

技術文章作成能力の育成を目指した教育支援ツールの開発

大野 博之[†] 稲積 宏誠[†]

[†] 青山学院大学 理工学部 情報テクノロジー学科 〒 229-8558 神奈川県相模原市淵野辺 5-10-1

E-mail: †{oono,hiro}@ina-lab.it.aoyama.ac.jp

あらまし 近年、日本語の乱れが注目されるとともに、日本語教育の重要性が指摘されている。しかし、大学カリキュラムにおいても、座学では身につかないことや教員の負担を理由にコース学習されることは少なかった。そこで、既存の形態素解析や係り受け解析技術を利用し、ルールベースに基づく文章校正と、書き手の意図や趣旨を正確に表現することを目的とした推敲を支援するシステムを提案する。校正機能としては、技術文章を書くうえでの基本的なルールを対象とし、推敲機能としては、修飾語と被修飾語の関係から書き手の意図を確認する方法に着目した。

キーワード 日本語、技術文章、推敲、校正、ルールベース

Development of an education support tool for improvement of ability for sentence making

Hiroyuki OONO[†] and Hiroshige INAZUMI[†]

[†] College of Science and Engineering, Aoyama Gakuin University Fuchinobe 5-10-1, Sagami-hara-shi, Kanagawa, 229-8558 Japan

E-mail: †{oono,hiro}@ina-lab.it.aoyama.ac.jp

1. はじめに

1978年に日本で最初のワードプロセッサが発売され、自然言語処理の実用的な取り組みとして、1980年以降から日本語を機械処理しようという研究が数多く行われている。そうした中で、形態素解析や係り受け解析といった基礎研究に加え、インテリジェントエディタや機械翻訳システムの開発など幅広い取り組みがなされてきた。特に近年においては、テキストマイニングとして、意味理解や文脈理解の技術を発展させた工学的な応用分野としても注目されている。

一方、日本語の乱れとともに日本語教育の重要性が改めて指摘されている。これは単に「国語」教育の問題だけではなく、すべての分野の教育レベルを左右する重大な問題と位置づけられ、大学においても、教育の根幹を揺るがしかねない課題となりつつある。高校までの「国語」教育とは異なり、実際に役に立つ文章を書くためには、一定のルールに従うこと、そして正確かつ簡潔な表現を行うことが求められる。

ここでは、各種専門分野や実務的な有用性から、レポート、マニュアルおよび技術系の論文など技術文章を対象として、それらを書き上げる能力をいかにして身につけさせるかに注目する。これらは、必ずしも単純に机上の学問で身につくものではなく、実践の中で身につくものであることから、従来は大学カ

リキュラムなど、まとまったコースとしての教育はなされてこなかった。その結果、近年になって、特色ある大学教育支援プログラム(特色GP)[1]においても、これらの取り組みが大きく取り上げられる結果に至っている。しかしながら、このようなテーマに対するコースウェアや教材の開発および教育そのものは教員の負担が増え、非常に大きなコストを要する。

そこで、本研究では、上記の問題解決を実現するための支援システムの構築を目的とする。そのために、まず本稿では既存の形態素解析や係り受け解析技術を利用し、ルールベースに基づく文章校正と、書き手の意図や趣旨を正確に表現することを目的とした推敲を支援するシステムを提案する。将来的には、目的に合致した文章作成を実現させることのできる推敲システムを構築し、教育システムツールへと展開させることを想定したものである。

なお、本稿において、文と文章という用語は、次のような意味で用いる。文とは1語もしくはそれ以上の語からなり、ひと区切りのまとまりある考えを示したもので句点によって終わるものとし、文章とはそれらの文が集まったものとする。また、各語の品詞についてはchasenで用いられている日本語辞書(ipadic2.6.3)[2]の品詞体系を用いる。これは、情報処理振興事業協会(IPA)で設定されたIPA品詞体系を一部修正したものである。本稿内では本文と区別がつくように[名詞-一般]のよう

に品詞を括弧で囲んで表記することとした。

2. 関連研究

関連研究の1つとして、日本語文章を字面でのみ解析し、推敲に活用できる情報を執筆者に提供する日本語文章推敲支援ツール「推敲」[3][4][5]がある。「推敲」の基本方針は、実際に推敲するのは執筆者であり、問題点を指摘できれば良いという点と、1万字規模の文章でも実用上の時間内に解析できる点としており、前者は本研究の基本方針と同じである。このツールでは、事前に日本語文章を調査し、さまざまな品詞などの情報を抽出するための表層的なデータベースを作成して、実際に推敲に必要な情報の提示方法を構築している。推敲に関する機能としては、カタカナの表記ゆれ、アルファベットの表記ゆれ、文末の特徴、長すぎる文、長い平がな列、文章の文字情報、非対応括弧、受身、指示詞、とりたて詞、接続助詞「が」、連用中止法、「は」を複数含む文、否定表現を複数含む文、「ように」+否定表現、などが挙げられる。

また、日本語校正支援システム「FleCS」[6][7][8]では、校正の対象を「単語そのものが誤っている」「単語は合っているが付近の単語との関係がおかしい」「表記揺れなど文章全体に関わる誤り」としており、これらに関する指摘を行うツールとなっている。仕組みとしては、文章を形態素解析する「文法解析部」、形態素の木構造化処理を行い、文を管理する「文書構造化部」、ヒューリスティックな誤り指摘パターンを保持する「ルールベース部」を核とし、基本的にパターンマッチングにより修正候補を出しながら、執筆者に修正を促す。校正に関する機能としては、入力での誤り、かな漢字変換の誤り、助詞抜けの誤り、禁止語・誤用語、表記揺れ、典型的な重言、二重否定や文の長さなど文体に関するもの、などに対応している。これらのルールは、一般用やマニュアル作成などに特化したルールに容易に変更できるように設計されている。

3. システム概要

本システムにおける処理の流れを図1に示す。まず、文章の入力画面(図2)に処理対象の文章を入力する。次に、入力された文章を文に分割する処理を行う。これは、次のステップで利用する係り受け解析ツールに対して、文章を与えた場合に、例えば句点があっても別々の文と解釈されず、1つのつながった文として解析されてしまうためである。そして分割処理後、すべての文ごとに対して形態素解析・係り受け解析を行う。形態素解析・係り受け解析には、cabocha[9]を利用した(cabochaはデフォルトで形態素解析にchasen[10]を利用している)。そして、形態素解析から得られた品詞情報をもとに、各校正チェック項目に関するルールベースに従った指摘を図3のように行う。また、係り受け解析から得られた係り受け情報を利用して、文章の意図を執筆者に確認させながら、推敲を支援する。

3.1 文章の分割処理

先に述べたように、入力された文章に対して基本的な形態素解析・係り受け解析を行うにあたり、文章を文に分割する必要がある。日本語文章では、句点として「。」と「。」が存在して

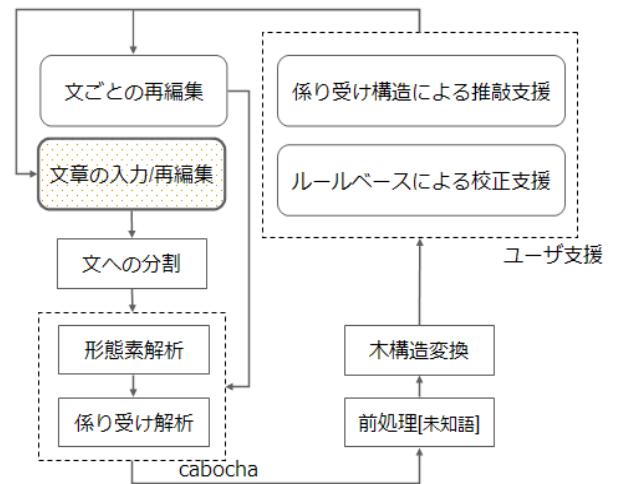


図1 システムの処理の流れ

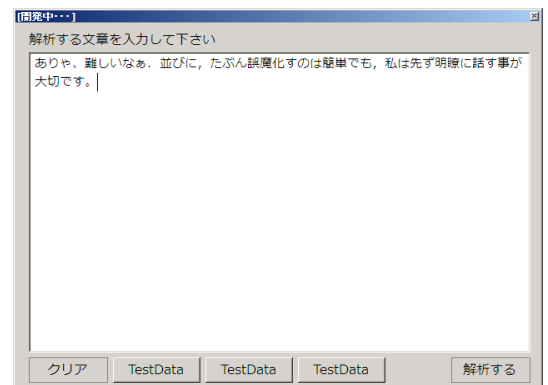


図2 文章入力画面

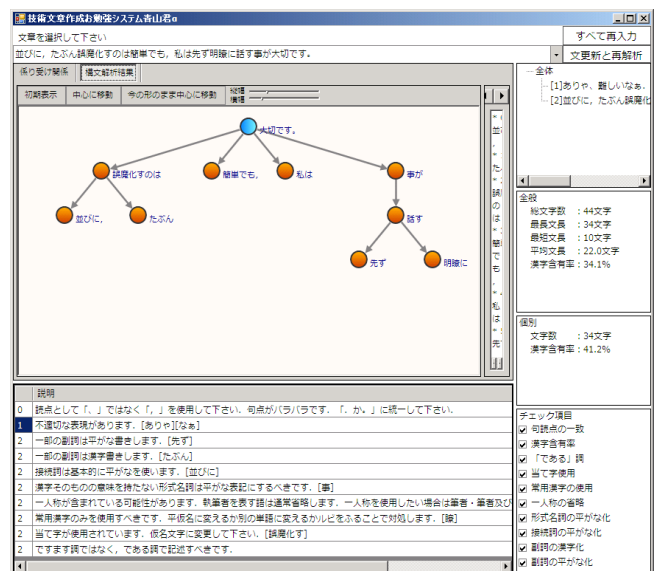


図3 解析結果後の画面概観

おり、技術文章としても執筆要綱などで特に指定されていない限り、どちらを使用しても構わない。そこで、「。」もしくは「。」が現れた場合に、文として分割するという簡潔なルールを用いた。

3.2 形態素解析と係り受け解析

形態素解析と係り受け解析には、cabochaを利用したcabocha

は、文節区切り及び係り受け関係の解析に Support Vector Machines を利用した係り受け解析器で、決定的な解析アルゴリズム (Cascaded Chunking Model) を採用しており、現実的な解析速度を実現している。入力を生文で受け付け、内部で chasen を用いた形態素解析を行った後、係り受け解析を行うため、複数のシステムを経由することなく、処理を簡略化できるため採用した。この chasen によって、文は品詞付きの形態素に分けられ、cabocha により、さらに文節ごとに区切られ、表 1 のような係り受け情報が付与される。

表 1 cabocha による係り受け解析結果例

* 0 1D 0/1 1.55444968				
論文	ロンブン	論文	名詞-一般	
の	ノ	の	助詞-連体化	
* 1 2D 0/1 0.00000000				
執筆	シッピツ	執筆	名詞-サ変接続	
は	ハ	は	助詞-係助詞	
* 2 -1O 0/0 0.00000000				
難しい	ムズカシイ	難しい	形容詞-自立	形容詞-イ段 基本形
.	.	.	記号-句点	

3.3 未知語に対する前処理

キーボードから通常入力可能な半角記号は、chasen では未知語として解釈されるため、全角記号の場合と同様の品詞付けを行った。それ以外の未知語に関しては、技術文章においては、その多くが専門用語であると思われるので [名詞-一般] を品詞付けした。

3.4 係り受け関係の木構造化

文の各文節をノードとし、文末をルートとした木構造に変換する。ノードとなる文節は複数の形態素からなり、その情報を保持している。木構造に変換する目的は、グラフィカル表示によって執筆者の文章構造認識を容易にするためと、推敲支援機能で利用するためである。

3.5 ルールベースによる校正支援

ルールベースに基づいた校正支援機能を提供する。ここでは、執筆者が自分の執筆状況に合わせられるように各チェック機能を任意に選択できるようにした。なお、各チェック機能項目に関する詳細は「4. 校正支援機能」で述べることとする。

3.6 修飾語に注目した推敲支援

係り受け関係による修飾語と被修飾語の関係を提示することにより、執筆者に文意を確認させながら推敲させる支援機能を提供する。この推敲支援に関して、詳細は「5. 文章推敲支援」で述べることとする。

3.7 文の再編集による確認

執筆者は、ルールベースの校正チェックや修飾語に注目した推敲支援機能の結果から、文の再編集を繰り返しながら、文を書き上げる。今回は、指摘するということを重視し、修正候補の提示などは行わないこととした。これは、本システムが執筆者の執筆作業を助けることが主目的ではなく、見落とししている点や正確さが不十分な点を自らが考え気づき、どのようにすれば良いかを発見することを体験してもらう狙いがあるためである。そのような理由から、文章全体でも、文単体でも簡単に再

解析できる GUI を設計し、修正した文の解析結果をすばやく表示し、随時確認できるようにした。

4. 校正支援機能

技術文章を書くうえで、いくつかの点は定型的なルールにのっとって記述しなければならないこともある。ありがちなミスとして挙げられるのは、例えば、常用漢字以外を使用してしまうことや当て字を使用してしまうこと、副詞について無意識のまま分かりづらい表記を使ってしまうことなどである。これらは、教科書のようなもので暗記して身につくものでもなく、繰り返し指摘を受けながら、経験的に身につけていくものである。そこで、校正チェック対象項目として、いろいろと考えられるが、今回はその中の 10 個程度のチェック機能を実装し、経験的な学習を支援することとした。

4.1 文章全体に関するチェック

まず、文章全体に対して句読点と漢字含有率、長文についてのチェックを行う。ここでのチェックに関しては、次の 4 つの基準 [11][12] に従っている。

- 読点「、」は通常縦書きの場合に用いられるものであるため、「、」が使用された場合には指摘する。
- 句点「。」はどちらを使用しても構わないが、どちらかに統一すべきであるので、統一されていない場合に指摘する。
- 漢字含有率は全体で 30% ~ 40% が望ましいとされている。よって、この範囲外であるならば指摘する。
- 日本語の場合、接続詞を使用することで文と文をつなぎ、長い文にすることができるが、あまりにも長過ぎると、文の構造が複雑になり、意味がわかりづらくなる。一般的には 50 文字程度、長くて 120 ~ 130 文字程度であり、やむを得ない場合には 150 文字前後に留めるべきだとされている。そこで、150 文字を超えた文がある場合には、120 ~ 130 文字程度におさえるよう指摘する。

なお、次節以降は、各文ごとに対してのチェックについて説明する。

4.2 不適切な表現のチェック

技術文章として、話言葉のような不適切な表現は避けるべきである。例えば「ありゃ、～」や語尾の「～かしら」「～ぜ」など、表 2 のような品詞が対象として挙げられる。そこで、これらの品詞を持つ場合には、不適切な表現を利用しているという指摘を与えることとした。

表 2 不適切な表現に該当する品詞

品詞	説明	例
[名詞-代名詞-縮約]	代名詞の口語表現	ありゃ、こりゃ
[接頭詞-動詞接続]	動詞の命令形の接頭語	お(読みなさい)
[接頭詞-形容詞接続]	形容詞の接頭語	お(寒い)
[助詞-終助詞]	文の終わりの働きかけ	(行こう) ぜ、(寒い) ね
[感動詞]	あいさつなど	おはよう、まあ
[フィラー]	話し言葉のあいづちなど	あの、えーと

4.3 半角と全角のチェック

多くの論文の執筆要綱において、英数字については半角が指定されているため、アラビア数字と英字の大文字・小文字にお

いて、全角が使用されている場合は指摘することとした。また、カタカナについては、半角カタカナはコンピュータで扱われる際に文字化けの原因ともなるので、これについては半角が使用されている場合に指摘することとした。

4.4 文末の口語文章体のチェック

技術文章においては、会話体である「です・ます調」ではなく、口語文章体の「である調」を用いるべきである。そこで、表3に従って文末表現のチェックを行い、不適切と判断した場合に指摘することとした。

表3 不適切な文末表現に該当する品詞

品詞	説明	例
[助動詞 五段・ラ行特殊]	断定の助動詞「ござる」	(忍者で) ござる
[助動詞 特殊・デス]	断定の助動詞「です」	(無理) です
[助動詞 特殊・ジャ]	断定の助動詞「じゃ」	(行くの) じゃ
[助動詞 特殊・マス]	譲渡・丁寧の助動詞「ます」	(行き) ます
[助動詞 特殊・ヤ]	断定の助動詞(関西方言)「や」	(そう) や

4.5 当て字のチェック

当て字は、字本来の用法を無視し、異なる語の表記に転用した文字を指している。一例としては、文化庁における常用漢字表[13]の付表に含まれており、これに参考文献[11][12]と合わせて考慮し、技術文章で使用されるであろう当て字を表4のように定義した。これらが各文における形態素の基本形に該当するならば、指摘することとした。ただし、副詞に該当するものは後述の「4.10 一部の副詞の漢字表記化と平かな表記化」において処理されるため、ここでは省いてある。

表4 当て字として判定する単語

単語	読み	単語	読み	単語	読み
灰汁	あく	湿気る	しける	吹雪く	ふぶく
如何	いかが	尻尾	しっぽ	不味い	まずい
色々	いろいろ	其の	その	真似る	まねる
上手い	うまい	従って	したがって	真面目に	まじめに
夥しい	おびただしい	仕舞う	しまう	浮腫み	むくみ
美味しい	おいしい	素早い	すばやい	滅茶苦茶	めっちゃくちゃ
鍛冶	かじ	素晴らしい	すばらしい	鍍金	めっき
瓦斯	ガス	容易い	たやすい	減り張り	めりはり
硝子	ガラス	序でに	ついでに	貰う	もらう
肌理	きめ	出来る	できる	易い	やすい
此处	ここ	出鱈目	でたらめ	所以	ゆえん
如く	ごとく	何処	どこ	を以て	をもって
誤魔化す	ごまかす	流行る	はやる		
流石	さすが	相応しい	ふさわしい		

4.6 常用漢字のチェック

常用漢字は、一般の社会生活で用いられる目安として1981年に内閣告示された1945文字であり、論文の多くの執筆要綱でも、常用漢字を使用するよう記載されている。しかし、実際に文章を書く際に常に1945文字を覚えておくことは容易ではなく、機械的なチェックに頼るべき点である。ただし、常用漢字以外でも人名に用いられる場合は例外であるため、品詞が[名詞-固有名詞]に該当する場合は、その旨も合わせて指摘することとした。[名詞-固有名詞]は、人名や地域などの固有名詞が用いられている場合に、chasenによって付与される品詞である。

4.7 一人称のチェック

通常、技術文章では著者自身のことを「私」や「僕」といった表現はすべきではない。文中で「私」や「僕」が使用される可能性が高いのは、主語の位置と思われるが、場合によっては目的語などにも使用されることがあるため、文中での位置に関係なく形態素レベルでチェックを行い、「私」や「僕」が使用されている場合には、「著者」や「著者ら」を使用するか、一人称がなくても通じる文への書き換えを行うよう指示することとした。

4.8 形式名詞の平かな表記化

形式名詞は、実質的な意味を持たず、常に連体修飾語を受けて用いられる名詞で、例えば「(解明した)事(が)」や「(扱ってはいけぬ)物(である)」などの「事」や「物」が該当する。これらは、漢字そのものに意味があるわけではないため、通常平かなで表記すべきである。そこで、形式名詞に対しては平かな表記にするように指摘するが、形態素解析結果の一意な品詞情報から単純に形式名詞を判断できないため、表5のような形式名詞ごとの判断基準を作成した。

表5 平かな表記化する形式名詞とその品詞

[名詞-非自立-一般]	訳, 物, 方, 筈, 積もり, 毎
[名詞-非自立-副詞可能]	所, 時, 通り, 度, 限り, 折り, 上
[助詞-副助詞]	程, 等, 位
[名詞-非自立-助動詞語幹]	様

4.9 接続詞の平かな表記化

接続詞は、前後の単語や文節、文同士を接続するための語である。技術文章では特に漢字表記の単語が多い傾向にあるため、単語同士を接続する場合に接続詞まで漢字で記述してしまうと、非常に分かりづらい。例えば、「頻出木及び縮約木では～」の場合、「頻出木および縮約木では～」とした方が読みやすい。そこで、品詞として[接続詞]として判断されたものが漢字になっていた場合は、平かな表記するよう指摘することとした。

4.10 一部の副詞の漢字表記化と平かな表記化

副詞とは、それだけで意味を表し、文節を作ることが可能で、活用がなく、後ろの用言を修飾する語であるが、漢字表記すべきか、平かな表記すべきかについて、日本語文法的に定まっているわけではない。しかし、文献[11]によると、漢字表記する基準は、読者の素早い理解を助けるために漢字の意味を保持しているものとし、平かな表記の基準としては逆に漢字の意味を保持しておらず、読者が見た時にわかりづらいものとしている。平かな表記の基準として具体的には、以下の3点が挙げられている。

- (1) 常用漢字にない音訓読みをするとき
- (2) 常用漢字以外の漢字が使用されるとき
- (3) 当て字で使用されるとき

そこで、上記の基準に従い、漢字表記化と平かな表記化対象の副詞として、品詞が[副詞-一般]または[名詞-副詞可能]、[副詞-助詞類接続]のいずれかであり、表記が表6もしくは表7に定めたものに該当するならば指摘することとした。

表6 漢字表記化対象の副詞

単語	読み	単語	読み	単語	読み
一応	いちおう	徐々に	じょじょに	特別	とくべつ
一概に	いちがい	少し	すこし	突然	とつぜん
一段と	いちだんと	全然	ぜんぜん	残らず	のこらず
一度に	いちどに	全部	ぜんぶ	後ほど	のちほど
概して	がいして	早々	そうそう	初めて	はじめて
必ず	かならず	絶えず	たえず	再び	ふたたび
必ずしも	かならずしも	確か	たしか	万一	まんいち
偶然	ぐうぜん	多分	たぶん	最も	もっとも
早速	さっそく	単に	たんに	要するに	ようするに
終始	しゅうし	当然	とうぜん		
十分	じゅうぶん	特に	とくに		

表7 平かな表記化対象の副詞

単語	読み	単語	読み	単語	読み
敢えて	あえて	暫く	しばらく	何故	なぜ
予め	あらかじめ	洪々	しぶしぶ	一先ず	ひとまず
如何にも	いかに	所詮	しょせん	一際	ひときわ
幾らか	いくらか	随分	ずいぶん	先ず	まず
些か	いささか	既に	すでに	益々	ますます
徒に	いたづらに	即ち	すなわち	真っ直ぐ	まっすぐ
一旦	いったん	全て	すべて	寧ろ	むしろ
今更	いまさら	総て	すべて	滅多に	めったに
概ね	おおむね	凡て	すべて	若しも	もしも
凡そ	おおよそ	大して	たいして	勿論	もちろん
却って	かえって	沢山	たくさん	専ら	もっぱら
且つ	かつ	唯	ただ	最早	もはや
嘗て	かつて	忽ち	たちまち	漸く	ようやく
殊更	ことさら	丁度	ちょうど	余程	よほど
然も	しかも	一寸	ちょっと	宜しく	よろしく
然るに	しかるに	遂に	ついに	態と	わざと
頻りに	しきりに	尚	なお	態態	わざわざ

5. 文章推敲支援

文章作成のポイントの1つに、短文が良いのか長文が良いのかという課題がある。誤りなく簡潔に伝えるためには短文が良いとされるが、リズムカルな展開や、より複雑な状況・内容を説明するためには長文も用いることが求められる。

その結果、文節間の係り受け関係、修飾関係をもとにした意味解釈の正確さが求められる。特に、レポート、マニュアル等においては、複数の曖昧な解釈が成り立つような文章表現は避けるべきであり、この係り受け関係、修飾関係が、執筆者の意図に合致しているか否かのチェック、さらに改善案検討のための支援機能が求められる。

そこで、係り受け関係に基づいた文構造を活用することとした。すなわち、係り受け関係の木表現化、それに基づいた文の骨格表現から段階的な修飾関係の明示、さらに文の修正による構造の変化を明示することにより、何が解釈の違いを導くのかについて検討することを可能としている。

表8 曖昧な文章と改善した文章例

例文1	このシステムでなぜ精度が低下する従来の測定器を卓越できるのかを考察する。
例文2	精度が低下する従来の測定器をこのシステムでなぜ卓越できるのかを考察する。
例文3	精度が低下する従来の測定器をなぜ卓越できるのかをこのシステムで考察する。

例えば、表8の例文1では、「このシステムで卓越できるの

か」ということを「考察する」のか、「このシステムで考察する」のか、判断できない曖昧な表現になっている。これは、修飾語と非修飾語の距離が非常に離れており、複数に係る危険性を増加させてしまっていることに原因がある。例文2は、「このシステムで卓越できるのか」という意図になるよう修正した文であり、例文3は、「このシステムで考察する」という意図になるよう修正した文である。一般的に、複数の解釈可能な曖昧表現を避ける方法として、修飾語と非修飾語の距離を近づけることや、1つの非修飾語に対して複数の修飾語が与えられる場合には、長い修飾節順に並べると良いとされている。

5.1 木構造のグラフィカル表示

図4は、これら例文1および例文2について、係り受け関係を木構造としてグラフィカルに表現したものである。係り受けの最後の用言をルートとし、葉からルートに向かって係り受け関係を表し、左から右へ文章を記述した順を表している。この図から、例文1は長い修飾節が右に寄っており、紛らわしい文である可能性を確認できる。

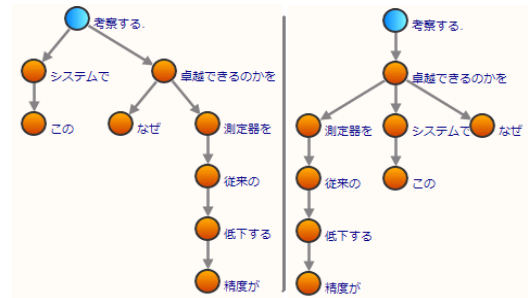


図4 例文1と例文2の木構造化

また、このグラフィカルな表現に加え、3種類の確認方法によって、これら係り受け関係をテキストベースで提示する。

5.2 2文節の係り受け関係の提示

係り受け関係にあるすべての2文節を一目で確認できるように提示している。これにより、どのあたりに不自然な係り受け関係があるかをすばやく把握することが可能となっている。

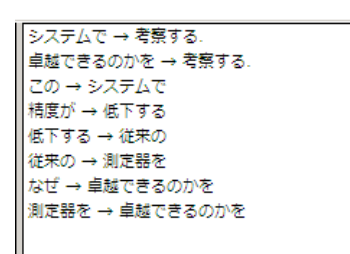


図5 2文節の係り受け関係の提示

5.3 段階的な修飾関係の提示

図4のような木表現を行うと、文の骨格表現が、ルートからの段階的な深さによって確認できることがわかる。例えば、例文1であると、基本骨格がルートの「考察する」であり、その基本骨格を残りの単語で修飾している。深さ2で見ると、「システムで考察する」と「卓越できるのかを考察する」の2文の骨格構造に対して残りの単語で修飾している。このように、文は

骨格に対して修飾の連続で成り立っていると見ることができる。そこで、図 6(例文 1 の結果)のように、木構造の深さによる骨格構造を段階的に提示することで修飾関係を容易に解釈できるようにした。また、例文 2 の意図のように「システムで考察する」わけではない場合には、この深さレベル 2 を確認するだけで、意図の間違いに気づくことも可能である。

この段階的な修飾関係の提示は、木構造のルートだけでなく、複数の修飾語を受ける被修飾語についても表示し確認できるようにしている。

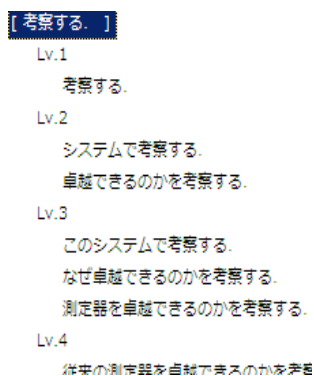


図 6 段階的な修飾関係の提示

5.4 段階的な修飾関係による文表現の提示

前節でも述べたように、文は骨格構造に対して、修飾しながら長くなっていく。そこで、図 7 のように、木構造における各深さレベルの状態ごとの文を表示する。例えば、深さレベル 2 の状態では、図 6 で示したように、骨格として「システムで考察する」と「卓越できるのかを考察する」の 2 文となるので、これらを結合し、「システムで卓越できるのかを考察する」といった文を提示する。これにより、骨格に対して修飾という肉付けをうまく行えているかを随時確認しながら、表現豊かな文を繰り返し創り上げていく支援が可能となる。

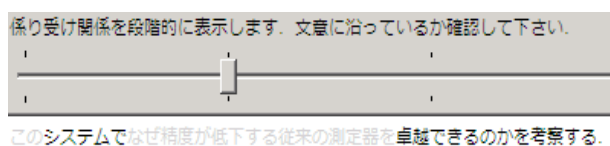


図 7 段階的な修飾関係による文表現の提示

6. ま と め

本研究では、レポート、マニュアルおよび技術系の論文など技術文章を対象として、執筆活動を通しながら学習するための支援ツールのプロトタイプを開発した。

本ツールは、一般的な校正内容を体験しながら身につけると、曖昧な文章を視覚的に理解し、改善するための方法を自ら発見し、誤解を与えない文を作成できることをねらいとした。そして、そのために校正支援機能と推敲支援機能の 2 側面を持ち、校正支援機能は、ルールベースの拡充により実現し、推敲支援機能は、係り受け関係を構造化および木表現を行い、文の段階的表示や構造情報をグラフィカルに提示することで、視覚

的な理解を補佐し、考えさせることを実現した。

係り受け解析ツールの解析結果は、100%の正解率を出すわけではないが、一般的な解釈、つまり多くの人が解釈するであろう係り受け結果と考えられるため、逆に執筆した文が思い通りの係り受け関係が得られていないならば、それはわかりづらい文である可能性が高いと考えられるので、本提案ツールの誤解を与えない文を作成するという趣旨に合致しており、高い効果を期待できると考えている。

今後は、校正支援機能の面は、さらなるルールベースの拡充を行う予定である。また、推敲支援機能としては、リズムカルな文章の作成支援や文間の接続詞による視覚的な文脈理解の支援、句読点位置による文意の変化についての理解支援などに加え、文章の木構造データを蓄積し、執筆者の文章の特徴抽出や文章作成技術が長けている者との比較などを行うことを考えている。

文 献

- [1] 岡山大学, 日本語力の徹底訓練による発想型技術者育成, <http://www.okayama-u.ac.jp/jp/pdf/GP.pdf> .
- [2] 浅原正幸、松本裕治, ipadic version 2.6.3 ユーザーズマニュアル, <http://chasen.aist-nara.ac.jp/chasen/manual.html.ja>, 2003 .
- [3] 菅沼明、石田朗子、倉田昌典、牛島和夫, 日本語文章推敲支援ツール『推敲』における字面解析手法とその評価, 情報処理学会自然言語処理研究会, 68-8, 1988.9 .
- [4] 下園幸一、菅沼明、牛島和夫, 日本語文章推敲支援ツール『推敲』における助詞「は」と「が」の抽出について, 情報処理学会自然言語処理研究会, 94-6, 1993.3 .
- [5] 菅沼明、牛島和夫, 字面解析を応用した日本語文章推敲支援ツールの開発, 情報処理学会デジタル・ドキュメント研究会, 5-1, 1997.1 .
- [6] 奥村薫、建石由佳、脇田早紀子、金子宏, 日本語校正支援システム「FleCS」, 情報処理学会自然言語処理研究会, 87-11, 1992.1 .
- [7] 奥村薫、脇田早紀子、金子宏, 日本語校正支援システム FleCS の新聞社における実用化, 情報処理学会自然言語処理研究会, 91-5, 1992.9 .
- [8] 脇田早紀子、奥村薫, 日本語校正支援システム FleCS - ミスタイプ検出について, 情報処理学会自然言語処理研究会, 97-19, 1993.9 .
- [9] 工藤拓、松本裕治, チャンキングの段階適用による係り受け解析, 情報処理学会論文誌, Vol 43 No. 6, pp.1834-1842, June 2002 .
- [10] 松本裕治, 形態素解析システム「茶釜」, 情報処理 Vol.41 No.11, pp.1208-1214, November 2000 .
- [11] 中島利勝、塚本真也, 知的な科学・技術文章の書き方, コロナ社, 1996 .
- [12] 八木和久, 例題で学ぶ原稿の書き方, 米田出版, 2004 .
- [13] 文化庁, 国語施策情報システム, <http://www.bunka.go.jp/kokugo/>, 国語表記の基準 .
- [14] 井上勝也, 科学表現-基本と演習-, 培風館, 1992 .
- [15] 山村広臣、菅沼明、牛島和夫, 日本語文における名詞句の並列構造の推定および推敲支援への適用, 情報処理学会自然言語処理研究会, 111-2, 1996.1 .
- [16] 牛島和夫、日並順二, 日本語文章推敲支援ツールの試作とその作成環境, 情報処理学会ソフトウェア工学研究会, 35-2, 1984.5 .
- [17] 牛島和夫、日並順二、尹志熙、高木利久, 日本語文章推敲支援ツールのプロトタイプング, 情報処理学会ソフトウェア工学研究会, 40-8, 1985.2 .
- [18] 箱森聡、佐川雄二、大西昇、杉江昇, 日本語の修飾構造を評価する添削支援システムを実現するための基礎研究, 情報処理学会論文誌, Vol.33 No.2 pp.153-161, 1992 .
- [19] 阿辺川武、八木豊、戸次徳久、傳亮、Slaven Bilacy、奥村学、仁科喜久子, 読解支援システムのための言語非依存フレームワーク構築, 言語処理学会大 10 回年次大会発表論文集, pp.144-147, 2004 .