

センター試験『国語』現代文の傍線部問題を解く ベースライン法

佐藤 理史^{1,a)} 加納 隼人² 西村 翔平² 駒谷 和範¹

概要: 大学入試センター試験『国語』の現代文で出題される、いわゆる「傍線部問題」を解く方法を定式化し、実装した。実装した方法は、「評論」の「傍線部問題」の半数を正しく解くことができた。

キーワード: 大学入試センター試験, 国語現代文, 傍線部問題, オーバーラップ率

A Baseline Method for Solving Questions about Underlined Segments in Contemporary Japanese Language, National Center Test for University Admissions

SATOSHI SATO^{1,a)} HAYATO KANO² SHOHEI NISHIMURA² KAZUNORI KOMATANI¹

Abstract: We have formalized and implemented a method for solving questions about underlined segments in contemporary Japanese language, National Center Test for University Admissions. The method can solve a half of these questions in “critical essay” correctly.

Keywords: National Center Test for University Admissions, contemporary Japanese language, questions about underlined segments, overlap ratio

1. はじめに

日本において、大学入試問題は、学力(知力および知識力)を問う問題として定着している。この大学入試問題を計算機に解かせようという試みが、国立情報学研究所のグランドチャレンジ「ロボットは東大に入れるか?」というプロジェクトとして2011年に開始された[1]。このプロジェクトの中間目標は、2016年までに大学入試センター試験で(東京大学の二次試験に進めるような)高得点を取ることである。

我々は、このプロジェクトに参画し、本年度(2013年度)より、大学入試センター試験の『国語』現代文の問題を解くシステムの開発に取り組んでいる。対象として、『国語』

現代文を選んだのは、2013年4月の時点でほとんど手つかずの状態であったこと、および、「自然言語処理の研究者ならば、『国語』現代文をやらずして何をやる」と第一著者が考えたことによる。

次節で述べるように、『国語』の現代文の設問の過半は、傍線部問題とよばれる設問である。研究を開始した4月下旬の時点で、次のような計画を立てた。

- (1) 本年度の主な対象を、「評論」の傍線部問題とする。
- (2) 「評論」の傍線部問題の50%を正しく解くシステムの実現を、本年度の目標とする。
- (3) 解き方として、船口明の攻略法[2]を採用する。
- (4) これと並行して、表層的な手がかりに基づく方法をベースライン法として実装する。

本稿では、上記のうち、(4)の進行状況について報告する。後で詳しく述べるように、表層的な手がかりに基づく方法の性能は、我々の予想をはるかに上回り、それ単体で、当初の目標(2)を達成してしまうこととなった。これを受

¹ 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学工学部
School of Engineering, Nagoya University

a) ssato@nuee.nagoya-u.ac.jp

- 問5 傍線部D「行為と行為をつなぐこの空間の密度を下げていくのが、現在の住宅である」とあるが、それはどういうことか。その説明として最も適当なものを、次の1～5のうちから一つ選べ。
- 1 現在の住宅では、仕事部屋や子ども部屋など目的ごとに空間が切り分けられており、それぞれの用途とはかかわらない複数の異なる行為を同時に行ったり、他者との関係を作り出したりするような可能性が低下してしまっていること。
 - 2 現在の住宅では、ゾーニングが普及することでそれぞれの空間の独立性が高められており、家族であってもそれぞれが自室で過ごす時間が増えることで、人と人とが触れあい、関係を深めていくことが少なくなっていること。
 - 3 現在の住宅では、空間の慣習的な使用規則に縛られない設計がなされており、居住者たちがそのときその場で思いついたことを実現できるように、各自がそれぞれの行為を同時に行えるようになっていること。
 - 4 木造家屋などかつての居住空間では、居間や台所など空間ごとの特性が際立っていたが、現代の住宅では、居住者が部屋の用途を交換でき、空間それぞれの特性がなくなっていること。
 - 5 木造家屋などかつての居住空間では、人体の運動と連動して空間が作り変えられるような特性があったが、空間ごとの役割を明確にした現在の住宅では、予想外の行為によって空間の用途を多様にすることが困難になっていること。

図1 傍線部問題の例 (2011年度本試験第1問の問5 (2011M-C5))

表1 センター試験『国語』の大問構成 (出典 [5])

近代以降の文章 (現代文)	第1問	評論	50点
		第2問	小説
古典		第3問	古文 50点
		第4問	漢文 50点

表2 使用する傍線部問題の一覧

	第1問「評論」		第2問「小説」	
		問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5
2001 本試験	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5
2003 本試験	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5
2005 本試験	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5
2007 本試験	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	—
2009 本試験	問2, 問3, 問4, 問5	—	問2, 問3, 問4, 問5	問2, 問3, 問4
2011 本試験	問2, 問3, 問4, 問5	問2, 問3, 問4, 問5	問2, 問3, 問4	問2, 問3, 問4
	計 40 問		計 38 問	

けて、我々は、目標を上方修正するとともに、この方法の性能の上限を明らかにすることに取り組んだ。

以下、本稿は、次のように構成されている。まず、2節で、大学入試センター試験の『国語』の構成と、それに含まれる傍線部問題について説明する。3節では、我々がベースライン法として採用する方法について述べ、4節ではその実装について述べる。5節では、実施した実験の結果を示し、その結果について検討する。

2. センター試験『国語』と傍線部問題

大学入試センター試験の『国語』では、毎年、大問4題が出題される [5]。その大問構成を表1に示す。この表に示すように、現代文に関する出題は、第1問の「評論」と第2問の「小説」であり、『国語』の半分を占めている。

第1問の「評論」は、何らかの評論から抜き出された文章(本文)と、それに対する6問の設問から構成される。6問の内訳は、通常、以下のようにになっている。

問1 漢字の書き取り問題が5つ出題される。

問2-問5 傍線部について、その内容や理由が問われる。

問6 本文全体にかかわる問題で、2006年以降は本文の論の進め方や本文の構成上の特徴などが問われる。

一方、第2問の「小説」は、何らかの小説から抜き出された文章(本文)と、それに対する6問の設問から構成される。6問の内訳は、通常、以下のようにになっている。

問1 語句の意味内容を問う問題が3つ出題される。

問2-問5 傍線部を参照し、登場人物や心情・人物像・行動の理由などが説明問題の形で問われる。

問6 本文全体の趣旨や作者の意図、表現上の特徴などが問われる。

これらの設問のうち、「評論」「小説」の両者の問2-問5

を傍線部問題と呼ぶ。傍線部問題の具体例を図1に示す。傍線部問題の配点は、2009年度本試験では、第1問32点、第2問33点の計65点であり、現代文の配点100点の約2/3を占める。船口 [2] が暗に指摘しているように、『国語』の現代文を「攻略」するカギは、傍線部問題の「攻略」にある。

使用する試験問題

本研究では、2001年度から2011年度の奇数年の大学入試センター試験の本試験および追試験の『国語』(2005年以前は『国語I-II』)を使用する。ただし、諸般の事情により、本文等が欠けているものがあり、それらは使用しない。表2に、本研究で使用する傍線部問題の一覧を示す。

なお、以降では、設問を指し示すIDとして、以下のような4つの情報を盛り込んだ形式を採用する。

- (1) 年度 (4桁)
- (2) 試験区分 (M:本試験, S:追試験)
- (3) 出題区分 (C:評論, N:小説)
- (4) 設問番号 (2,3,4,5)

たとえば、図1に示した、2011年度本試験第1問「評論」の問5は、「2011M-C5」と表す。((2)と(3)の間に、ハイフン「-」を挟む。)

3. 傍線部問題の定式化

我々の知る限り、大学入試の『国語』の傍線部問題を計算機に解かせる試みは、これまでに存在しない*1。そのため、色々な方法を試すまえに、比較的単純な方法で、どのぐらいの正解率が得られるのかを明らかにしておく必要がある。言い換えるならば、今後の研究のベースラインとなるような方法の設定と、その性能の調査が必要である。

3.1 含意関係認識問題とみなす

図 1 に示した典型例を用いて、傍線部問題を解く方法を考えよう。この設問は、本文中に出現する傍線部 D 「行為と行為をつなぐこの空間の密度を下げているのが、現在の住宅である」を参照し、「それはどういうことか」という問い方で、傍線部を別の言い方で説明した文を選択肢の中から選ぶことを求めている。

このような設問においては、傍線部を別の言い方で表現した箇所が、問題の本文中のどこか*2に書かれていることが多い。「本文中のどこかに書かれている」を仮定すれば、「本文は、正解である選択肢を含意する(正解である選択肢は本文に含意される)」とみなすことができる。ここで「含意」という用語を用いるのは、本文中の言い方と選択肢の言い方は、表現としては必ずしも一致しないからである。

日本語テキストの含意関係認識は、評価型ワークショップである RITE [6] および RITE2 [7] で取り上げられている。これらのワークショップの BC サブタスクは、「2つのテキスト t_1 と t_2 が与えられたとき、 t_1 が t_2 を含意するか否かを判定する」タスクである。傍線部問題を上記のようにみなせば、それは、本文 (t_1) と選択肢 (t_2) の含意認識問題となる。ただし、センター試験の傍線部問題は 5 択の選択問題なので、含意するか否かの二値分類として定式化するのではなく、含意される可能性が最も高い選択肢を選ぶという選択問題として定式化するのが自然である。

以上の考え方にに基づき、傍線部問題を図 2 のように定式化する。この定式化では、本文 T と選択肢集合 C ののみを使用し、問題を、 T が c_i を含意する可能性の高さを表すスコア $\text{score}(T, c_i)$ に帰着する ($\text{score}(T, c_i) \geq 0$ を仮定する)。このスコアを、以下では、(本文と選択肢の) 照合スコアと呼ぶ。

3.2 設問文の極性を考慮する

ほとんどの傍線部問題の設問文は、「最も適当なものを、次の 1～5 のうちから一つ選べ」という形式となっている。しかし、2001S-C4 のように、「適当でないものを、次

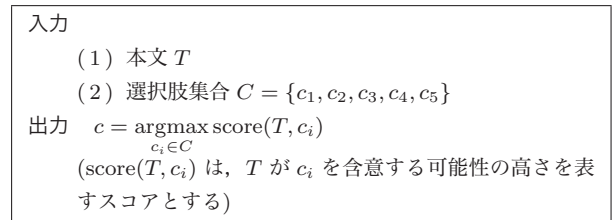


図 2 傍線部問題の定式化 (1)

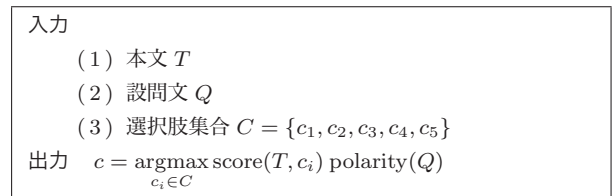


図 3 傍線部問題の定式化 (2)

の 1～5 のうちから一つ選べ」という形式も存在する。このような「適当でないもの」を選ぶ設問に対しては、照合スコアが最も小さなものを選択するのが自然である。

これを採用するために、設問文 Q の極性を判定する関数 $\text{polarity}(Q)$ を導入する。この関数は、設問が「適当なもの」を要求している場合に +1 を、「適当でないもの」を要求している場合に -1 を返すものとする。この関数を用いて、定式化を図 3 のように変更する。

3.3 本文の一部と照合する

傍線部を別の言い方で表現した箇所は、多くの場合、本文中の傍線部の周辺にあると考えるのが妥当である。実際、我々人間が傍線部問題を解くとき、本文中の傍線部の前後に注目するのは、標準的な戦略である。

このような戦略は、選択肢 c_i を本文 T 全体と照合するのではなく、あらかじめ、本文から、傍線部の言い換えが書かれていそうな部分 \hat{T} を抜き出し、 \hat{T} と c_i の照合スコアを計算することで、具体化できる。

本文の一部を取り出す方法として、次のような方法が考えられる。

- (1) 本文の先頭から、当該傍線部までを \hat{T} とする。
- (2) 一つ前の設問で参照された傍線部から、当該傍線部までを \hat{T} とする。
- (3) 傍線部の前後のある範囲を \hat{T} とする。

これらの方法と、採用する単位(文または段落)を組み合わせることにより、多くのバリエーションが生まれることになる。

ここでは、それらに関数として抽象化しよう。本文 T と設問文 Q から、本文 T の一部 \hat{T} を選択・抽出する関数を extract を導入する。すなわち、 extract は、次のような関数である。

$$\hat{T} = \text{extract}(T, Q) \tag{1}$$

このように、本文 T の一部を選択してからスコアを計算

*1 CLEF2013 では、QA4MRE のサブタスクの一つとして、Entraces Exams が実施され、そこでは、センター試験の『英語』の問題が使用された。

*2 かならずしも、一箇所にまとまって書かれているとは限らない。

<p>入力</p> <p>(1) 本文 T</p> <p>(2) 設問文 Q (本文中のどこを参照しているかの情報が含まれるものとする)</p> <p>(3) 選択肢集合 $C = \{c_1, c_2, c_3, c_4, c_5\}$</p> <p>出力 $c = \operatorname{argmax}_{c_i \in C} \operatorname{score}(\operatorname{extract}(T, Q), c_i) \operatorname{polarity}(Q)$</p>
--

図 4 傍線部問題の定式化 (3) 本文照合法

<p>入力</p> <p>(1) 本文 T</p> <p>(2) 設問文 Q (本文中のどこを参照しているかの情報が含まれるものとする)</p> <p>(3) 選択肢集合 $C = \{c_1, c_2, c_3, c_4, c_5\}$</p> <p>出力 $c = \operatorname{argmax}_{c_i \in \operatorname{ps}(C, Q)} \operatorname{score}(\operatorname{extract}(T, Q), c_i) \operatorname{polarity}(Q)$</p> <p>関数</p> <p>(1) $\operatorname{score}(\hat{T}, c_i)$: テキスト \hat{T} と c_i から、照合スコアを計算する</p> <p>(2) $\operatorname{polarity}(Q)$: 設問文の極性を判定する</p> <p>(3) $\operatorname{extract}(T, Q)$: 本文 T から、その一部を抽出する</p> <p>(4) $\operatorname{ps}(C, Q)$: 選択肢集合 C から、その部分集合を選ぶ</p>

図 5 傍線部問題の定式化 (4) 本文照合法+選択肢事前選抜

する方法を採用すると、傍線部問題は、図 4 のように定式化される。この定式化を、本文照合法と名付ける。

なお、厳密に言えば、設問文 Q には、一つ前の設問の傍線部の情報は含まれないが、その情報は、本文を参照することによって得られるものと仮定する。事実、センター試験では、傍線部には、A, B, C, D の記号が順に振られるため、当該傍線部が B であれば、一つ前の設問の傍線部は A であることがわかる。

3.4 選択肢の事前選抜を導入する

もれ聞くとくによれば、「本文を読まなくても、選択肢さえ見れば、正解がわかる」と豪語する猛者がいるとのこと。それが真実であるどうかはさておき、「設問文と5つの選択肢をよく読めば、本文を参照せずとも、正解にはならない選択肢のいくつかをあらかじめ排除できる」場合は存在すると考えられる。

そのような考えに基づき、選択肢の事前選抜を導入する。事前選抜 ps (pre-select) は、選択肢集合 C と設問文 Q から、 C の部分集合を返す関数として定式化する。

$$\hat{C} = \operatorname{ps}(C, Q), \quad \hat{C} \subseteq C \quad (2)$$

これを先の本文照合法と組み合わせると、図 5 の定式化が得られる。この定式化が、本論文で提案するベースライン法の枠組である。

4. 実装

前節の定式化に基づいて傍線部問題ソルバーを実装するためには、 score , $\operatorname{polarity}$, $\operatorname{extract}$, ps の4つの関数を実装

表 3 実装法の概要

$\operatorname{score}(\hat{T}, c_i)$	A	$\operatorname{overlap_ratio}_D(A; \hat{T}, c_i)$ (文字オーバーラップ率)
	A^2	$\operatorname{overlap_ratio}_D(A^2; \hat{T}, c_i)$ (文字 bigram のオーバーラップ率)
	W	$\operatorname{overlap_ratio}_D(W; \hat{T}, c_i)$ (形態素表層形のオーバーラップ率)
	L	$\operatorname{overlap_ratio}_D(W; \hat{T}, c_i)$ (形態素原形のオーバーラップ率)
$\operatorname{polarity}(Q)$		表層の手がかりに基づき、極性判定を行なう
$\operatorname{ps}(C)$	no	選択肢の事前選抜を行わない
	yes	選択肢のうちの1つを除外する
$\operatorname{extract}(T, Q)$	P-a-e	本文すべて
	P-a-0	本文先頭から、当該傍線部を含む段落まで
	P-b-0	ひとつ前の傍線部を含む段落から、当該傍線部を含む段落まで
	P-b-e	ひとつ前の傍線部を含む段落から、本文末尾まで
	P-m-n	当該傍線部を含む段落と、その前 m 段落、後 n 段落
	S-m-n	当該傍線部を含む文と、その前 m 文、後 n 文
\bar{X} (bar)		(抽出法 X で、当該傍線部を含む文を除外)

する必要がある。表 3 に、今回実装した方法の一覧を示す。なお、前節の定式化では、正解と考えられる選択肢を一つ出力する形になっているが、実際のシステムは、選択肢を照合スコア順にソートした結果(すなわち、それぞれの選択肢の順位)を出力する仕様となっている。なお、照合スコアが一致した場合は、選択肢番号の若いものを上位とする。

4.1 オーバーラップ率

照合スコア score , および、選択肢の事前選抜 ps の実装には、オーバーラップ率を用いる。

我々は、オーバーラップ率の定義として、何を単位としてオーバーラップ率を計算するかという、その単位を抽象化した定義を採用する [3], [4]。まず、ある集合 E を仮定する。この集合の要素が、オーバーラップ率を計算する際の単位となる。集合 E の具体例としては、たとえば、文字集合 A , 形態素集合 W , あるいは、文字 n -gram の集合などを想定する。

オーバーラップ率の算出の出発点となる式は、2つの文字列 t_1 と t_2 に共通に出現する要素の数を求める次式である。

$$\operatorname{overlap}(E; t_1, t_2) = \sum_{x \in E} \min(\operatorname{fr}(x, t_1), \operatorname{fr}(x, t_2)) \quad (3)$$

ここで、 $\operatorname{fr}(x, t)$ は、文字列 t における $x (x \in E)$ の出現回数を表す。この値を、 t_2 の長さ、あるいは、 t_1 と t_2 の長さの

和で正規化することにより、オーバーラップ率を定義する。

$$\text{overlap_ratio}_D(E; t_1, t_2) = \frac{\text{overlap}(E; t_1, t_2)}{\sum_{x \in E} \text{fr}(x, t_2)} \quad (4)$$

$$\text{overlap_ratio}_B(E; t_1, t_2) = \frac{2 \text{overlap}(E; t_1, t_2)}{\sum_{x \in E} \text{fr}(x, t_1) + \sum_{x \in E} \text{fr}(x, t_2)} \quad (5)$$

前者の overlap_ratio_D は、 t_2 の長さのみで正規化したもので、方向性を持った (directional) オーバーラップ率となる。後者の overlap_ratio_B は、 t_1 と t_2 の長さの和で正規化したもので、方向性を持たない、双方向性 (bidirectional) のオーバーラップ率となる。

4.2 照合スコア

本文の一部 \hat{T} と選択肢 c_i の照合スコアには、上記で定義したオーバーラップ率 overlap_ratio_D を用いた。

$$\text{score}(\hat{T}, c_i) = \text{overlap_ratio}_D(E, \hat{T}, c_i) \quad (6)$$

ここで、オーバーラップを測る際の単位 (要素) 集合 E とし、以下の 4 種類を実装した。

- (1) A : 文字集合
- (2) A^2 : 文字 bigram の集合
- (3) W : 形態素表層形の集合
- (4) L : 形態素原形の集合

いずれの場合も、句読点は要素に含めなかった。

形態素解析器には mecab-0.994 を、形態素解析辞書には、ipadic-2.7.0 または unidic-2.1.0 を用いた。すなわち、 W と L は、それぞれ 2 種類存在することになる。

4.3 設問文の極性判定

設問文の極性判定は、文字列マッチングで実装した。対象とした問題は限られているので、極性判定結果は、人間の判断とすべて一致する。

4.4 本文の一部の抽出

段落 (P) 単位および文 (S) 単位の抽出を実装した。抽出する領域は、連続領域を採用した。すなわち、抽出単位、抽出開始点、抽出終了点の 3 つの情報によって、抽出領域は定まる。

抽出開始・終了点は、当該傍線部を含む単位 (段落または文) を基準点 0 とし、その前後何単位であるかを、整数で表す。たとえば、S- m - n は、当該傍線部を含む文と、その前 m 文、後 n 文を表す (全部で $m+1+n$ 文となる)。この他に、本文先頭 (a)、前問の傍線部の位置 (b)、本文末尾 (e) という 3 種類の特別な位置を指定できるようにした。さらに、当該傍線部を含む文を除外するというオプション (\bar{X}) も実装した。

4.5 選択肢の事前選抜

選択肢の事前選抜には、次の方法を採用した。

- (1) それぞれの選択肢 c_i において、他の選択肢 c_j との双方向文字オーバーラップ率 $\text{overlap_ratio}_B(A; c_i, c_j)$ をすべて計算し、その平均を求める。
- (2) 得られた平均値が最も低い選択肢を、選択肢集合から除外する。(最終順位付けでは、かならず 5 位とする)

5. 実験と検討

5.1 実験結果

実装した傍線部問題ソルバーを用いて、評論 40 問を解いた結果を表 4 および表 5 に示す。この表の各行の先頭の欄 (ID) は、本文抽出法 (extract) に対応しており、次の 2 つの数字は、その抽出法 (ID) で抽出された文数 (40 問の平均値)、および、該当傍線部を含む文を除外した場合 (\bar{ID}) で抽出された文数を示す*3。斜線で区切られた 4 つの数字は、ある要素集合を単位としてオーバーラップ率を計算した場合に対応し、それぞれの数字は、順に、以下の場合の正解数を示す。

- (1) 抽出法 ID + 事前選抜なし (no)
- (2) 抽出法 \bar{ID} + 事前選抜なし (no)
- (3) 抽出法 ID + 事前選抜あり (yes)
- (4) 抽出法 \bar{ID} + 事前選抜あり (yes)

表 4 の 2 行目 (P-a-0) の A 欄の最初の数字 20 が、我々に衝撃を与えた数字である。これは、

本文の先頭から当該傍線部を含む段落までを \hat{T} とし、抽出し (P-a-0)、 \hat{T} と各選択肢 c_i との照合スコアを文字オーバーラップ率 (A) で計算して、スコアが最大値を取る選択肢を選んだ場合、「評論」の傍線部問題の半分 (20/40) が正しく解けることを意味する。

センター試験の設問は 5 択問題であるので、解答する選択肢をランダムに選んだとしても 1/5 の確率で正解する。40 問においてランダムに解答を選んだ場合、正解する問題数は、 8 ± 4.96 ($p = 0.05$) である。この値と比べ、正解数 20 問は有意に多い。

我々は、このような性能が得られることを、まったく予期していなかった。この結果を受けて、我々は、色々な設定 ($84 \times 6 \times 4 - 12 = 2004$ 通り) での性能を網羅的に調べた。こうして得られたのが、先に示した表 4 と表 5 である。これらの表では、正解数 20 以上をボールド体で表示している。さらに、正解数が 22 以上となった 16 の設定とその設定における正解の順位分布 (第 n 位として出力された正解がいくつあるか) を、表 6 に示した。

*3 S- m - n で文数が $m+1+n$ を越えるのは、2003S-C5 の設問文が複数の傍線部 (正確には、波線部) を含むためである。この場合、最初に現れる波線部の前方 m 文から、最後に現れる波線部の後方 n 文までを抽出する。

表 4 「評論」 に対する実験結果 (その 1)

ID	ave. sent.	A		A ²		W-ipadic		W-unidic		L-ipadic		L-unidic	
	ID / \bar{ID}	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	no / yes	
P-a-e	68.9/67.8	14/ 14/ 18/ 18	14/ 12/ 16/ 14	15/ 17/ 17/ 17	14/ 15/ 16/ 16	15/ 15/ 16/ 16	13/ 13/ 15/ 15	11/ 13/ 12/ 13	13/ 15/ 15/ 16	11/ 12/ 12/ 13	13/ 13/ 15/ 15		
P-a-0	40.0/39.0	20/21/21/22	15/ 13/ 17/ 15	14/ 14/ 15/ 15	11/ 13/ 12/ 13	13/ 15/ 15/ 16	11/ 12/ 12/ 14	14/ 16/ 17/ 18	12/ 14/ 13/ 15	12/ 14/ 13/ 15	10/ 10/ 13/ 13		
P-b-0	21.5/20.5	19/ 19/ 22/22	14/ 13/ 16/ 16	13/ 13/ 15/ 16	11/ 12/ 12/ 14	14/ 16/ 17/ 18	12/ 14/ 13/ 15	17/ 16/ 19/ 17	13/ 13/ 15/ 15	13/ 13/ 15/ 15	13/ 13/ 15/ 15		
P-b-e	50.4/49.3	14/ 14/ 18/ 18	13/ 12/ 15/ 14	13/ 14/ 16/ 16	13/ 14/ 16/ 16	14/ 14/ 16/ 16	12/ 14/ 15/ 16	17/ 16/ 19/ 17	13/ 13/ 15/ 16	13/ 13/ 15/ 16	13/ 13/ 15/ 16		
P-0-0	6.0/ -	14/ -/17/ -	12/ -/17/ -	12/ -/15/ -	11/ -/14/ -	8/ -/13/ -	13/ -/16/ -						
P-1-0	10.8/ 9.7	16/ 17/ 19/ 19	14/ 14/ 16/ 17	12/ 15/ 17/ 18	15/ 17/ 17/ 18	14/ 15/ 18/ 17	13/ 15/ 14/ 16						
P-1-1	15.2/14.2	15/ 16/ 16/ 16	14/ 16/ 15/ 18	15/ 15/ 17/ 17	16/ 18/ 17/ 18	18/ 17/ 19/ 18	13/ 13/ 14/ 15						
P-2-0	15.2/14.2	18/ 18/ 21/21	15/ 14/ 17/ 18	12/ 13/ 16/ 16	13/ 13/ 15/ 15	12/ 14/ 15/ 16	13/ 13/ 15/ 15						
P-2-1	19.6/18.6	17/ 17/ 18/ 18	12/ 14/ 13/ 15	14/ 14/ 16/ 16	14/ 13/ 16/ 15	17/ 16/ 19/ 17	13/ 13/ 15/ 15						
P-2-2	24.2/23.2	16/ 15/ 19/ 18	16/ 16/ 16/ 16	17/ 17/ 19/ 19	19/ 19/ 21/21	19/ 19/ 20/19	16/ 15/ 18/ 17						
P-3-0	19.1/18.1	18/ 16/ 21/19	15/ 16/ 17/ 18	11/ 12/ 14/ 14	13/ 14/ 15/ 16	12/ 14/ 14/ 15	10/ 12/ 13/ 15						
P-3-1	23.6/22.5	18/ 16/ 19/ 17	11/ 12/ 12/ 13	11/ 10/ 12/ 11	12/ 12/ 13/ 13	14/ 14/ 16/ 15	10/ 10/ 12/ 12						
P-3-2	28.1/27.1	17/ 16/ 20/18	13/ 14/ 14/ 15	14/ 13/ 15/ 14	16/ 16/ 17/ 17	16/ 15/ 17/ 16	13/ 13/ 15/ 15						
P-3-3	31.3/30.3	15/ 14/ 18/ 16	12/ 12/ 14/ 14	11/ 12/ 13/ 13	14/ 14/ 15/ 15	14/ 14/ 14/ 14	13/ 12/ 15/ 14						
P-4-0	22.8/21.8	18/ 16/ 21/18	15/ 14/ 17/ 16	9/ 10/ 11/ 11	10/ 12/ 12/ 14	11/ 13/ 13/ 15	13/ 14/ 15/ 16						
P-4-1	27.2/26.2	17/ 16/ 19/ 17	9/ 9/ 10/ 10	12/ 12/ 13/ 12	11/ 12/ 11/ 12	14/ 13/ 15/ 14	12/ 12/ 14/ 14						
P-4-2	31.9/30.8	16/ 14/ 19/ 17	12/ 12/ 12/ 12	15/ 14/ 15/ 14	15/ 16/ 16/ 17	17/ 16/ 17/ 16	13/ 13/ 15/ 15						
P-4-3	35.0/34.0	14/ 13/ 19/ 17	11/ 10/ 12/ 11	12/ 13/ 13/ 13	13/ 14/ 14/ 15	15/ 14/ 15/ 14	13/ 13/ 15/ 15						
P-4-4	38.2/37.1	14/ 14/ 17/ 16	10/ 10/ 11/ 10	14/ 14/ 14/ 14	15/ 16/ 15/ 16	17/ 16/ 17/ 16	14/ 14/ 16/ 16						
S-1-0	2.3/ 1.2	10/ 11/ 14/ 13	11/ 15/ 16/ 17	10/ 9/ 16/ 14	10/ 11/ 13/ 12	6/ 11/ 11/ 14	9/ 11/ 12/ 12						
S-1-1	3.3/ 2.2	12/ 14/ 14/ 15	10/ 10/ 14/ 13	10/ 11/ 14/ 14	11/ 13/ 14/ 16	11/ 8/ 15/ 12	9/ 8/ 11/ 12						
S-2-0	3.3/ 2.2	16/ 11/ 20/15	14/ 13/ 17/ 18	13/ 11/ 16/ 12	12/ 11/ 15/ 12	12/ 13/ 16/ 16	11/ 11/ 12/ 15						
S-2-1	4.3/ 3.2	14/ 14/ 15/ 17	12/ 14/ 15/ 16	16/ 13/ 17/ 14	11/ 13/ 13/ 15	10/ 11/ 13/ 13	12/ 12/ 14/ 15						
S-2-2	5.2/ 4.2	14/ 14/ 15/ 16	12/ 15/ 14/ 17	14/ 13/ 16/ 14	14/ 15/ 15/ 16	13/ 12/ 16/ 14	14/ 12/ 16/ 14						
S-3-0	4.3/ 3.2	12/ 11/ 16/ 15	16/ 18/ 18/ 20	17/ 16/ 19/ 19	14/ 14/ 15/ 15	12/ 15/ 16/ 18	13/ 13/ 14/ 16						
S-3-1	5.3/ 4.2	14/ 15/ 16/ 17	15/ 16/ 16/ 19	16/ 14/ 18/ 16	12/ 12/ 15/ 15	10/ 11/ 13/ 14	15/ 12/ 17/ 15						
S-3-2	6.2/ 5.2	16/ 14/ 18/ 17	15/ 17/ 16/ 19	16/ 16/ 18/ 18	15/ 16/ 16/ 18	13/ 12/ 16/ 15	14/ 14/ 15/ 15						
S-3-3	7.1/ 6.0	12/ 14/ 14/ 15	16/ 15/ 17/ 16	16/ 16/ 17/ 17	14/ 15/ 14/ 16	16/ 14/ 18/ 15	14/ 15/ 15/ 16						
S-4-0	5.3/ 4.2	13/ 13/ 18/ 19	15/ 14/ 18/ 20	17/ 14/ 20/18	14/ 12/ 17/ 15	10/ 12/ 13/ 17	16/ 15/ 17/ 18						
S-4-1	6.3/ 5.2	16/ 17/ 18/ 20	12/ 16/ 15/ 18	14/ 13/ 18/ 17	15/ 14/ 18/ 16	12/ 13/ 17/ 16	15/ 17/ 18/ 17						
S-4-2	7.2/ 6.2	17/ 18/ 19/ 20	11/ 12/ 13/ 14	14/ 14/ 18/ 17	15/ 16/ 17/ 17	13/ 12/ 18/ 17	14/ 15/ 16/ 17						
S-4-3	8.1/ 7.0	15/ 17/ 17/ 19	13/ 14/ 15/ 16	14/ 15/ 17/ 17	14/ 15/ 15/ 17	15/ 14/ 19/ 17	14/ 16/ 16/ 18						
S-4-4	8.9/ 7.9	16/ 15/ 18/ 16	13/ 14/ 15/ 16	16/ 15/ 19/ 18	13/ 14/ 13/ 14	17/ 16/ 20/18	14/ 14/ 16/ 16						
S-5-0	6.2/ 5.2	13/ 15/ 18/ 21	13/ 15/ 18/ 20	15/ 13/ 19/ 17	15/ 16/ 18/ 16	12/ 13/ 16/ 16	15/ 15/ 17/ 17						
S-5-1	7.2/ 6.2	15/ 17/ 17/ 20	13/ 17/ 16/ 20	13/ 15/ 16/ 18	15/ 18/ 18/ 18	12/ 13/ 17/ 16	12/ 16/ 15/ 16						
S-5-2	8.2/ 7.1	16/ 18/ 18/ 20	12/ 14/ 14/ 17	15/ 14/ 19/ 18	16/ 16/ 19/ 19	13/ 12/ 17/ 16	14/ 18/ 16/ 19						
S-5-3	9.1/ 8.0	15/ 16/ 16/ 17	12/ 14/ 14/ 16	15/ 15/ 19/ 17	14/ 16/ 16/ 18	14/ 14/ 18/ 17	15/ 17/ 17/ 19						
S-5-4	9.9/ 8.8	15/ 16/ 17/ 17	11/ 13/ 14/ 15	15/ 16/ 18/ 19	13/ 15/ 14/ 16	14/ 13/ 17/ 16	14/ 14/ 16/ 16						
S-5-5	10.7/ 9.7	16/ 17/ 18/ 18	11/ 14/ 13/ 15	14/ 14/ 16/ 16	13/ 14/ 14/ 15	15/ 14/ 18/ 17	13/ 13/ 15/ 15						
S-6-0	7.2/ 6.1	14/ 18/ 19/ 22	17/ 18/ 20/22	18/ 17/ 21/20	17/ 17/ 19/ 18	15/ 16/ 18/ 19	18/ 15/ 19/ 18						
S-6-1	8.2/ 7.1	16/ 18/ 18/ 20	15/ 19/ 18/ 22	16/ 17/ 19/ 19	15/ 18/ 17/ 19	15/ 16/ 19/ 19	14/ 17/ 17/ 18						
S-6-2	9.1/ 8.1	17/ 20/18/21	12/ 14/ 15/ 16	15/ 16/ 18/ 19	16/ 16/ 17/ 17	16/ 16/ 19/ 18	16/ 19/ 17/ 20						
S-6-3	10.0/ 8.9	17/ 20/19/20	12/ 14/ 15/ 16	14/ 16/ 17/ 18	13/ 15/ 15/ 17	15/ 15/ 18/ 16	14/ 15/ 15/ 16						
S-6-4	10.8/ 9.8	16/ 18/ 17/ 18	12/ 13/ 14/ 14	15/ 16/ 17/ 18	12/ 15/ 12/ 15	16/ 16/ 17/ 17	13/ 14/ 14/ 15						
S-6-5	11.7/ 10.6	16/ 17/ 17/ 18	12/ 13/ 13/ 14	15/ 14/ 16/ 15	11/ 13/ 11/ 13	16/ 15/ 17/ 16	13/ 14/ 14/ 15						
S-6-6	12.5/ 11.4	16/ 18/ 17/ 18	13/ 13/ 14/ 15	14/ 14/ 16/ 16	12/ 13/ 12/ 13	16/ 16/ 19/ 19	13/ 13/ 16/ 16						

表 5 「評論」に対する実験結果 (その 2)

ID	ave. sent.		A		A ²		W-ipadic		W-unidic		L-ipadic		L-unidic	
	ID	\overline{ID}	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
S-7-0	8.1/	7.0	15/18/19/21		16/18/20/22		15/16/19/19		16/17/20/19		13/15/17/19		16/13/19/17	
S-7-1	9.1/	8.1	19/19/20/20		14/19/17/22		15/18/18/19		15/18/17/19		14/17/17/19		14/16/17/18	
S-7-2	10.0/	9.0	19/21/19/21		12/14/16/17		15/16/18/18		15/15/17/17		16/15/19/18		15/18/18/21	
S-7-3	10.9/	9.8	18/21/19/20		12/14/15/16		15/16/18/17		13/15/15/17		16/15/19/17		15/16/16/17	
S-7-4	11.8/10.7		19/19/19/18		14/15/16/17		16/16/18/17		11/13/12/14		16/15/18/16		13/13/14/14	
S-7-5	12.6/11.5		18/17/18/17		15/14/16/15		16/15/17/16		11/12/11/12		16/15/18/16		13/13/14/14	
S-7-6	13.4/12.3		19/17/18/17		14/14/15/16		15/16/16/17		12/13/12/13		16/16/20/19		14/14/17/17	
S-7-7	14.2/13.2		19/18/19/18		16/16/16/16		16/16/18/18		13/14/13/14		17/17/21/20		13/13/17/17	
S-8-0	9.0/	8.0	16/15/20/19		15/17/21/22		16/15/20/18		18/18/22/19		15/17/18/21		19/15/20/18	
S-8-1	10.0/	9.0	19/17/20/18		14/19/19/22		16/18/18/18		18/19/19/18		15/17/19/20		16/18/19/20	
S-8-2	10.9/	9.9	20/19/20/20		15/17/18/18		16/18/19/19		16/17/17/18		16/17/20/20		15/18/18/21	
S-8-3	11.8/10.8		20/20/20/19		13/16/15/18		15/16/18/17		13/16/15/18		15/16/18/17		15/17/16/18	
S-8-4	12.7/11.6		19/18/19/17		15/17/17/19		16/17/19/20		13/16/13/16		16/17/18/18		14/14/15/15	
S-8-5	13.5/12.4		17/16/17/16		16/16/17/17		15/16/18/19		12/14/13/15		16/17/18/18		14/14/15/15	
S-8-6	14.3/13.3		20/18/19/18		14/14/15/15		14/16/17/18		12/14/13/15		16/17/20/20		14/14/16/16	
S-8-7	15.2/14.1		20/19/20/19		15/15/16/16		15/16/19/19		13/15/14/16		17/18/21/21		13/13/16/16	
S-8-8	16.0/14.9		19/18/20/18		15/16/16/17		18/18/21/21		15/17/16/18		19/19/22/22		14/14/17/17	
S-9-0	9.9/	8.9	17/16/21/20		16/18/20/23		15/16/20/20		15/17/19/18		17/19/19/23		18/15/18/17	
S-9-1	10.9/	9.9	18/17/20/18		14/17/19/21		14/17/17/19		15/17/17/17		14/18/18/21		14/16/17/19	
S-9-2	11.8/10.8		20/20/20/19		16/17/19/19		14/16/18/18		15/16/17/17		14/16/18/19		14/15/17/18	
S-9-3	12.7/11.7		19/18/19/17		13/15/15/17		13/15/17/16		14/15/17/17		12/13/15/14		15/15/16/16	
S-9-4	13.6/12.5		18/18/19/18		14/16/16/17		14/15/17/18		14/15/15/15		13/14/15/15		13/13/14/14	
S-9-5	14.4/13.3		17/17/17/16		14/14/15/15		13/14/16/17		13/14/15/15		13/14/15/15		13/13/14/14	
S-9-6	15.2/14.2		20/19/20/19		13/13/14/14		13/13/16/16		12/14/14/15		13/14/17/17		14/14/16/16	
S-9-7	16.1/15.0		19/19/20/19		14/14/15/15		13/14/17/17		13/14/15/15		16/17/20/20		14/14/16/16	
S-9-8	16.9/15.8		18/18/19/18		15/15/16/16		16/17/19/19		15/16/17/17		18/18/21/21		15/15/17/17	
S-9-9	17.7/16.6		18/18/19/18		15/16/15/16		17/19/20/20		17/18/20/20		19/20/21/22		16/16/19/19	
S-10-0	10.8/	9.8	17/16/21/20		15/16/19/21		14/14/18/17		15/18/19/19		15/16/17/19		16/15/18/18	
S-10-1	11.8/10.8		18/17/19/18		14/17/19/21		13/15/15/17		14/17/16/17		15/17/19/20		15/16/18/19	
S-10-2	12.8/11.7		20/20/20/19		14/16/17/18		13/14/16/16		14/16/16/17		16/16/20/19		16/16/19/19	
S-10-3	13.6/12.6		20/19/20/18		13/15/15/17		12/13/15/14		13/15/16/17		15/15/18/16		16/16/17/17	
S-10-4	14.5/13.4		19/19/19/18		13/15/15/16		14/15/17/18		14/16/15/16		14/14/16/15		15/15/15/15	
S-10-5	15.3/14.2		18/18/18/17		13/14/14/15		13/14/16/17		13/15/15/16		14/14/16/15		15/15/15/15	
S-10-6	16.1/15.1		21/20/21/20		13/14/15/16		13/13/16/16		13/16/15/17		14/14/17/16		16/16/17/17	
S-10-7	16.9/15.9		19/19/21/20		15/16/15/16		13/14/17/17		14/16/16/17		17/17/20/19		16/16/18/18	
S-10-8	17.8/16.7		18/18/20/19		16/17/16/17		16/17/19/19		14/16/16/17		18/18/21/20		16/16/18/18	
S-10-9	18.6/17.5		18/18/20/19		17/18/16/17		16/18/19/19		16/18/19/20		19/20/21/21		18/18/20/20	
S-10-10	19.4/18.3		17/17/19/18		15/17/16/16		15/15/18/18		16/18/19/20		18/19/20/20		17/17/19/19	

5.2 実験結果を検討する

表 4 と表 5 を観察すると、以下のことに気づく。

- (1) 照合するテキスト \hat{T} が極端に短い場合を除き、ほとんどの場合 (2004 通り中 1828 通りの設定) で、正解数はランダムな方法より有意に多い。すなわち、「評論」の傍線部問題に対しては、ベースライン法は、有効に機能する。
- (2) 照合スコアのオーバーラップ率の計算には、文字 (A) を用いると相対的に成績がよい場合が多い。文字オーバーラップ率を使用するのであれば、長さに関するパラメータの設定が不要な P-a-0 や P-b-0 が有望である。

- (3) 照合スコアのオーバーラップ率の計算に、文字 bigram (A^2)、形態素出現形 (W)、形態素原形 (L) を用いた場合は、比較的短い \hat{T} のいくつかに対して、成績がよい。今回の実験で最も成績がよかった正解数 23 は、抽出法 $\overline{S-9-0}$ 、照合法 A^2 または L-unidic、事前選抜あり (yes) の場合に得られた。
- (4) 照合テキスト \hat{T} から、当該傍線部を含む文を除外した方が、除外しなかった場合よりも、成績は若干よい傾向を示す。
- (5) 選択肢の事前選抜は、正解数を増やす効果が見られる。なお、今回使用した 40 問において、正解が事前選抜

表 6 正解数が 22 以上の設定と正解の順位分布 (「評論」)

ID	bar	match	ps	ave.	1	2	3	4	5
S-9-0	bar	A ²	yes	8.9	23	8	2	6	1
S-9-0	bar	L-ipadic	yes	8.9	23	6	2	8	1
P-a-0	bar	A	yes	39.0	22	5	7	5	1
P-b-0		A	yes	21.5	22	6	5	6	1
P-b-0	bar	A	yes	20.5	22	6	6	5	1
S-6-0	bar	A	yes	6.1	22	5	6	6	1
S-6-0	bar	A ²	yes	6.1	22	9	3	5	1
S-6-1	bar	A ²	yes	7.1	22	8	4	5	1
S-7-0	bar	A ²	yes	7.0	22	8	2	7	1
S-7-1	bar	A ²	yes	8.1	22	6	5	6	1
S-8-0	bar	A ²	yes	8.0	22	8	2	7	1
S-8-0		W-unidic	yes	9.0	22	5	5	7	1
S-8-1	bar	A ²	yes	9.0	22	6	5	6	1
S-8-8		L-ipadic	yes	16.0	22	3	5	9	1
S-8-8	bar	L-ipadic	yes	14.9	22	5	3	9	1
S-9-9	bar	L-ipadic	yes	16.6	22	7	3	7	1

によって除外される設問は、1 問だけ存在した。

(6) 用いる形態素解析辞書によって、得られる結果は若干異なる。これは、形態素として認定する単位、および、原形の認定法の違い*4による。今回の実験では、ipadic を使用した方が、相対的によい結果が得られた場合が多かった。

5.3 性能の上限を見積もる

今回採用したベースライン法で、どの程度の性能が達成可能であるかを見積もってみよう。性能の上限は、それぞれの設問において、

- (1) 最も適切な本文の一部 \hat{T} が選択でき、かつ、
- (2) 最も適切な照合スコアを選択できると仮定した場合の正解率で与えられる。ここでは、

- 形態素解析辞書には unidic を用いる
- 選択肢の事前選択は行なわない*5

こととした 668 (= 84 × 4 - 4) 通りの設定を採用し、各設問毎に 668 通りの設定の成績 (正解の順位) を集計した。その結果を表 7 に示す。

この表に示すように、668 通りの設定のいずれにおいても正解を出力できなかった設問は、2 問 (2001S-C5 と 2009M-C4) のみであった。すなわち、38/40(=95%) の設問に対して、ベースライン法は正解を出力できる可能性がある。

この表において、1 位となった設定数が多い設問は、言わば「ストライクゾーンが広い」設問である。つまり、パラメータ選択に「鈍感」であり、機械にとってやさしい設問である。たとえば、2003S-C3 (1 行目) は、668 通り中 664 通りの設定で正解が得られている。逆に、1 位となった設

表 7 各設問における正解順位分布 (「評論」)

Q	1	2	3	4	5
2003S-C3	664	4	0	0	0
2001S-C4	612	51	3	1	1
2001M-C2	582	78	8	0	0
2005M-C5	558	90	15	5	0
2007S-C3	553	82	26	5	2
2007M-C5	544	52	50	14	8
2001M-C5	543	96	25	4	0
2005S-C2	438	187	43	0	0
2005S-C5	420	211	32	5	0
2009M-C3	396	86	98	52	36
2011M-C2	377	178	62	37	14
2003S-C5	359	305	4	0	0
2001S-C3	358	78	217	8	7
2001M-C4	348	150	147	23	0
2007S-C2	285	140	152	63	28
2003M-C3	258	171	235	4	0
2003S-C2	251	383	34	0	0
2005M-C2	223	252	105	64	24
2007M-C3	215	132	227	18	76
2007M-C2	205	266	183	12	2
2003M-C4	202	103	8	34	321
2011M-C3	195	159	105	80	129
2005M-C4	181	156	133	130	68
2011M-C5	177	290	149	42	10
2003S-C4	174	143	171	148	32
2001S-C2	172	268	189	36	3
2007S-C5	168	386	79	32	3
2007S-C4	123	353	173	19	0
2005S-C3	102	25	74	212	255
2005M-C3	100	150	248	107	63
2003M-C2	77	279	47	265	0
2005S-C4	76	74	79	156	283
2007M-C4	45	66	267	207	83
2001M-C3	44	102	219	198	105
2009M-C2	25	117	236	286	4
2011M-C4	20	151	290	193	14
2009M-C5	14	197	120	254	83
2003M-C5	14	28	86	124	416
2001S-C5	0	3	37	31	597
2009M-C4	0	1	14	151	502

定数が少ない設問は、正解を出力するのが難しい設問である。たとえば、2003M-C5 (下から 3 行目) は、668 通り中 14 通りの設定でしか正解が得られない。

これらのことを考慮して、次に、もうすこし現実的な到達目標を考えよう。正解率 1/5 でランダムに 668 回の試行を行なった場合の正解数は 133.6 ± 20.26 ($p = 0.05$) である。この値の上限をひとつの目安として、これよりも多い正解数が得られた設問は、設問に応じた適切なパラメータ選択により、正解を導ける可能性が高いとみなそう。このような設問は、40 問中 27 問 (67.5%) である。実際、今回の実験で得られた最大正解数は 23 であり、正解数 27 は、

*4 unidic では、語彙素を原形として採用した。

*5 事前選抜を上限の計算に含めるのは複雑なので、除外した。

現実的に到達可能な範囲にあると考えられる。

5.4 好成績の理由を考える

比較的単純なベースライン法でも、半数以上の設問が正しく解けるのは、どうしてだろうか。その理由は、おそらく、「センター試験がよく練られた試験問題である」ということになろう。センター試験の問題は、当然のことながら、「正解が一意に定まる(大多数の人が、正解に納得できる)」ことが必要である。

答の一意性を保証できる『数学』の問題とは異なり、『国語』の傍線部問題は、潜在的には多数の「正解文」が存在する。作問者の立場に立てば、そのうちの一つを選択肢に含め、それ以外を選択肢に含めないように問題を作らなければならない。そのため、正解選択肢とそれ以外の選択肢の間に、明示的な差異を持ち込まざるを得なくなる。そして、そのために持ち込まれた差異は、オーバーラップ率のような表層的な指標においても、識別できる差異として現れてしまうのであろう。もし、この推測が正しいとすれば、「良い問題であれば、機械にも解ける」ということであり、本稿で示したベースライン法は、センター試験ならではの性質を利用していることになろう。

5.5 「小説」に適用する

本ベースライン法を、そのまま「小説」の傍線部問題に適用すると、どのような結果が得られるであろうか。その疑問に答えるために、「評論」と同様の実験を、「小説」に対しても実施した。対象とした「小説」の傍線部問題は計38問なので、ランダムに解答すると、 7.6 ± 4.83 問 ($p = 0.05$) の正解が得られることになる。

実験において、統計的に有意な結果(正解数13以上)が得られたのは、2004通り中13通りの設定のみであった。これらを表8に示す。さらに、「評論」と同様に、各設問に対しても668通りの実験結果を集計した*6。正解数が 133.6 ± 20.26 の上限を越えたのは、38問中10問であった。これらの結果より、「小説」に対しては、ベースライン法の性能は、チャンスレベルと大差がないとみなすのが妥当であろう。

5.6 まとめ

以上の結果をまとめると、次のようになる。

- (1) 今回実装したベースライン法は、「評論」の傍線部問題に対しては有効に機能し、半数以上の設問に対して正解を出力することができる。今回の実験から推測される性能の上限は95%、現実的に到達可能な性能は65-70%である。
- (2) 本ベースライン法は、「小説」の傍線部問題に対しては

表8 正解数が13以上の設定と正解の順位分布(「小説」)

ID	bar	match	ps	ave.	1	2	3	4	5
S-1-0	bar	W-unidic	yes	1.2	15	6	4	7	6
S-2-0	bar	L-unidic	yes	2.2	15	4	2	11	6
P-0-0		A ²	no	6.0	14	5	6	7	6
P-0-0		A ²	yes	6.0	14	6	7	5	6
S-1-0	bar	L-unidic	yes	1.2	14	6	5	7	6
S-1-1	bar	L-unidic	yes	2.2	14	4	7	7	6
S-1-0		A ²	no	2.3	14	4	7	7	6
S-1-0	bar	A ²	yes	1.2	13	7	3	9	6
S-1-0	bar	W-unidic	no	1.2	13	6	5	6	8
S-1-1	bar	W-unidic	yes	2.2	13	8	4	7	6
S-3-1	bar	A ²	yes	4.2	13	5	7	7	6
S-3-2	bar	A ²	yes	5.2	13	5	6	8	6

機能しない。その性能はチャンスレベルと同等である。本ベースライン法で「『評論』が解ける」という事実を言い換えるならば、それは、「『評論』では、本文に書かれていることが問われる」ということである。これに対して、「『小説』が解けない」という事実は、その裏返し、すなわち、「『小説』では、本文に書かれていないこと(心情や行間)が問われる」ということを示している。このような差異の存在を、船口[2]も指摘しているが、表層的なオーバーラップ率を用いる比較的単純な方法においても、その差異が明確な形で現れることが判明した。

謝辞 本研究では、国立情報学研究所のプロジェクト「ロボットは東大に入れるか?」から、データの提供を受けて実施した。

参考文献

- [1] 新井紀子, 松崎拓也: ロボットは東大に入れるか?—国立情報学研究所「人工頭脳」プロジェクト—, 人工知能学会誌, Vol. 27, No. 5, pp. 463-469 (2012).
- [2] 船口 明: きめる! センター国語現代文, 学研教育出版(1997).
- [3] 服部昇平, 佐藤理史: 多段階戦略に基づくテキストの意味関係認識: RITE2 タスクへの適用, 情報処理学会研究報告 2013-NLP-211 No.4/2013-SLP-96 No.4, 情報処理学会(2013).
- [4] Hattori, S. and Sato, S.: Team SKL's Strategy and Experience in RITE2, *Proceedings of the 10th NTCIR Conference*, pp. 435-442 (2013).
- [5] 教学社編集部: センター試験過去問研究 国語 (2014年版 センター赤本シリーズ), 教学社 (2013).
- [6] Shima, H., Kanayama, H., Lee, C.-W., Lin, C.-J., Mitamura, T., Miyao, Y., Shi, S. and Takeda, K.: Overview of NTCIR-9 RITE: Recognizing Inference in TExt, *Proceedings of NTCIR-9 Workshop Meeting*, pp. 291-301 (2011).
- [7] Watanabe, Y., Miyao, Y., Mizuno, J., Shibata, T., Kanayama, H., Lee, C.-W., Lin, C.-J., Shi, S., Mitamura, T., Kando, N., Shima, H. and Takeda, K.: Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10, *Proceedings of the 10th NTCIR Conference*, pp. 385-404 (2013).

*6 表8の結果に基づき、形態素解析辞書には ipadic ではなく unidic を採用した。