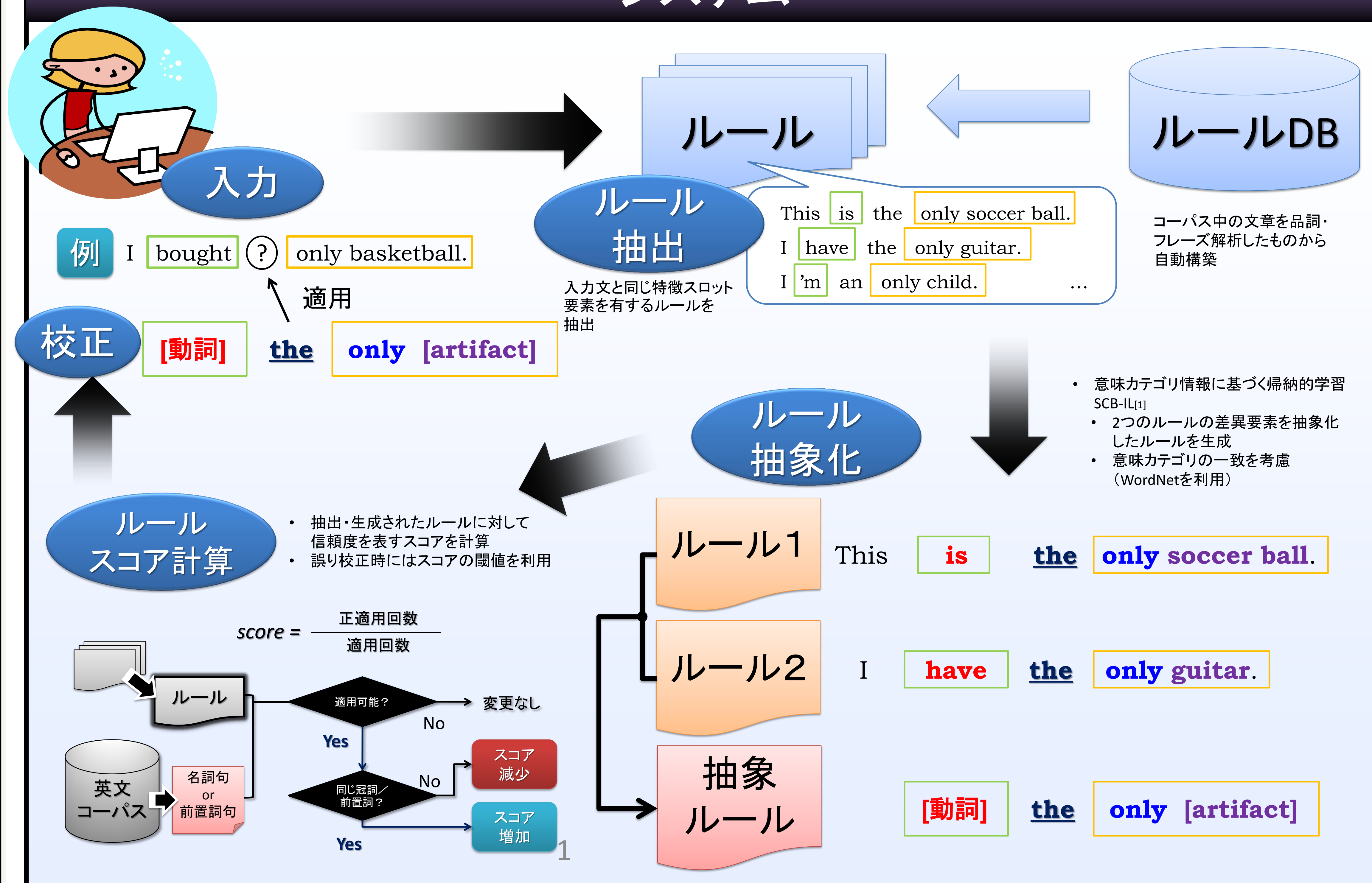
本発表では、英文コーパスから自動生成されるルールに基づいた英語冠詞・前置詞誤りの校正手法、および提案手法を用いた印象評価実験について述べる。

提案手法は、従来の統計的手法にみられるルールを手で作成する労力が必要ない利点に加え、ルールを用いているため、ユーザに校正理由を提示しやすい利点があると考えられる。

本実験は約300名の被験者に対して調査を行い、英文誤り校正システム一般に関する選択形式のアンケート、および提案手法を実装したシステムを利用した感想・意見・提案等を自由記述形式で収集した。

その結果、英文誤り校正システムに興味を持って利用したいという意見が非常に多いことが確認され、その利用方法としては英語論文の執筆や英語学習の補助が挙げられた。

システム



校正ルール

冠詞

This is the **only** **soccer ball** which I bought yesterday.

前置修飾

後置修飾

特徴スロット

ルール

Target	Head	ball (artifact)
	Preceding noun	soccer
	Phrase	NP
	Preposition	-
	Preceding verb	be (stative)
	Following verb	-
	Number	singular
Preceding	Proper noun	no

Preceding	Modifier	only
	Modifier POS	RB

Following	Preposition	Preposition	-
		Determiner	-
		Nouns	-
		Head	-
	Infinitive	Modifier	-
		Verb	-
		Determiner	-
		Object	-
	Relative	Adverb	-
		Subject	I
		Verb	buy (possession)
		Determiner	-
		Object	-
		Adverb	yesterday

Article
the

前置詞

I decided not to **take part** **in** the event.

前置

後置

特徴スロット

ルール

Preceding	Phrase Type	VP
	Head	part
	Head POS	NN
	Modifier	-
	Verb	take
Following	Preposition	-
	Head	event
	Head POS	NN

Preposition
in

評価実験

実験手順

- 冠詞および前置詞について、正しく提示できるかどうかを評価

トレーニングデータ

- Reuters Corpus
英文記事約2億語

テストデータ

- Reuters Corpus
約22万個の冠詞を含む記事
- Reuters Corpus
約2万個の前置詞を含む記事
- トレーニングデータとは別データ

実験システム

- ベースライン
(常に無冠詞, “of”を提示)
- 最大エントロピー分類器
(OpenNLP maxentパッケージ)
- SCB-IL[1]

性能評価

印象評価

実験手順

- 本システムのWebインターフェイスを実装
- 被験者が本システムを利用
 - 大学1年生293名
 - 英語講義中に実施
- 無記名でアンケートに回答
 - 選択形式アンケート
 - 自由記述式の感想・意見等

感想・意見等

- 全体の57%
 - ぜひ利用したい
 - 冠詞・前置詞は日本人にとって苦手
 - 既存の文書作成ソフトには未搭載
- より実用性を高めるために
 - 処理時間の改善(16%)
 - 例文提示の改善(8%)

1.

英語の文法や単語を修正してくれるシステムがあったと仮定します。その架空のシステムは無料ですが、たまに間違っただ言をしましょう。あなたは、どの程度の間違いなら我慢できますか？

i. ほとんど全部を間違えても構わない	1.0%
ii. しばしば間違えても構わない	7.2%
iii. たまに間違える程度なら構わない	38.2%
iv. ごくまれに間違えるなら構わない	37.9%
v. けっして間違ってはならない	15.7%

2.

システムから得られる助言の分量と正確さのどちらを、より重視しますか？

i. 助言が多いかわりに、ときどき誤った助言が混ざるシステムが望ましい	19.9%
ii. 助言が少ないかわりに、誤った助言もほとんど混ざらないシステムが望ましい	80.1%

3.

システムから助言が得られるまでの時間は、どのくらいまで待てますか？

i. 瞬時	9.9%
ii. 数秒	51.4%
iii. 十数秒	28.1%
iv. それ以上でも待てる	10.6%

WebインターフェイスURL:

<http://hkt.tl.fukuoka-u.ac.jp/ac>

ユーザ名: actest

パスワード: actest2009

4.

英語を書くときに、文法や単語を確実に修正してくれるシステムがあったら、使いますか？

i. 必ず使うだろう	33.8%
ii. しばしば使うだろう	40.3%
iii. たまに使うだろう	21.5%
iv. ごくまれに使うだろう	3.1%
v. 使わないだろう	1.4%

5.

英語を書くときに、あなたの書いた英語を添削してくれるシステムと、あなたの書いた日本語を英語に訳してくれるシステムの2種類があったら、どちらを使いますか？

i. 英語を添削するシステムを使う	43.3%
ii. 日本語を英語に訳すシステムを使う	15.7%
iii. 両方とも混ぜて使う	39.9%
iv. いずれも使わない	1.0%

校正例

* 現代英語冠詞事典[2], 英辞郎 on the WEB より

input 冠詞をとらない固有名詞の成功例

Where is George Washington buried?

output 適用ルール: be [NONE] Washington

Where is **[none]** George Washington buried?

関係詞節による指示物の同定の成功例
(実際はenjoyを含むルールから判断)

input I really enjoyed the book I have just finished reading.

output 適用ルール: enjoy THE book

I really enjoyed **[the]** book I have just finished reading.

文脈考慮ができなかったための、**校正失敗例**
(実際はencouragementは特定の行為などを指示する場合、不定冠詞をとる)

input The teacher's words were a great encouragement to him.

output 適用ルール: be **NONE** great encouragement to

The teacher's words were **[none]** great encouragement to him.

input 前置詞校正の成功例

The picture reminds me of my old friend.

output 適用ルール: remind of

The picture reminds me **[of]** my old friend.

input 前置詞校正の**失敗例**

I exchanged dollar for yen.

output 適用ルール: dollar **against** yen

I exchanged dollar **[against]** yen.

今後の課題

- 印象評価の結果を踏まえて
 - 処理速度の向上(現在約10秒/1文)
 - 校正結果と共に出力する情報の検討と評価
- 性能の改善
 - 冠詞では最大エントロピー分類器との併用で改善の可能性
 - 前置詞の選択にはもっと情報が必要
- 動詞一名詞のコロケーション誤りに対応

[1] H. Otake et al., "English Article Correction System Using Semantic Category Based Inductive Learning Rules," LNAI Vol. 5866

[2] 樋口昌幸, [例解]現代英語冠詞事典, 大修館書店