

日本語を簡単な日本語に言い換える

— 福祉的コミュニケーション支援に向けて —

乾 健太郎

inui@ai.kyutech.ac.jp

1 背景と目的

計算機ネットワークの普及と電子化テキストの爆発的増加に伴い、多様で大量の情報が個人でも収集できるようになってきた。その一方で、聴覚障害や視覚障害といった障害によってこの恩恵を十分に享受できない人々は、いわゆる情報弱者としてますます情報の流通から疎外される傾向にある。たとえば、手話言語を母語とする先天的聾者の中には、日本語などの音声言語（自然言語）を使用する訓練が不十分なために自然言語テキストの読解能力が不十分な人が少なくない。計算機ネットワーク上を流れる情報の多くが自然言語テキストである現状を考えると、こうした人々に対する情報保障が急務の課題であると言える。

このような背景から、本研究では、上述のような読者を対象とした読解支援の要素技術となるテキスト単純化技術に関する基礎的研究を行った [7, 11, 12]。テキスト単純化技術とは、一般のテキストに構文的・語彙的変換操作を適用することにより、これをより平易なテキストに言い換える技術である。

本研究で取り組む言い換え技術は、上述の応用にとどまらず、様々な応用の可能性を秘めている。たとえば、テキスト読み上げによる視覚障害者向け読解支援では、入力テキストを言語レベルで単純化しておくことで、出力音声の理解容易性が向上する。テキストから手話に翻訳する場合も同様である。これら福祉的コミュニケーション支援の分野はもとより、翻訳や要約、推敲支援といった応用においても、言い換え技術の導入は有用であると期待されている。しかしながら、このような背景があるにもかかわらず、これまで言い換えに関する研究は少なく、その全体像はほとんど解明されていない。

読解支援を目的とするテキスト単純化では、原文のままでもユーザが理解できる部分は言い換える必要がないし、むしろ言い換えるべきでない。一方、可読性が低い部分を言い換える場合には、その言い換えが可読性を向上させるという目的に合うものでなければならない。このように、テキスト単純化は、「原文のどの部分に問題があり、それを解決するにはどのように言い換えればよいか」という判断を行う問題検出・解決型のタスクと見なすことができる。

以上より、本研究では次の課題に取り組んだ。

- **技術的課題**：テキスト単純化という事例研究を通して言い換えのメカニズムに対する理解を深める。具体的には、次の2種類の技術に関する基礎を確立する。
 - **言い換え生成**：与えられた言語表現に対して、言語的に適格な種々の言い換えを網羅的に生成する技術
 - **テキスト評価**：与えられた評価基準に基づいてテキストを評価する技術
- **社会的課題**：テキスト単純化による読解支援という新しいアプリケーションを開拓する。具体的には、対象とするユーザおよびその支援者に技術可能性を広報するとともに、アプリケーションのあり方を共同で模索する。

2 研究成果

本研究の成果は次の4点に要約できる。

1. 大規模な言い換え生成実験によって言い換え事例を効率的に蓄積する環境を整備した。
2. 言い換えるタイプごとにそれぞれ事例研究を行い、実験を通じて言い換えの問題の性質を部分的に明らかにした。
3. 聾学校教諭を対象に大規模なアンケート調査を実施し、回答データから文の可読性を評価する計算モデルを構築できることを示した。

4. 読解支援プロトタイプシステムを作成し、「テキスト簡単化による読解支援」のイメージを具現化した。
以下、それぞれについて概要を述べる。

2.1 言い換え生成実験環境 KURA の構築

言い換えは、テキストの翻訳の一種と見なせるので、その実現方法は、従来から広く研究されてきた異言語間機械翻訳、とくに構造変換（トランスファ）方式の翻訳に範を求めるのが自然であろう。ただし、同一言語内翻訳である言い換えには、次のような異言語間翻訳にない性質もある。

- 一つの単語の言い換えから文の分割・併合まで、一つひとつの構造変換がそれ単体で一つの言い換えを構成し得る。
- したがって、多くの場合原文の大部分が言い換え後も保存される。
- 態の変換や文の分割のように原文の統語構造を特定の構造に変換すること自体が言い換しの動機になる場合も多く、構造変換において表層や構文の情報を過度に捨象する抽象化は望ましくない。

言い換え固有の性質や問題はまたこの他にもあるかもしれない。言い換しの研究を着実に進めるためには、そういった問題の全体像を明らかにする必要があるが、本研究開始時には、そもそもそれを調べるためのソフトウェアツールさえ存在しないのが現状であった。

このような背景から、本研究では次のような特長を持つ言い換え生成実験環境／言い換えエンジン KURA を整備した [15, 20, 21] (図 1 参照)。

Lexico-structural paraphrasing: システムは、語彙・構文的言い換えを扱う。語彙・構文的言い換えは、たとえば次のような言い換えである。

- 副詞節や並列節を含む長文の分割:
ようやく回復の足取りを固めたとはいえ、日本経済の先行きは不透明感を免れない。⇒日本経済は、ようやく回復の足取りを固めた。しかし、先行きは不透明感を免れない。
- 連体修飾節の言い換え（主節化）:
きっかけは中村屋がインド独立運動の闘士であったビハリ・ボースをかくまうようになったことである。⇒ きっかけは中村屋がビハリ・ボースをかくまうようになったことである。ビハリ・ボースはインド独立運動の闘士であった。
- 複雑な否定表現の言い換え:
彼はベジタリアンなので、パーティーでも野菜しか食べなかった。⇒ 彼はベジタリアンなので、パーティーでも食べたのは野菜だけだった。
- 複雑な比較表現の言い換え:
学生時代の友人と飲む酒ほど楽しい酒はない。⇒ 学生時代の友人と飲む酒が一番楽しい。
- 受動・使役表現の言い換え:
太郎は駅前で警官に呼びとめられた。⇒ 警官が駅前で太郎を呼びとめた。
- 機能語相当表現の言い換え:
たばこをやめてからというもの、食欲が出て体の調子がとてもいい。⇒ たばこをやめてから後ずっと、食欲が出て体の調子がとてもいい。
- 慣用句などの言い換え:
学生時代は毎日のように映画館に足を運んだ。⇒ 学生時代は毎日のように映画館に行った。
- 難しい内容語／句の言い換え:
多数意見は立法府（⇒国会）の裁量判断を尊重する立場を示した。
足を運ぶ ⇒ 行く。

Three-layered architecture: 言語的に適切な言い換えを網羅的に生成する処理と可読性評価などアプリケーションに依存する処理を独立に開発できる。

Revision-based morpho-syntactic transformation: 構造変換規則の複雑さを抑えるために、変換規則に非決定性を持たせるとともに、適用条件や変換処理の記述が不完全 (underspecified) な規則を許す。これにより、変換規則が過度に複雑になるのを防ぐことができる。

変換規則が非決定的あるいは不完全な場合、変換結果の言語的適格性は必ずしも保証されない。そこで、変換規則の適用によって生じる不適格な変換結果は、変換後に「言語モデル」の知識に基づいて修正、棄却、あるいはランクづけすることによって吸収する (図 1 の linguistic transformation layer)。これにより、知識のモジュール性が増し、同じ知識が様々な言い換えで利用可能になると期待できる。このことにより、システム全体の知識開発のコストを抑えることができる。

Data-driven control of semantic analysis: 構造変換あるいは選択・修正でどのような構文・意味情報が必要になるかは、処理の開始時点では予測できない。そこで、入力構文・意味情報は、それが必要になった時点でオンデマンドに解析・収集する。これにより、「いつどの部分の何に関する意味解析をするか」という制御を自然に最適化することができる。

Declarative and uniform knowledge representation: 変換規則および言語モデルからなる言い換え知識を図 2 に示すような統一かつ宣言的な規則として記述できる。また、高い抽象度で規則が記述できるため、システム内部に必ずしも精通しない言語の専門家に規則記述をアウトソーシングすることが可能である。

Process efficiency: 大規模な言い換え実験に耐えるように種々の高速化の工夫が施されている。

GUI for process monitoring: 実験結果をすべてリレーショナル・データベース上で管理し、図 4 に示すような誤り分析や規則開発を効率化する環境を提供する。

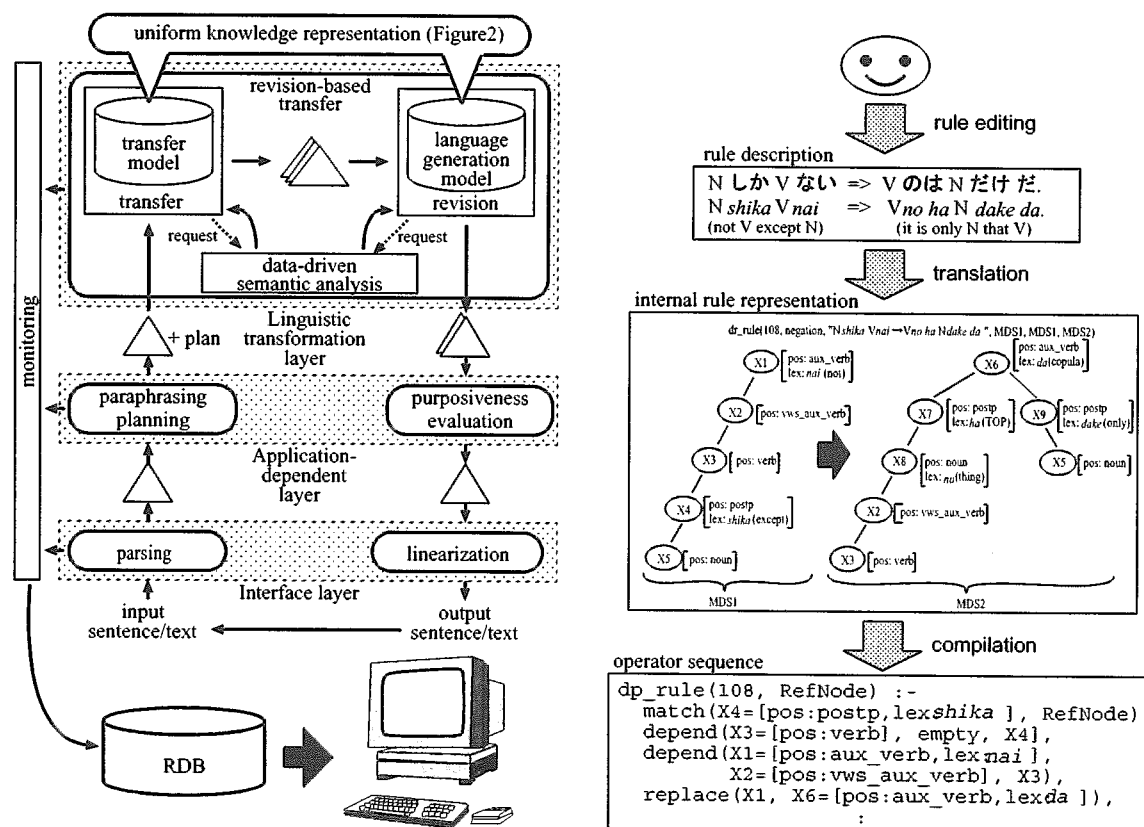


図 1: 言い換え生成実験環境

本システムの開発に当たっては、本研究で次第に明らかになった言い換え生成に関する知見に基づき、本研究の期間を通して何度も設計・実装を繰り返した。最終的に辿りついた現在のシステムには、言い換え知識の蓄積・共有、言い換え事例の蓄積を効率的に行うための様々な工夫が施されており、その設計上の知見は本研究の主要な成果の一つである。本システムは、次のアドレスでソースとともに公開している。

www.pluto.ai.kyutech.ac.jp/plt/inui-lab

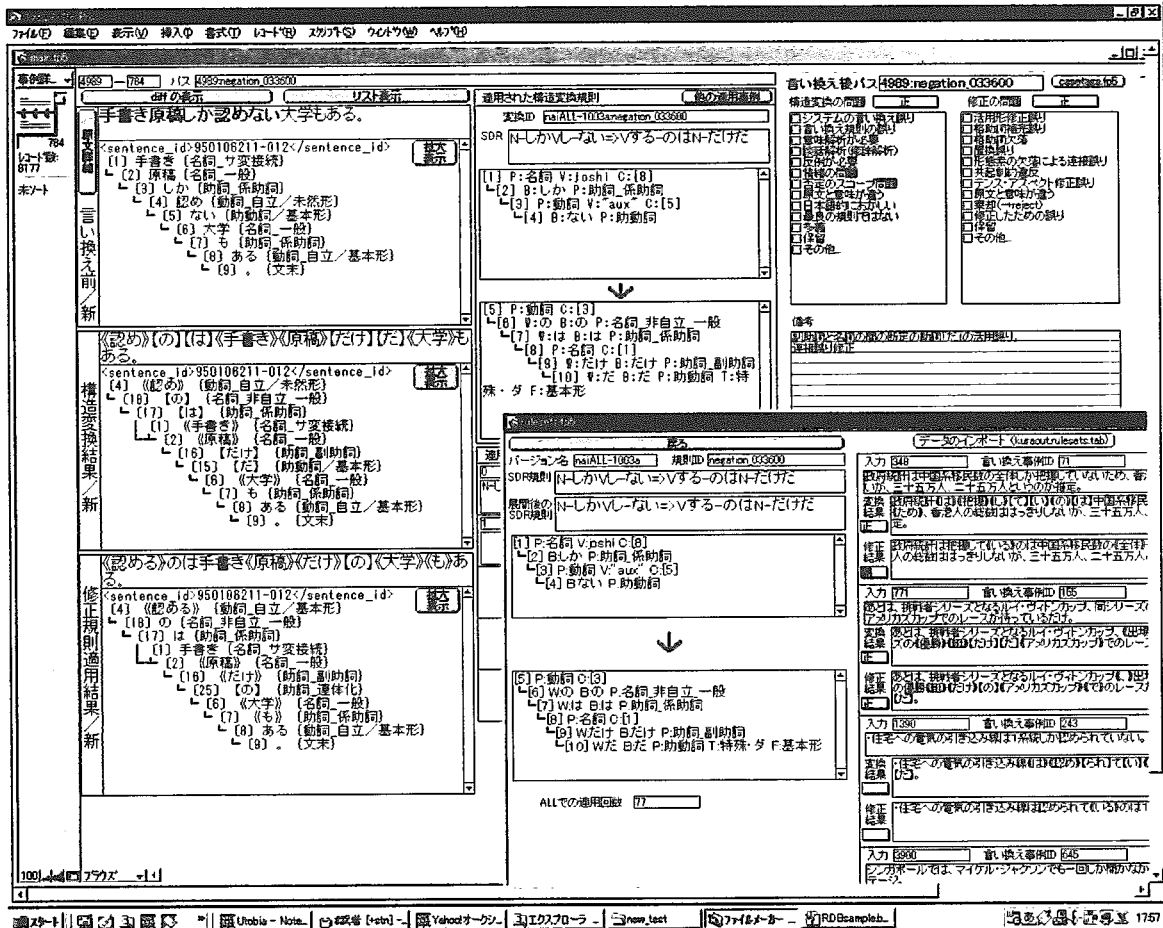


図 3: RDB による言い換え生成結果の管理

2.2 言い換え生成の個別的事例研究

言い換え生成の問題は、(a) 変換パタンの多様性と獲得コスト、(b) 言い換えによって変化する意味の性質の違い、の観点から性質の異なる複数の問題の複合であることが本研究を通じてわかってきた。本研究でとったアプローチは、次のように要約できる。

1. 言い換えの問題は多様な性質の部分問題の複合であり、十分に複雑であるため、単純に言い換えるの訓練事例から必要な知識を組み合わせを直接的に自動獲得しようとするアプローチではうまくいかないと考えた。
2. そこでまず、問題の性質が異なると考えられる言い換えるの種類ごとに、人手で典型的な言い換えパターンを記述し、前述の言い換え実験環境を用いて、大規模なコーパス上でその振舞いを分析する。同時に言い換えるの事例を蓄積する。

3. 事例の分析を通じて問題の性質（全体の問題を構成する部分問題）を明らかにするとともに、個々の部分問題に必要な知識を類型化・構造化する。
4. 言い換え知識の種類ごとにそれに合った知識獲得の方法を選択し、個別に知識の構築・大規模化をはかる。

具体的には、以下の4種類の言い換えに個別的に取り組んだ。

2.2.1 節間構造レベルの言い換え

典型的には、連体節や連用節を含む複文を分割したり、複数の文を一文に統合したりといった、節間の組み合わせ方を変更するレベルの言い換えである。この種の言い換えは一般にテキストの結束構造を変化させるので、どのようにして言い換え後の結束性を保証するかという課題に取り組んだ [13, 19]。たとえば、次のテキスト (1a) の第1文は、(b1) や (b2) のように言い換えても結束性は維持されるが、(b3) のように言い換えると後続文脈との繋がりが悪くなってしまふ。

- (1) a. スウェーデンの首都ストックホルムから南西部に位置するスモーランド地方は別名「ガラスの王国」とも呼ばれている。この地方にある二つの大きな町カルマルからベクショーにかけて、十六ものガラス工場が点在しているからだ。
- b1. スモーランド地方は、スウェーデンの首都ストックホルムから南西部に位置し、別名「ガラスの王国」とも呼ばれている。この地方にある二つの...ガラス工場が点在しているからだ。
- b2. スモーランド地方はスウェーデンの首都ストックホルムから南西部に位置する。この地方は別名「ガラスの王国」とも呼ばれている。この地方にある二つの...ガラス工場が点在しているからだ。
- b3. *スモーランド地方は別名「ガラスの王国」とも呼ばれている。スウェーデンの首都ストックホルムから南西部に位置する。この地方にある二つの...ガラス工場が点在しているからだ。

この問題は、(a) 構文的・意味的に適格、かつ多様な言い換え候補を並列に生成する問題と (b) 生成された言い換え候補を結束性の良さの観点から評価する問題に分けることができる。我々は、まず (i) の言い換えエンジンを構築し、構文・意味レベルの適格性が保証された多様な言い換え候補を生成できる環境を整えた。(ii) については、まず、コーパス中の連体修飾節に対し、言い換えエンジンを使って多様な言い換え事例をシステムティックに生成し、これに人手による結束性評価を加えたタグつき事例集合を作成した。現在の事例集合の規模は、正例 497 件、負例 1007 件である。まだまだ不十分な規模だが、それでも文献 [13, 19] で報告したように、結束性の具体的な評価基準について興味深い知見が得られつつある。また、これらの作業と並行して、照応・省略表現の適格性の評価および選択に焦点を絞った研究も進めており [16]、今後は両者の知見の統合をはかるとともに、結束性評価モデルの自動学習と訓練事例の選択的収集を統合する枠組みの構築を目指す。

2.2.2 節内構造レベルの言い換え

節内構造レベルの言い換える例題として、次の例のような、否定と副助詞が呼応する表現の言い換えを取り上げた。これらの言い換えは、主辞交替を伴う言い換える代表的なものであり、副詞と否定の呼応、副助詞と格助詞の相互作用、動詞の格交替といった様々な言語現象を引き起こす点で興味深い例題である。

- (2) a. りんごしか食べない→りんごだけ食べる（副助詞が否定「ない」を含む述語に係る）
- b. お酒を飲まない人は彼だけだ→彼以外は全員がお酒を飲む（否定「ない」を含む関係節の被修飾名詞が副助詞を伴う述語に係る）
- c. 彼ほどお酒を飲む人はない→彼が一番お酒を飲む（副助詞を含む関係節の被修飾名詞が否定「ない」を含む述語に係る）

具体的には、新聞記事例文集約 4 万文を対象に、前述の実験環境 KURA を利用して、手作業で次の例のような言い換えパターンを実装することを試みた [3]。

- (3) a. N- だけ V- する → N- 以外は V- する.
 b. N- しか V- しない → V- するのは N- だけ.

現時点で、構造変換規則約 800 件、修正規則約 150 件を実装し、約 4 万の例文に対して、正例・負例を含む約 8,600 件の言い換え事例を蓄積している。これまでの事例分析から、少くとも次の 2 点が明らかになった。

- この種の言い換えでは、多義性の問題に起因する言い換え誤りはそれほど頻繁には起こらない。さらに、多義性による誤りが生じるパターン（たとえば、「～だけでない」）の数は、個別的な対処が可能な程度に限定されている。この点は、機械翻訳における内容語の訳語選択の問題とは大きく異なる。
- 構造変換規則は入力文の一部の部分木を書き換えるものであるが、このとき書き換え部分とその周囲（典型的には、修飾要素やテンス・アスペクト表現）との整合性が、書き換えによって損われるケースが目立った。これは、次のような修正規則の拡充によって効果的に対処できる可能性が高い。

- (4) a. ほどだけ V- する → ほど V- するだけ.
 b. N1- だけ N2 → N1- だけという N2.

2.2.3 機能語相当表現レベルの言い換え

機能語相当表現とは、いわゆる「表現文型」という術語で一括りにされる助詞・助動詞相当表現である。このレベルでの言い換えは節内・節間レベルの言い換えに比べ語彙的な性格が強い。これらの表現の中には「涙ながらに→涙を流しながら」、「～しがてら→～するのを兼ねて」、「～のきらいがある→～の傾向がある」のように局所的な情報を参照するだけで言い換えられるものも多い。

このレベルの言い換えについては、まず「日本語表現文型言い換え辞書」の構築に取り組んだ。具体的には、「日本語能力試験 1・2 級の機能語の類のリスト」に準拠した 3 種類の日本語教材に掲載されている解説や例文をもとに、機能語相当表現言い換え辞書の作成を進めた。すでに、機能語相当表現 323 項目について、985 個の言い換えパターンと約 2,600 件の例文に対する言い換え事例約 3,800 件（正例 3,314 件、負例 500 件）が収集できている。この種の辞書の構築はこれまでに例がなく、読解支援だけでなく種々の言語処理に有用な知識になると期待できる。

次に、これらのうち「1 級」相当の表現について、279 個の言い換え規則と 53 個の修正規則を KURA に実装し、上記辞書中の例文と新聞記事例文集を対象に言い換え実験を行った [3]。これまでのところ、辞書中の例文（クローズドテスト）について再現率、適合率ともに 92%、新聞記事例文集について適合率 39% という結果が得られている。事例分析の結果、前述の否定表現の言い換えの場合とほぼ同様の性質を持っていることがわかった。

2.2.4 内容語レベルの言い換え

内容語間の言い換え、あるいは内容語の組み合わせによる慣用的表現の言い換えのレベル。このレベルの変換パターンは個々の内容語ごとに記述する必要があるため、その獲得に当たっては、パラレルコーパスのアライメントも含め、シソーラスや対訳辞書などの既存の資源を最大限に利用する手段を検討すべきである。具体的には、次の 2 つの課題に取り組んだ。

- 普通名詞の同概念語への言い換えの可否を判定するというタスクを取り上げ、(a) 単語間共起制約、(b) 語釈文の比較により抽出される単語対の共通 / 個別特徴、(c) 共通 / 個別特徴と文脈の重なり、といった手がかりを利用してこれを解決する試みを進めた [1, 2]。
- 「足を運ぶ ⇒ 行く / 通う」のような慣用的な内容語相当表現の言い換え知識を、(a) 対訳つき格フレーム辞書、(b) verb alternation、(c) 語釈文、などを組み合わせることによって安価に獲得する方法について検討し、いくつかの予備実験を行った。現在、人手で作成した格交替規則 40 個、国語辞典から半自動的に作成・洗練した慣用句言い換え規則約 4,100 個、機械翻訳用対訳辞書から抽出・洗練した動詞格フレーム言い換え規則約 4,000 個を KURA 上に実装し、その振舞いを調べる実験を進めている。

	A	B	C	D	E	F	G
400	～に(は)あたらぬ						
401	交換規則	動詞+(は)あたらぬ	動詞+には及ばない		動詞+必要はない		動詞+ほどのことではない
402		彼の實力を考えると、今回の彼の受賞も驚くにあたらぬ。	彼の實力を考えると、今回の彼の受賞も驚くには及ばない。		彼の實力を考えると、今回の彼の受賞も驚く必要はない。		彼の實力を考えると、今回の彼の受賞も驚くほどのことではない。
403	～に即して「た」						
404	交換規則	名詞+に即して	名詞+に従って		名詞+に応じて		
405	選択制限	違反者は、法律に即して処分されます。	違反者は、法律に従って処分されます。		*違反者は、法律に応じて処分されます。	「法律」はルールを表す	
406		実情に即して対応策を検討いたします。	*実情に従って対応策を検討いたします。て	「実情」はルールを表すのではない	実情に応じて対応策を検討いたします。		
407							
408	交換規則	名詞+に即した	名詞+に従った		名詞+に応じた		
409	選択制限	実情に即した対応策を検討いたします。	*実情に従った対応策を検討いたします。	「実情」はルールを表すのではない	実情に応じた対応策を検討いたします。		
410		非常事態でも、人道に即した行動がとれるようになりたい。	非常事態でも、人道に従った行動がとれるようになりたい。		*非常事態でも、人道に応じた行動がとれるようになりたい。	「人道」はルールを表す	
411	～に即る						
412	交換規則	々変名詞+(する+)に足る	々変名詞+する+だけの値がある		々変名詞+できる		
413		あの人は信賴するに足る人物です。	あの人は信賴するだけの値がある人物です。		あの人は信賴できる人物です。		
414	選択制限	これはわざわざ議論するに足る問題だろうか。	これはわざわざ議論するだけの値がある問題だろうか。	?	*これはわざわざ議論できる問題だろうか。	?	
415		彼は今度の数学オリンピックで十分満足に足る成績がとれるだろう。	*彼は今度の数学オリンピックで十分満足するだけの値がある成績がとれるだろう。	?	彼は今度の数学オリンピックで十分満足できる成績がとれるだろう。		
416							
417							

図 4: 表現文型言い換え辞書

2.3 可読性基準のモデル化

可読性基準のモデル化については、聾者の読解能力について豊富な経験的知見を持つ聾学校の語学担当教諭（国語、英語）を対象に文の可読性に関するアンケート調査を行い、形態・構文的要因と可読性の関係のモデル化を試みた[14].

この調査では、様々な構文的言い換え事例を予め用意し、言い換え前と言い換え後の可読性の相対的な差を被験者（聾学校語学担当教諭）に判定させる（図5参照）。こうして得られたデータを半自動的に分析することにより、文の形態・構文的特徴と可読性の関係のモデル化が可能になると期待できる。1999年夏の予備調査の結果[23]を踏まえ、2000年夏に全国の聾学校中等部および高等部の約半数に当たる50校に調査を依頼した[24]。この結果、延べ4,200件の言い換えの組（異なり数780組、1組当たり平均3.8文、1問当たりの平均回答者数7.7人）について、各文の相対的な難易度データを収集することができている。

また、上の作業と並行して、回答データから可読性評価モデルを半自動獲得する研究にも着手した。各調査対象文を統語的属性の束で表現し、統語的属性の任意の組み合わせと相対的可読性の関係を、決定リスト学習のような何らかの機械学習の方法で獲得することを考える。具体的には、構文的言い換えの関係にある任意の2文（言い換え対）を入力として、どちらの文がより可読性が高いかを判定するというタスクを設定し、言い換え対の構文的属性から可読性の差を推定する規則集合を回答データから獲得する実験を行った。個々の規則は、たとえば次のようなものである（Dの方がEより理解が容易であることを表す）。

- (5) D: <補足節「こと」内容>かつ <語彙「不可能」>
 (例. 私は、太郎が今度のテストで100点を取ることは不可能だと思う)
- E: <可能「ことができる」>かつ <事態の否定>
 (例. 私は、太郎は今度のテストで100点を取ることができないと思う)

今回の実験では、事例集合の過疎性や規則獲得アルゴリズムの稚拙さなど、いくつかの予想された問題が残っているものの、交差検定で適合率0.94、再現率0.58の精度を得ることができた。

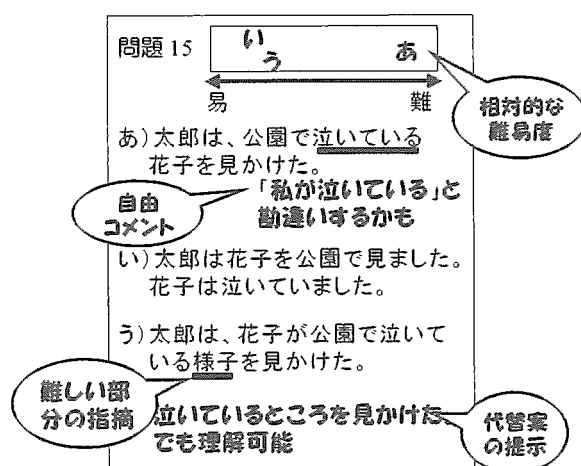


図 5: アンケートの調査問題の形式

以上から、今後同様のアンケート調査を継続し、事例集合を拡大していけば、少なくとも聾学校教諭の可読性判定に関する限り、これをおある程度模倣するモデルの構築が可能になると期待できる。聾者の文理解能力に関する従来の研究はもっぱら聾者を被験者とする調査に基づいており、今回のような聾学校教諭を対象とする調査は皆無であった。聾者を被験者とする調査は、聾者の文理解能力をより直接的に観察できる反面、調査対象の構文パターンを大きく制限せざるを得ず、包括的な可読性基準の構築に至るのは難しい。我々の調査方法は、当該分野のエキスパートである聾学校の語学教諭の専門知識をアンケート調査によって外在化させ、これをモデル化しようとするもので、読解支援という工学的目的には妥当な方法と考えている。

2.4 読解支援プロトタイプシステム

上述の言い換えエンジン KURA と Web 上のプロキシサーバを接続し、読解支援プロトタイプシステムを構築した。現在、kura.pluto.ai.kyutech.ac.jp:8088 にあるプロキシサーバを通して Web ページをブラウズすると、選択したテキストの単純化が実行できるようになっている。現在、このシステムをたたき台として、そのデザインについて現場からのフィードバックを集める努力をしている。

現在のシステムの振舞いのスナップショットを図 6 に示す。ウィンドウの右側が入力テキストで、左側中段が単純化後のテキストである。ユーザは任意のプレーンテキストを選択することができる。操作のオプションには〈一括言い換え〉と〈対話的言い換え〉がある。〈一括言い換え〉は、適当なデフォルトの可読性基準に基づいてテキスト全体を言い換える。一方〈対話的言い換え〉では、ユーザが入力テキストあるいは〈一括言い換え〉後のテキストを読み進めながら、難解な箇所をマウスで自由に選択し、システムに問い合わせることができる。システムは、問い合わせに応じて、指定された箇所をさらに平易な表現に言い換えたり、注釈的説明を付け加えたりする。〈対話的言い換え〉のねらいは、デフォルトの可読性基準では対処できないユーザの読解能力の個人差を部分的に吸収することである。

3 まとめと今後の課題

またたく間に過ぎた 3 年間だった。自然言語の言い換えは、言語の理解と生成の両方の側面を持つ複雑な問題であり、もとより 3 年間で実用レベルの実現に到達できるとは考えていなかった。そうした認識のもと、本研究期間で何とか達成したいと考えていた目標は、技術的には、可能なかぎり多くの言い換え事例にまみれることによって言い換えの問題の構造を解明し、個々の部分問題へのアプローチの仕方を確立すること、そして社会的には、聾者や失語症患者、その支援者らとコミュニケーション支援技術開発の人的ネットワークを育てることであった。

前者については、まずは、最終的に現在の言い換えエンジン KURA の開発にたどり着いたことが重要な成

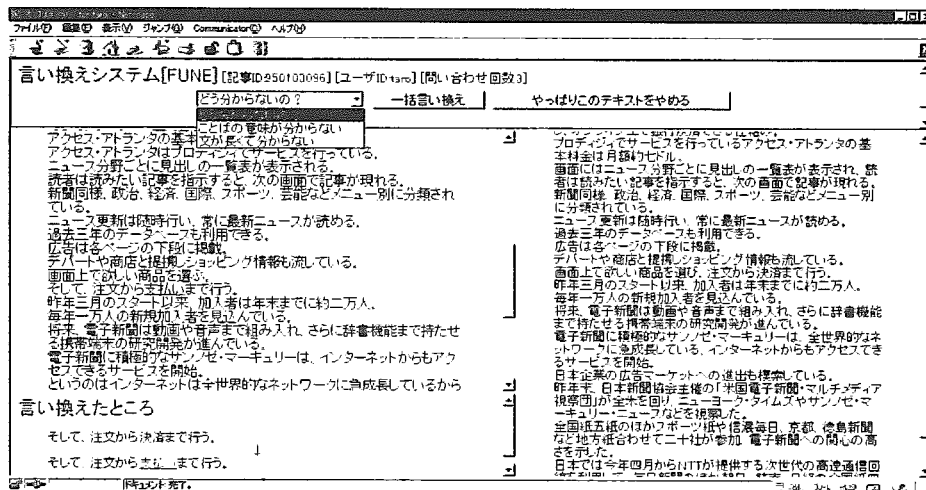


図 6: 読解支援システムのプロトタイプ

果と考えている。また KURA あるいはその前世代のシステムを利用して数多くの言い換え生成実験を実施し、大量の言い換え事例を実際に自分の目で観察したことによって、少くとも「事例にまみれる」という目標は達成できた。本報告は、あるいは、脈絡のない個別的な試みの寄せ集めのように見えたかもしれない。しかし、言い換え技術の着実な進歩は、少なくとも初期の段階では、こうした個別の事例と格闘する努力なしにはあり得ないと私は信じている。今後は、実験を通して得られた、あるいは得られつつある知見や知識・事例の蓄積を他の研究者と共有する仕組みを早急に構築する必要があるだろう。また、個別の部分問題に必要な知識について、その大規模化をはかる手段を確立していく必要がある。

後者の人的ネットワークの構築については、福岡県内の聾学校、聴覚障害者センター、手話サークルなどの多くの関係者、また特殊教育分野の研究者ら、言語療法分野の専門家ら、実に多くの方々と本研究を通して交流を持つことができた。福祉的コミュニケーション支援技術は今後じっくりと腰を据えて取り組むべき課題であり、こうした関係分野各位との交流は今後の重要な財産になると確信している。

これまでに得た知見と資源を Web サーバ上に統合し、高等聾学校生を対象とする使用実験を行う。また、成果をしかるべき場所に公表するとともに、特殊教育分野および医療リハビリ分野への広報活動を行う。

謝辞

情報と知領域の総括・アドバイザーの先生方、その他の研究者の皆様には、本研究に対し示唆に富む実に多くのアドバイスをいただき、言葉にできないくらい多くのことを学ばせていただきました。この場をお借りして厚くお礼申し上げます。また、領域事務所の皆様には、3年間にわたって終始温かく本研究のサポートをいただきました。深く感謝するとともに、事務手続き等で何度もご迷惑をおかけしたことを、お詫び申し上げます。

文可読性のアンケート調査に快くご協力くださいました全国聾学校の先生方に心から感謝申し上げます。同調査の実施に当たっては、上越教育大学の我妻敏博氏から多くの有益なご助言とご協力をいただきました。また、福岡県内の聴覚障害者センター、聾学校、手話サークルなど、多くの関係者からも有意義なご助言をいただきました。千葉労災病院の言語聴覚士、安田清先生にも現場からの説得力のあるアドバイスを多くいただきました。結束性評価の研究については、フランス LIMSI の Michael Zock 氏から有益なご助言をいただきました。言い換えエンジン KURA の開発に当たっては、東京大学辻井研究室で開発されたプログラミング言語 LiLFeS、奈良先端大松本研究室で開発された形態素解析器 ChaSen、および構文解析器 Cabocha を使用させていただきました。また言語モデルの開発に当たっては NTT コミュニケーション科学研究所の日本語語彙大系を利用させていただいています。同辞書は、慣用句の言い換え知識の獲得にも利用させていただきました。表現文型言い換え辞書の作成に当たっては、(有) インターナショナルランゲージウェアの衛藤純司氏にご協力

いただきました。普通名詞の言い換えの実験では、京都大学の河原大輔氏、黒橋禎夫氏が作成された格フレーム用例データを利用させていただきました。ここに深く感謝申し上げます。

成果の公表

- [1] 藤田篤, 乾健太郎, 乾裕子. 名詞言い換えコーパスの作成環境. 電子情報通信学会思考と言語研究会, TL2000-32, 2000.
- [2] 藤田篤, 乾健太郎. 語釈文を利用した普通名詞の同概念語への言い換え. 言語処理学会第7回年次大会, 2001.
- [3] 飯田龍, 徳永泰浩, 乾健太郎, 衛藤純司. 言い換えエンジン KURA を用いた節内構造および機能語相当表現レベルの言い換え. 第63回情報処理学会全国大会, 3H-03, 2001.
- [4] Inui, K. and Wakigawa, H. POS-tag conversion algorithm for reusing corpora. *The 5th Natural Language Pacific-Rim Symposium*, 1999.
- [5] 乾健太郎, 山本聡美, 野上優, 藤田篤, 乾裕子. 聾者向け文章読解支援における構文的言い換えの効果について. 電子情報通信学会福祉情報工学研究会, WIT99-2, 1999.
- [6] 乾健太郎. コンピュータが言葉を話せるようになるまでに一言語生成研究の現状と課題一. 人文学と情報処理, No.21, 1999.
- [7] 乾健太郎. テキスト単純化による聾者向け文章読解支援一現状と展望一. 電子情報通信学会福祉情報工学研究会, WIT2000-34, 2000.
- [8] 乾健太郎, 白井清昭. 例文を使って文の解析をしよう. 情報処理, Vol. 41, No. 7, 2000.
- [9] Inui, T. and Inui, K. Committee-based decision making in probabilistic partial parsing. *The 18th International Conference on Computational Linguistics*, 2000.
- [10] Inui, K., Virach Sornlertlambanich, Tanaka, H. and Tokunaga, T. Probabilistic GLR parsing. In H. Bunt and A. Nijholt (eds), *Advances in Probabilistic and Other Parsing Technologies*, Kluwer Academic Press, 2000.
- [11] 乾健太郎. コミュニケーション支援のための言い換え. 言語処理学会第7回年次大会併設ワークショップ「言い換え/パラフレーズの自動化に向けて」, 2001.
- [12] 乾健太郎. 読解支援を目的とするテキスト単純化の実現に向けて一課題と方法論一. 電子情報通信学会思考と言語研究会, TL2001-8, 2001.
- [13] Inui, K. and Nogami, M. A paraphrase-based exploration of cohesiveness criteria. *The 8th European Workshop on Natural Language Generation*, 2001.
- [14] Inui, K. and Yamamoto, S. Corpus-based acquisition of sentence readability ranking models for deaf people. *The 6th Natural Language Processing Pacific Rim Symposium*, 2001.
- [15] 岩倉友哉, 高橋哲朗, 乾健太郎. 不完全な内部表現および変換規則を用いた言い換えの実現方法, 第15回人工知能学会全国大会, 2001.
- [16] 橋本さち恵, 乾健太郎, 白井清昭, 徳永健伸, 田中穂積. 日本語文生成における照応表現の選択. 情報処理学会自然言語処理研究会, NL-143, 2001.
- [17] 神田慎哉, 藤田篤, 乾健太郎. 連用節主節化に関する規則の追試と洗練, 第15回人工知能学会全国大会, 2001.
- [18] 野上優, 藤田篤, 乾健太郎. 文分割による連体節の言い換え. 言語処理学会第6回年次大会, 2000.
- [19] 野上優, 乾健太郎. 結束性を考慮した連体修飾節の言い換え. 言語処理学会第7回年次大会, 2001.
- [20] 高橋哲朗, 岩倉友哉, 乾健太郎. 不完全な構造変換規則による言い換えの実現方法. 言語処理学会第7回年次大会, 2001.
- [21] 高橋哲朗, 岩倉友哉, 飯田龍, 乾健太郎. KURA: 統一的かつ宣言的記述法に基づく言い換え知識の開発環境. 電子情報通信学会思考と言語研究会, TL01-12, 2001.
- [22] Takahashi, T., Iwakura, T., Iida, R., Fujita, A. and Inui, K. KURA: A transfer-based lexico-structural paraphrasing engine. *NLPRS2001 Workshop on Automatic Paraphrasing: Theories and Applications*, 2001.
- [23] 山本聡美, 乾健太郎, 野上優, 藤田篤, 乾裕子. 聾者向け文章読解支援のための文可読性基準の調査. 情報処理学会自然言語処理研究会, NL-135, 2000.
- [24] 山本聡美, 乾健太郎, 乾裕子. 聾者向け文章読解支援のための文可読性基準のモデル化. 情報処理学会自然言語処理研究会, NL-141, 2001.