

「やさしい日本語」を用いたユニバーサルコミュニケーション実現のための予備的考察

庵 功雄・岩田 一成・筒井 千絵・森 篤嗣・松田 真希子

要旨

地域社会において在住外国人と地域住民とのコミュニケーションを考える際には、在住外国人側にも一方的に日本語習得、日本文化への同化を強いることは不相当である。そうした問題意識に立ち、本稿では「やさしい日本語」を媒介とするコミュニケーションを提案する。これは、在住外国人側にも日本語習得のための一定の努力を求める一方、地域住民側にも自らの日本語使用を在住外国人にわかりやすいように調整することを求めるものである。

初めに、学習者に習得してもらいたい「ミニマムの文法」とその発想に基づく教材について提案する。次に、地域住民側の在住外国人への接近のあり方の一例として公的文書の翻訳を検討する。これは本稿で日々ほんやくコンニャクプロジェクトと呼ぶものの中核をなすものである。まず、日本語教育経験が豊富な日本語教師が公的文書を「逐語訳、意識、要約」に分けて書き換えた内容について報告する。次に、その書き換えの内容を語彙の面から検討する。これは公的文書を読むための基本語彙を設定するための基礎的作業である。最後に、作成されたコーパスを機械翻訳ツールに載せた分析結果を報告する。これは今後本格的に研究を行い、公的文書の日日翻訳を自動化するための前段階をなすものである。

キーワード：日々ほんやくコンニャクプロジェクト（HK プロジェクト）、おしゃべり型教材、書き換え、基本語彙、統計翻訳

1. 日々ほんやくコンニャクプロジェクト

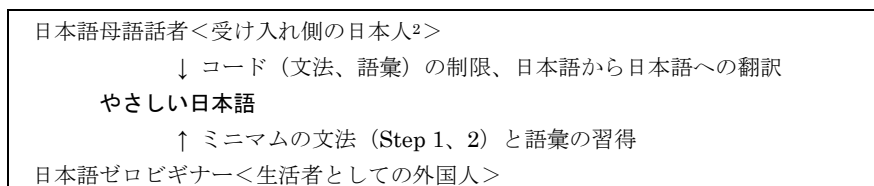
日本国内に在住する外国人の人口は 200 万人を超え、その人々に対する公的言語保障の問題が重要になってきている (cf. 山田 2002)。この問題に対して、本稿では「やさしい日本語」¹の規格化とその活用のための教材の開発、および、公文書の書き換え（の自動化）というアプローチを提示する。

本研究グループでは先に、地域日本語教育の「初級」を現在の「学校型」日本語教育のそれよりも狭く限定すべきことを日本語教育文法の観点から論じた (庵 2009)。また、広島市発行の市民向け公文書を「やさしい日本語」で書き換えた結果から、文法項目に関する限り、こうした「ミニマムの文法」で十分であることを示した (庵・岩田・森 2009)。

¹ 阪神大震災などの経験をもとに、災害時における外国人に対する情報提供を目的とする「やさしい日本語」(以下、「減災 EJ」)の必要性が提唱されている (佐藤 2004 ほか)。減災 EJ と本稿の「やさしい日本語」の相違点については庵・岩田・森 (2009) を参照されたい。

以上の成果を踏まえ、本稿では、地域日本語教育における「やさしい日本語」を次のように位置づける。

(1) 地域日本語教育における「やさしい日本語」の位置づけ



(1)のうち、「普通の」日本語から「やさしい」日本語への翻訳の部分を「日々ほんやくコンニャクプロジェクト（HKプロジェクト）」と仮称する。この名称は「やさしい日本語」という共通の言語のもとに、地域住民（受け入れ側の日本人）と生活者としての外国人がコミュニケーションを行うという理想を述べたものである。さらに、本稿で対象としている日日翻訳は高齢者などにとっても「やさしい」ものであり、日本に居住する全ての人にとってのディスコミュニケーションを解消し、ユニバーサルコミュニケーションを実現することを目指すものでもある。

2. 地域日本語教育に適した初級教材の在り方

従来の「やさしい日本語」は外国人をメッセージの受け手と考えてきたが、HKプロジェクトでは、テキスト製作を通じて外国人を能動的な主体として捉えている（イ 2009）。ここでは地域日本語教育の教材としてどういったものが必要であるかを論じる。地域日本語教育は様々な活動を内包するが、米勢（2002）はその役割を以下のように3分類している（括弧は外国人参加者の視点：表は岩田が再構成したもの）。

教室運営	① 日本語教育（日本語学習）
	② 相互学習・交流（日本語習得）
	③ 生活支援・社会認識（生活上の問題の解決）

我々の主張は、②タイプを活性化することで、日本語教育に貢献しようとするものであり、決して①の日本語教育をボランティアに押し付けようとするものではない³。

² 高度な日本語力を有し、日本語が不自由な在在外人と日本語でコミュニケーションを取る外国人（日本国籍を持たない者）もここでいう「日本人」に準じて考える。
³ ただし、表の3分類ははっきりと区別できるものではないことも確認しておきたい。例えば日本語教育学会（2009：248-50）で授業分析として金田氏が紹介している例は、日本人が離乳食の作り方を話しているのに、外国人側は「ぬらす」という語彙の意味を確認しており、表の①～③にまたがっている。

ここからの議論は、尾崎（2004）で指摘されているものを出発点としたい。さらに我々の実践経験を加えて地域日本語教育で求められていると考える教材像をここで述べたい。以後尾崎（2004）に倣い、大学や日本語学校で行われている日本語教育を「学校型」、ボランティア主導の日本語教室での日本語教育を「地域型」と呼ぶ。

2.1 シラバス：文法圧縮

小林（2009：28）では、「初級（3・4級レベル）では、助詞や活用などの日本語の基本的な文型に関わる要素をひと通り教える。→中級（2級レベル）以降では、複合辞や機能語を教える。」という文法教育観があることを指摘している。野田編（2005）でも現行のシラバスが日本語学に引きずられていることを指摘している。こういう指摘から考えると、日本語の初級と呼ばれるレベル（学習時間 300 時間）は、「地域型」には重すぎるのである。

この問題に関しては、すでに庵（2009）において、一機能一形式に限定してしまえば、初級がずっと軽くなるということを指摘し、STEP1・2 という文法項目を提案している。一例を示せば、「と・ば・たら・なら」という表現を「たら」一つに絞ってしまえば、文法項目は圧縮できる上に、類似文法の使い分けといった説明も必要なくなるわけである。

提案 1：地域型では初級の文法項目を大幅に圧縮して教材に採用すべきである

2.2 方法：「教授型」から「おしゃべり型の教育」へ

「学校型」に見られる教育方法を「教授型」と呼びたい。「学校型」は一般的に、文法積み上げ方式によって教員が学生に教育内容を指導する形で授業は進行する。このスタイルの前提となるのは、教員は教案を書くなどの授業準備をすること、学生と教員の間に金銭関係があること（直接的であれ間接的であれ）、学生は学業を生活の主な目的としていることなどが挙げられる⁴。こういった前提が「地域型」では当てはまらないわけであり、そのような状況に「学校型」の方法論を適応してもうまく機能しない。こういった背景から我々は「おしゃべり型の教育」が適していると考え⁵。ここで注意したいのは、単なる「おしゃべり」ではなく「おしゃべり型の教育」という点である。「おしゃべり型の教育」とは、話題中心に広がっていく授業の裏に、文法項目を散りばめていくというものである。これを我々はトピックシラバス（隠れ文法付き）と呼んでいる。この隠れ文法の部分に 2.1 で紹介した STEP1・2（庵 2009）を組み込めばよいのではないだろうか。

提案 2：地域型にはトピックシラバス（隠れ文法付き）による「おしゃべり型の教育」が適している

⁴ これに関しては日本語教育学会（2009：247）で尾崎氏が「学校型」は「教える方も習う方もそれに責任を持っていることが前提になっている」とし、それが「地域型」との違いであると指摘している。本稿の主張は、それを具体化したものである。

⁵ 用語は様々であるが多くの文献で同様の指摘がなされている。その流れについては日本語教育学会（2008：91）などに詳しくまとめている。

2.3 活動の限界・参加者のバリエーションに対応

ここでは参加者の特性について考察したい。便宜上日本人参加者、外国人参加者という呼び方をする。日本人参加者は必ずしも日本語教師である必要はなく地域住民に広く開かれている。その参加者は毎週必ず来るわけではなく、都合のいい時に来る。これは外国人参加者も同じであり、両者の関係は固定的ではありえない。こういった活動の限界を考えると、「地域型」の教材は、はじめから順番に進めていくタイプよりも、どこからでも始められるモジュール型が適していると考えられる。

モジュール型でありながら、2.2 で述べたような隠れ文法を埋め込もうとすると、1 トピックに1 文法という対応ではなく、1 トピックに複数の文法が使われ、かつ同じ文法が複数のトピックに何度も出てくるような形式を取らねばならない。これを「文法スパイラル」と呼ぶ⁶。こういった方策は、外国人参加者の多様性に対応するためにも必要である。

「地域型」は、様々なレベルで様々な母語を持つ外国人参加者にできるだけ広く対応しなければならない。特に「やさしい日本語」の発想では、初級レベルに焦点をあて、ある一定のレベルまで高めていけるような教材でなければならない。レベルや母語の違いを広く受け入れるには、イラストを多用して、指差し会話のようなコミュニケーションを出発点とすべきである。

提案3：モジュール型で「文法スパイラル」を採用し、イラストが多用された教材が適している

ここまでの提案を踏まえて作成された教材が『にほんごこれだけ！1』（庵監修 2010）である。この教材で我々は「地域型」初級教材の一つの形を提案している。

3. 書き換えコーパスの作成

日本語教育経験者（10年以上）2名により、生活者が読み理解する必要がある文書を対象に、「逐語訳」「意識」「要約」の3段階で「やさしい日本語」への書き換えをおこなった。ここでは、実際の書き換え例を挙げ、書き換えの観点や問題点などを報告する。

3.1 対象とした文書

生活者が読んで理解する必要があると思われる通知類（B5版またはA4版で1～2枚程度を基準とした）、計116文書10万字弱の書き換えをおこなった。書き換えの対象とした文書は以下のようなものである。

例：学校のおたより、学童保育所申込みの案内、予防接種の案内、がん検診のお知らせ、断水のお知らせ、給付金の案内、振り込め詐欺に対する注意、など

⁶ 文法を繰り返し提示する必要性は小山（2007）でも指摘されており、その教材に採用されている。

3.2 書き換えの基準

書き換えに際しては、STEP1・2（庵 2009）の文法を原則として使用し、生活者としての外国人が読むことを想定してできる限りやさしく言い換えることを意図した。

また、書き換えは原文の「逐語訳」「意識」「要約」の3段階でおこなった。ここでいう「逐語訳」「意識」「要約」の基準を以下に示す。

- ①逐語訳……語・句のレベルでやさしく言い換える。元の文の流れは変えない
- ②意 訳……段落内で文や節をわかりやすく再構成する
- ③要 約……複段落レベルで再構成をおこなう。文章全体の伝達意図のみを残す

3.3 書き換えの観点

ここでは「やさしく」するためにどのような言い換えの方略を用いたのかについて報告する。

- ①長く複雑な文はできるだけ短く単純な文にする

例 1 原 文：ワクチン接種は多くの方々に重症化予防というメリットをもたらしますが、接種後、はれたり、熱が出るなどのケースもあり、まれではありますが、重篤な症状を引き起こす可能性もあります。

逐語訳：ワクチンで軽くなる人が多いです。でも接種の後、はれたり、熱が出たりすることもあります。少しですが、病気が悪くなることもあります。

- ②具体化する

例 2 原 文：冷えることが予想されますので、防寒の用意をお願いいたします。

逐語訳：たぶん寒いので、暖かい服を着て来てください。

例 3 原 文：フッ化物洗口薬品代について（文書のタイトル）

要 約：口を洗う薬のお金を集めます

- ③重要度の高い情報を先に持ってくる

例 4 原 文：介護保険制度は3年を一期として改定されます。平成21年度4月1日より介護保険制度が改正となりました。Xにおきましても、改正された介護保険制度に伴い平成21年4月1日のサービス利用分より、利用料金の改定をさせていただくこととなりました。※Xは施設名

例 5 意 訳：Xの利用料金が、平成21年4月1日のサービス利用分から、変わります。介護保険制度が平成21年度4月1日から新しくなった（3年に1度変わります）からです。

- ④不要な情報は削除する（時候の挨拶、前置き、重複した情報など）

例 6 通学区域外の学校に入学を希望する場合のみ、選択の申し出の手続きが必要となります。通学区域内の学校に入学を希望する場合は、特に手続きの必要はありません。（情報の重複）

→意識：通学区域以外の学校に入学したい人だけ手続きが必要です。

→要約：削除

3.4 書き換え上の問題点

文書を書き換えるということは元文書に手を加えることであるため、情報の取捨選択や正確さに関する問題は避けられない。以下に書き換え作業を通して生じた問題点について報告する。

①書き換えにより文字量が増える

例 7 校区外の幼稚園等に通っている場合

→住所で決まっている幼稚園じゃない幼稚園に通っている場合

②制度、病気、書類の名称などは書き換えができず、「やさしく」ならない

例 8 前年の所得金額が 300 万円以上の白色事業所得のある方

→前年の所得金額が 300 万円以上の白色事業所得がある人

③書き換えにより正確さが失われる可能性がある

例 9 虐待→暴力/いじめ、世帯→家族

④表現の与える印象の問題

例 10 受給資格決定→お金をあげることができるか決めます。

⑤文法的制約⁷による書き換えの困難さ

例 11 日頃より保護者の皆様にはご協力いただきまして、ありがとうございます。

→いつも保護者の皆さんの協力、ありがとうございます。（敬語、「てもらう」の回避）

例 12 （子どもに）外出を控えご家庭でゆっくり過ごさせるようお願いいたします。

→外に行かないで家でゆっくりしてください。（使役形の回避）

3.5 その他の課題

①収集する文書の種類

家族形態（子どもや要介護者の有無など）により手元に来る文書の種類や数が大きく異なるため、その点に配慮した文書の収集が必要である。

②通知類の理解に役立つ表現

やさしいとは言えないが、公的文書に高頻度で出現し、簡潔に言い換えられない語や文型に関しては、今後書き換えをおこなうか「通知類の理解に役立つ表現」として習得を促進するか検討が必要である。

⁷ 書き換えに際しては、STEP1・2（庵 2009）の文法を原則として使用し、その範囲を超える文法の使用はできるだけ避けた。

例：手続き、登録、申請、機関、場合、証明、緊急、書類、地域、年度、保護者までに、について、～先（問合せ先、振込先など）、とは など

③背景知識の必要な語への配慮

学校の配布物の語彙は多様かつ理解に背景知識が必要なものが多いため、外国人住民の負担は大きいと思われる。よって基本語彙の選定や書き換えに際しては配慮が必要である。

例：個人面談、お楽しみ会、書き初め、しおり、学年だより、など

4. 公的文書の書き換えのための基本語彙の抽出

4.1 日常生活における日本語の問題

日本における日常生活において、日本語ができるかできないかで、大きな不利益を被る可能性があるものの一つとして、日常生活で目にする公的文書（市役所や学校からの文書）や自治体 HP がある。この問題を解決するには、「日本人側がやさしく書く」か、「外国人側が読めるようになる」かのいずれかしかない⁸。

ここでは、「日本人側がやさしく書く」を目指し、われわれが作成した試用版書き換えコーパスを用い、日常生活で目にする公的文書（以下、公的文書とする）に含まれる語彙の現状と、日本語教育経験者がやさしく書き換えた場合の語彙について述べる。

4.2 公的文書の書き換えのための基本語彙

基本語彙とは何か。その定義について、林（1971:2）では以下のように示している。

- (1) 基礎語彙 意味の論理的分析によって求められた半人工的な語彙
- (2) 基本語彙 特定目的のための「〇〇基本語彙」
- (3) 基準語彙 標準的社会人としての生活に必要な語彙
- (4) 基調語彙 特定作品の基調を作るのに働く語彙
- (5) 基幹語彙 ある後集団の基幹部として存在する語彙

いずれも興味深いですが、ここでは基礎語彙と基本語彙の違いについてのみ考える。基礎語彙については、C. K. Ogden が 1932 年に発表した Basic English (850 語) と、Basic English の影響を受けた土居（1933）の基礎日本語（1,000 語）が有名である⁹。ただし、林（1971）が指摘するように、Ogden の Basic English も土居の基礎日本語も、「実用目的の文章において」、前者は「外国人のため」、後者は「母国語の一般大衆のため」という特定の目的

⁸ しかし、現状では「日本人側がやさしく書く」という対応は十全ではない。例えば、国立国語研究所(2004)による広報誌編集担当者への調査によると、「広報誌の編集で高齢の読者に配慮していること」で「特に配慮はしていない」は 3.9%だが、「広報誌の編集で外国人の読者に配慮していること」で「特に配慮はしていない」は 78.9%である。

⁹ 土居は 10 年後の 1943 年に基礎語彙を 100 語増やした 1,100 語を公開している。

があることを考えると、基礎語彙というより基本語彙であるとも言える。そう考えると、基礎語彙をどう定義づけるかは極めて難題である。

一方、基本語彙については「特定目的のため」という定義でよいだろう。教育のためであれば、教育基本語彙ということになる。今回のターゲットは公的文書であるため、公的文書基本語彙と呼ぶこととする。

4.3 語彙数についての目安

公的文書の書き換えを行う場合、語彙をどこまで制限するかで相当に印象が変わってくる。ここでどこまで語彙を制限するかということが悩みとなる。語彙を制限しすぎると、「不自然な日本語」「出来損ないの日本語」など簡約日本語が受けた批判と同様の批判を受けかねない。しかし、語彙を多くリストアップしてしまうと、「どこがやさしい日本語なのか」という批判をやはり受けることになる。

先行研究を簡単に概観してみると、土居（1933）の 1,000 語が最も少なく、次に水野（2006）など減災 EJ が想定する日本語能力試験 3 級の 1,500 語（国際交流基金・日本国際教育支援協会（2002））、さらに野元（1993）の簡約日本語や国立国語研究所（1984）の「基本語二千」の 2,000 語と続く。少し離れて、日本語能力試験 2 級と国立国語研究所（1984）の「基本語六千」の 6,000 語となる。

公的文書の伝達目的は、「書き言葉で過不足なく情報を伝達する」と考えられる。その意味で、語彙制限を厳しくした結果、過不足のない情報伝達が不十分ということになってしまうと本末転倒である。そう考えると、どれだけの語彙で公的文書のどのくらいをカバーできるのかという被覆率を基に目安を考えるということが必要になってくる。

被覆率に関しては、林（1971:9）では国立国語研究所の現代雑誌九十種調査をもとに、5,000 語で 81.7%、10,000 語で 91.7%という結果を示している。また、我々も松田ほか（2010）において、Wikipedia、Yahoo! 知恵袋、日本経済新聞、livedoor Blog というタイプの違う 4 つのコーパスを組み合わせ、上位 5,000 語の被覆率の調査を行った。

表 1 4 種コーパスの基本情報（松田ほか（2010）より）

	総単語数 ¹⁰	異なり語数	累積頻度 90% 超えの単語順位	コーパス のタイプ
Wikipedia ¹¹	171,772,307	792,685	7,278	学術文、文語的
Yahoo! 知恵袋	57,679,832	162,097	3,870	会話文、口語的
日本経済新聞	692,754,923	315,267	3,599	報道文、文語的
livedoor Blog	1,424,278,654	86,296	4,210	日記文、口語的

¹⁰ 未知語、記号、固有名詞を除く。

¹¹ 人文科学・社会科学・自然科学の分野から収集した。

松田ほか（2010）では、ChaSen-2.4.2 及び、IPAdic-2.7.0 を用いて表 1 のコーパスの形態素解析を行い、各コーパスにおいて、出現頻度の高かった語の上位 5,000 語をリスト化した。また、4 種のコーパスに現れた語彙について、各コーパスでの頻度を加算し、その頻度の総和の順に並べた場合の上位 5,000 語をリスト化した¹²。この上位 5,000 語による各コーパスの被覆率は、Wikipedia が 82.26%、Yahoo! 知恵袋が 87.33%、日本経済新聞が 89.68%、livedoor Blog が 87.98%であった。

もちろん、これらの結果は公的文書コーパスによる結果ではないため、公的文書基本語彙の目安にはなり得ないが、日常生活に目にする語彙数のある種の目安になり得ると考えられる。

4.4 試用版書き換えコーパスにおける公的文書の実態と書き換えの効果

3 で見たとおり、現在、公的文書の書き換えコーパスの構築を進めている。このコーパスが大規模になれば、その原文部分から公的文書基本語彙を抽出する予定である。今回は、この試用版書き換えコーパス（116 文書）に基づき、公的文書の実態と書き換えの効果について検討する。

表 2 試用版書き換えコーパスの基本情報と累積頻度（被覆率）

	総文字数	総単語数 ¹³	異なり語数	累積頻度 80% 超えの単語順位	累積頻度 90% 超えの単語順位
原文	96,280	51,827	4,078	519	1,176
逐語訳	87,160	49,967	2,774	418	888
意訳	84,424	42,203	2,591	173	315
要約	83,536	39,610	2,547	129	226

まず、総文字数を見てみると、原文から逐語訳の段階で差がある。3 で「書き換えにより文字量が増える」という問題点が指摘されたが、データから見ると文字数減は達成されている。ただ、要約でそれほど文字数が減少していないのは意外である。

何より注目したいのは異なり語数である。原文の 4,078 語に対し、逐語訳では 2,774 語に圧縮されている。「やさしく書き換えよう」という意識に基づく統制のかかった書き換えにより、異なり語数がかかなり抑えられることがわかった。3 において「制度、病気、書類の名称などは書き換えができない」とあったように、公的文書の性格から書き換え困難な語彙があるにもかかわらず、かなりの効果を挙げていると言える。

¹² Livedoor blog のコーパス規模が大きいため、頻度の総和ではなく、各コーパスにおける各語彙の出現頻度の割合を求め、その総和から上位 5,000 語を選出したが、結果はほぼ同じだった。

¹³ 未知語、記号、固有名詞を除く。

頻度情報の詳細をまず機能語から見てみると、格助詞「が」は原文では 834 回（10 位）、逐語訳では 1,154 回（7 位）、格助詞「を」は原文で 1,333 回（2 位）、逐語訳で 1,468 回（3 位）と、いずれも逐語訳で増加している¹⁴。これは、複雑な構文が基本的な構文に書き換えられていることを示唆している。逆の傾向を示している語としては、接続助詞「が」は原文では 84 回（78 位）であるが、逐語訳では 63 回（115 位）と減少している。これは不必要な複文が整理されていることを示唆している。

次に頻度情報の詳細を実質語から見てみる。ChaSen-2.4.2 及び、IPAdic-2.7.0 により原文のうちで名詞とカウントされた語の上位 10 語（数、接尾、非自立を除く）と、その逐語訳の対応は以下の通りである（頻度の右の矢印は、原文に対しての頻度の上下である）。

表 3 試用版書き換えコーパスにおける原文名詞上位 10 語と逐語訳の対応

	原文	頻度	逐語訳	頻度		原文	頻度	逐語訳	頻度
場合	24 位	311	50 位	133 ↓	利用	54 位	136	172 位	41 ↓
学校	42 位	188	31 位	266 ↑	必要	55 位	135	61 位	104 ↓
保護	47 位	156	56 位	121 ↓	相談	59 位	120	60 位	106 ↓
お願い	51 位	148	83 位	83 ↓	提出	60 位	109	100 位	71 ↓
接種	53 位	138	94 位	75 ↓	確認	64 位	105	106 位	67 ↓

このあと上位 20 位までは、65 位 電話（103）、66 位 申請（100）、69 位 機関（96）、71 位 証明（90）、72 位 児童（89）、75 位 健康（86）、76 位 連絡（85）、76 位 対象（85）、78 位 保険（84）、81 位 記入（83）が続いており、これらの語が現在の公的文書の実態を示している。

原文の頻度上位 10 語のうち、「学校」だけが頻度が増加しており、残り 9 語は軒並み減少している。書き換えにより、公的文書に特有な漢語は抑えられる傾向にあると言える。一方で、唯一増加した「学校」と同様の傾向を示すのが、「子ども」である。「子ども」は原文では 81 回（82 位）だが、逐語訳では 247 回（34 位）と急増している。「子ども」の場合は、書き換え対象となったと思われる「児童」が原文では 89 回（72 位）だが、逐語訳では 27 回（265 位）まで減少している。名詞の中でも、基本語と考えられる語へと集中して書き換えがされていると言える。

また、「場合」は名詞ではあるが、「～の場合」とった接続助詞相当として使われていることが多いと考えられ機能的であると言える。「場合」は原文では 311 回（24 位）だが、逐語訳では 133 回（51 位）まで落ちる。したがって、接続助詞「が」と同様に複雑な構文の書き換えを示唆していると思われる。

¹⁴ ちなみに原文、逐語訳、意識、要約の全てで大差の 1 位は助詞「の」である。

なお、累積頻度（被覆率）については、原文の異なり語数が 4,078 語と、もともと少ないため、80%や 90%の単語順位も低くなり、あまり参考にならない。今後、大規模書き換えコーパスの構築の結果を待ち、分析を進めたい。

4.5 本節のまとめ

このように、分量は少ないとはいえ、試用版書き換えコーパスから公的文書語彙の実態の片鱗と、書き換えの効果が見えてきた。松田ほか（2010）では、基礎語彙の重要な観点として「1）出現頻度が高いこと」、「2）意味領域を十分にカバーしていること」、「3）意味的な排他性が高いこと」の三つを挙げた。このうち、1)については従来のコーパスでも計ることができるが、2)と 3)については従来のコーパスでは困難である。

しかし、本研究の公的文書の書き換えコーパスでは、日本語教師の経験と直観により「やさしく書き換えよう」とした結果の語彙が集約される。「やさしく書き換えよう」という意識は、上記の 2)と 3)に対応する「これ以上は書き換えられない」という語彙の頻度を上げ、そうではない語彙の頻度を下げることから、書き換えコーパスの頻度情報は、1)だけでなく 2)と 3)もカバーし得る可能性がある。そうであれば、書き換えコーパスの頻度情報は、公的文書基本語彙の抽出にはもちろん、今まで困難を極めた基礎語彙の抽出にも貢献し得る可能性があり、社会貢献と同時に学術的価値も期待できると考えられる。

5. やさしい日本語への自動言い換えシステムの開発

本節では、日本語教師によるやさしい日本語書き換えコーパスを用いた、やさしい日本語への自動言い換えシステムの開発状況について述べる。

自動言い換え技術には、言い換え規則を手で構築して変換するルールベースの方法と、日本語（難）－日本語（易）の対となっている文（対訳文）を大量に用意して、統計的手法を用いて変換する統計ベースの方法がある。統計ベースの技術は google 翻訳などの機械翻訳で用いられている。

本研究では統計的機械翻訳（統計翻訳）の技術を利用して言い換えを行う。なぜなら、ルールベースの手法はルールを書くのに非常に手間がかかるが、統計翻訳においては、大量の対訳文を用意することができれば、以降の手続きは機械的な学習によって自動的に行われる。また、統計翻訳は日本語と韓国語のように、文の構造が類似する言語同士の翻訳においては有効に働くことが多いため（Hwang et al. 2005）、日-日翻訳でもあるやさしい日本語への言い換えもある程度の精度が出せると予想されるからである。

5.1 統計翻訳とは

統計翻訳の基本的な流れを図 1 に示す。統計翻訳では、まず大量の対訳文を収集・作成する。その対訳文を形態素解析した後、対訳文中の単語や句の対応付けを統計的にを行い、

翻訳モデルを作成する。その一方で翻訳先の言語のコーパスを収集し、形態素解析を行ったのちに統計分析を行い、最も確率が高い形態素の並びを推測したモデルを作る。これが言語モデルとなる。これらの処理が行われたのち、その二つをもとに、デコーダを用いて翻訳を行い、訳文を生成する。これが基本的な流れである。

5.2 実験方法

今回は 3700 の原文と逐語訳のペアを翻訳モデル用に使い、日本経済新聞のコーパス約 300 万文を言語モデル用に用いた。システム開発には、GIZA++[1]、Moses[2]、SRILM[3]、Chasen[4]を用いた。

次にテスト文 200 文について言い換え実験を行い、言い換え精度を評価した。精度評価には BLEU (Papineni et al. 2003) を用いた。BLEU は翻訳結果と正解データである参照訳との一致度を測る手法で、0-1 の値で示される。これは翻訳精度を測る手法で最も一般的なものである。

5.3 結果と考察

言い換え結果の例を表 4 に示す。精度評価の結果 BLEU 値は 0.199 (約 20%) と低くとどまった。また人手でも変換結果を評価したところ (表 5)、比較的良質な変換フレーズは 15 (7.5%) にすぎず、ほとんどが見出しのような短い句であった。

翻訳実験の精度が低かった理由は、三つあると考えられる。一つ目は翻訳モデルを作成するために使用した対訳文の量の問題、二つ目はコーパスの質の問題、三つ目は言語モデルの問題である。

対訳文の量については、今回用いた 3,700 ペアのうち、実際に言い換えているものは 2,663 ペアであった。統計翻訳を用いる場合、一般的にパラレルコーパスのサイズは 100 万ペア以上で行われる。そのため今回のサイズでは言い換えのためのモデルを作成するための情報が足りなかったと考えられる。対応策としては、ドメインを絞ることで使用される単語や句の数を制限するか、より多くの対訳文を用意する必要がある。

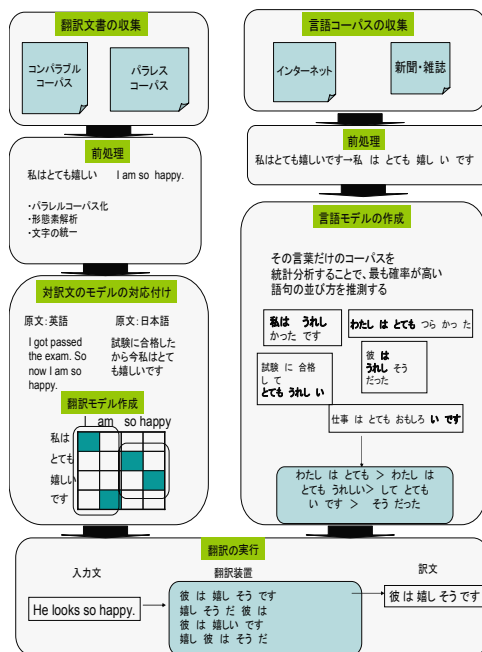


図 1 統計翻訳の流れ

表4 統計翻訳を用いたやさしい日本語への言い換え結果の例

	原文	日本語教師の言い換え	機械翻訳による言い換え
比較的良質	保護者各位	保護者の皆さんへ	保護者の皆さん
	Q6 接種費用の助成はありますか？	Q6 この注射のお金の助成はありますか。	Q?6 接種のお金を助けるはありますか
	Q3 医療機関に持参するものは？	Q3 病院にもって行くものは何ですか。	Q?3 する病院に持って行くものは
変換ミス	なお、修学旅行の詳細につきましては、後日生徒に配布される「修学旅行のしおり」をご覧ください。	それから、修学旅行の詳しいお知らせは、後で子どもに配る「修学旅行のしおり」を見てください。	なお、修学旅行の詳しい説明は、あと生徒にを変えたい「修学旅行のしおり」をので見て下さい。
	2. 悪天候の際には、大雨警報、暴風警報、大雪警報、暴風雪警報が発令されていなくても、周囲の状況で危険な場合は、保護者の判断で登校を見合わせてください。	2. 天気がとても悪いとき、大雨警報、暴風警報、大雪警報、暴風雪警報が出ていなくても、危ないと思ったら、学校に来ないでください。	2. で来悪天候お金を出すとき（ワクチンがあるか、大雨警報、周りにのているかなど、かもしいいと言つて危ない発令ていなくても、大雪暴風を、雪警報が暴風てください。見合わせ警報警報

表5 人手による評価結果

	比較的良質	解読不能・翻訳エラー	変化せず	合計
フレーズ数	15	172	13	200

コーパスの質の問題については、「①」「●」、「TEL」「→」等の記号がノイズとなり、翻訳エラーが起きている例が多くみられた。今回用いた書き換えコーパスは機械翻訳に適したフォーマットではなかったため、質を改善することで、翻訳精度が上がると予想される。また最も大きなノイズの原因になっているものは文中の括弧であった。そもそもやさしい日本語は括弧を使わないことが望ましいので、括弧内の文を展開する言い換え技術の開発が望まれる。

言語モデルについては、モデル構築に用いたコーパスが不相当であったと考えられる。今回は日本経済新聞を用いて言語モデル構築を行ったが、そもそも言語モデルは変換先の言語に対して有効に働くものなので、やさしい日本語に近いコーパスを用いることが望ましい。しかし、日本経済新聞は変換元の公文書に近い言語コーパスであったため、変換エラーを導きやすくなったと考えられる。そのため今後は日本語能力試験3級の問題テキストのような、やさしい日本語に近い言語コーパスを収集し、そのコーパスで言語モデルを構築することが重要となる。

5.4 本節のまとめ

本節では「やさしい日本語コーパスを用いた自動言い換え技術の開発の試みについて述べた。現状では十分な言い換えは行えていないが、コーパスの構築、改善やドメインの絞り込みを行うことで言い換え精度が上がる事が予想される。

6 まとめ

本稿では、「やさしい日本語」を媒介とするコミュニケーションの実現に向けた2つの試みについて提案した。一つは外国人に習得してもらうべき「ミニマムの文法」とその発想に基づく教材についてである。もう一つは公的文書の日日翻訳の自動化に関わる研究である。日日翻訳の自動化については、日本語教師による書き換えコーパス開発、基本語彙抽出、自動翻訳システム開発の三つの取り組みについて報告した。日日自動翻訳は現状では実用化までいたっていないが、今後本格的に研究を行い、3年後にはシステム化を実現する予定である。

7. 付記：地域日本語教育と「やさしい日本語」—国立市における1事例—

2で述べたように、本稿で述べた「やさしい日本語」の理念を具現化した教材として『にほんごこれだけ！』がある。現在『1』が出版されており、今後『2』が出版されることにより、1で述べたような「ミニマムの文法」が完結することになる。ここでは、『にほんごこれだけ！1』（庵監修 2010）の試行版を試用していただいた国立地域のボランティアの方の感想をもとに、本研究と地域日本語教育の関わりの可能性について述べたい。

『にほんごこれだけ！1』は大人の学習者の使用を想定しているが、今回の試用は子どもたちを対象としたものである。まず、ボランティアの方から寄せていただいた感想を引用してみたい。

私たちが活動しています日本語の充分ではない子供達への日本語支援活動「ハンズ」はテキストをひろげて日本語を勉強するというスタイルの日本語教室ではありません。外国から転校して来た子供達は、学校ではクラスの生徒と離れて別の部屋で日本語指導をうけ、教室に戻るといわゆる学校ルールという異文化の中でストレスがたまっていますので、ハンズでは自然体で遊びながら、同じことばを繰り返し耳で聞き、口から発声させるということに力を注いでいます。このような活動の中で『にほんごこれだけ！』はありがたい教材でした。子供達は出欠をとったり、来ることに義務づけられるということはありませんから、まず、本のページを順番に追ってやらなくてもよいという事は、とてもありがたいです。

2で述べたように、『にほんごこれだけ！1』は「おしゃべり型の教育」を目指したものである。これは大人の学習者を想定したものであったが、上記の感想を見ると、年少者教育においても十分その特長を活かせるものと考えられる。

このように、『にほんごこれだけ!』によって具現化される「やさしい日本語」という取り組みは、大人と子どもの違いを超えて、地域日本語教育にとって一つの方向性を示しているものと考えられる。

参考文献

- イ・ヨンスク (2009) 「外国人が能動的に生きるための日本語教育」『AJALT』32、pp.10-13
- 庵功雄 (2009) 「地域日本語教育と日本語教育文法」『人文・自然』3、pp.126-141、一橋大学
- 庵功雄監修 (2010) 『にほんごこれだけ! 1』ココ出版
- 庵功雄・岩田一成・森篤嗣 (2009) 「やさしい日本語」を用いた公文書の書き換え」『2009 年度日本語教育学会秋季大会予稿集』 pp.135-140、日本語教育学会
- 尾崎明人 (2004) 「地域型日本語教育の方法論的試論」『言語と教育』 pp.295-310、くろしお出版
- 国際交流基金・日本国際教育支援協会 (2002) 『日本語能力試験出題基準 [改訂版]』凡人社
- 国立国語研究所 (1984) 『日本語教育のための基本語彙調査』秀英出版
- 国立国語研究所 (2004) 『行政情報を分かりやすく伝える工夫に関する意識調査 (自治体調査)』国立国語研究所 研究プロジェクト「日本語の現在」意識調査グループ
- 小林ミナ (2009) 「文法研究と文法教育」小林ミナ・日比谷潤子編『日本語教育の過去・現在・未来 第5巻文法』凡人社、pp.17-38
- 小山悟 (2007) 「著者との対談」『J BRIDGE FOR BEGGINNERS vol.1』凡人社
- 佐藤和之 (2004) 「災害時の言語表現を考える」『日本語学』23-8、pp.34-45、明治書院
- 土居光知 (1933) 『基礎日本語』六星館
- 土居光知 (1943) 『日本語の姿』改造社
- 日本語教育学会 (2008) 『外国人に対する実践的な日本語教育の研究開発 (「生活者としての外国人」のための日本語教育事業)』日本語教育学会
- 日本語教育学会 (2009) 『外国人に対する実践的な日本語教育の研究開発 (「生活者としての外国人」のための日本語教育事業)』日本語教育学会
- 野田尚史編 (2005) 『コミュニケーションのための日本語教育文法』くろしお出版
- 野元菊雄 (1993) 「簡約日本語語彙の意味分野」『日本語学』12-5、pp.40-48、明治書院
- 林四郎 (1971) 「語彙調査と基本語彙」『国立国語研究所報告 39 電子計算機による国語研究 III』 pp.1-35、秀英出版
- 松田真希子・児玉茂昭・竹元勇太・石坂達也・森篤嗣・川村よし子・山本和英 (2010) 「コーパスの異なりと単語親密度を活用した日本語共通基礎語彙の抽出」『言語処理学会第16回年次大会予稿集』 pp.579-582、言語処理学会
- 水野義道 (2006) 「災害時のための外国人向け「やさしい日本語」」『月刊言語』35-7、pp.54-59、大修館書店

- 山田泉(2002)「第8章地域社会と日本語教育」『ことばと文化を結ぶ日本語教育』, pp.118-135、凡人社
- 米勢治子(2002)「地域社会における日本語習得支援—愛知県における活動—」『日本語学』21-6、pp.36-48、明治書院
- Ogden, C. K. (1932) *The ABC of Basic English*. London: Kegan Paul. [高田力訳 1936 『ベーシックのABC』 研究社]
- Young-Sook Hwang, Taro Watanabe, & Yutaka Sasaki. (2005) Empirical Study of Utilizing Morph-Syntactic Information in SMT. *Natural Language Processing –IJCNLP 2005*, pp.474-485, Springer.
- Kishore Papineni, Salim Roukos, Todd Ward, & Wei-Jing Zhu.(2002) BLEU: a method for automatic evaluation of machine translation. *Proc. of ACL2002*, pp.311-318

使用した言語ツール

- [1]GIZA++ <http://www-i6.informatik.rwth-aachen.de/colleagues/och/software/GIZA++.html>
- [2]SRILM toolkit <http://www.speech.sri.com/>
- [3]MOSES <http://www.stamt.org/moses/20/02/2008> version
- [4]ChaSen <http://chasen-legacy.sourceforge.jp/>

〔謝辞〕5節について有益なご助言をいただいた NHK 放送技術研究所の田中英輝氏に感謝の意を表す。また、付記に関しては国立市在住の新井由利子氏から寄せていただいたご意見を参考にし、引用させていただいた。あわせて感謝の意を表す。

(いおり いさお 国際教育センター准教授
いわた かずなり 広島市立大学
つつい ちえ フェリス女学院大学
もり あつし 国立国語研究所
まつだ まきこ 金沢大学)

* 本研究は平成 22 年度～25 年度日本学術振興会科学研究費補助金基盤研究 (A) 「やさしい日本語を用いたユニバーサルコミュニケーション社会実現のための総合的研究」(課題番号 22242013) (研究代表者: 庵 功雄) の成果の一部である。