

パネル討論 2

制限言語と言い換え：
多言語展開の改善へ向けて

影浦 峯

東京大学 教授

制限言語と言い換え：多言語展開の改善へ向けて

影浦峽[㊦], 宮田玲[㊦], Anthony Hartley[†], Cécile Paris[‡]

㊦ 東京大学 † 東京外国語大学 ‡ CSIRO, Australia

1 言語とテキスト／文書

最初に、言語表現を前にしたときに、二つの大きく異なる見方があることを確認しておきましょう。それらは、フーコーの言葉を借りるならば、次のようになります。

[...] 言語 [ラング] とは、無限数個の言語運用 [パフォーマンス] をゆるす有限な規則の集合であるのだ。これに対して、言説とは、じっさいに述べられた言語的まとまり [シークエンス] のみからなる、つねに有限で現働的 [アクチュアル] に限定された集合である。[...] 言語の分析が、ある言説の事実に関して問う問いとはつねに、「どのような規則にもとづいて、このような言表がつけられたのか？ またしたがって、どのような規則によって他の同じような言表をつくることができるのか？」であるのに対して、言説の記述が立てるのはまったく別の問いであって、「このような言表が出現した、しかも、他のいかなる言表もその代わりには出現しなかったのは、どのようなわけなのか？」という問いなのだ [1]。

いわゆる言語学は前者のように言語を見ているとともに、有限の語彙と規則による無限の適格文の生成を考えるというチョムスキーの言葉に代表されるように、文（あるいはそれより小さな単位）を中心的な対象に据えます（もちろん「テキスト言語学」といった領域もありますが、例外的です）。

一方、言説の記述は、一般的に言語現象の単位としてテキストや文書が起点となります。フーコーの研究では（少し単純化して言うと）テキストや文書の集合が扱われますが、テキストや文書を構成する要素として命題を構成する基本単位である文もまた、言説の記述において対象となる資格を有します。言説を扱う場合、言語学のように背後のメカニズムを探るというよりも存在の形を捉えることが主な課題となることが多

いのですが（そのことは上の引用における「分析」と「記述」という言葉の使い分けにも反映されています）、アプローチの仕方によっては、テキストや文書を生成する規則を考えることもできます。ただしその場合でも、言語の研究のように共時的な体系が許容する潜在的に無限の適格な表現を対象とするのではなく、いわば現実的存在可能性の境界を探ることが主な課題となることは、テキストや文書の性質を考えるならば、明らかです。私たちは、その場で適格な文を作ることができますが、その場で適格な「テキスト」や「文書」を試しに作ることは、その性格上困難である（あるいは定義によっては論理的に不可能である）からです。

2 言い換えと制限言語

私たちが「言い換え」と呼ぶ操作は、どこに位置づけられるものでしょうか。「何か」を「何か」に言い換えるという「言い換え」の基本的な性格を踏まえるならば、言い換えは常に、ある一定の言い換えられるべき「何か」すなわち「言語的まとまり」が「じっさいに述べられた」のち、あるいは「じっさいに述べられ」つつあるときに作動する操作であることが確認できます。その意味で、いわゆる言語学が想定するような、適格文の背後にある有限の語彙と規則と同様のレベルに位置づけられるものではなく、ある具体的な言語表現が実際に出現したとき、あるいは出現するときに作用する、むしろ言説の分析に近いところで捉えられるべき現象であるということが出来ます。

では、「制限言語」についてはどこに位置づけられると考えるべきでしょうか。これについては、二つの可能性があります。制限言語の規則が人工的な言語のように、適格文を生成する語彙と文法を絞り込むかたちで設定され、そのようなかたちで用いられることが想定される場合には、それらは言語学が捉えるような言語の規則に対してさらなる制約を課すものとみなすことができるでしょう。それに対して、制限言語の制限を、具体的な言語表現が作られつつあるとき、あるいは作られたあとで適用するというかたちで定義する

ならば、これは言説の分析に近いレベルで制限言語の位置づけを設定していることとなります。

本パネル討論のテーマは「言い換えから見た日本語」であり、その中の一つの話題提起として「制限言語と呼ばれる規則による言い換え」が設定されています。「言い換え」が「じっさいに述べられ」たり述べられつつあるときに作動するものであることを踏まえるならば、パネルのテーマの枠組みで捉えられる「制限言語」もまた、対応したかたちで、言説の分析に近いレベルで位置づけておくことが妥当です。私たちの研究は、幸いにして（というのも最初にお話しがあったときには、パネル全体の枠組み設定をきちんと把握しないままお引き受けすることにしたので）、明示的にそのように位置づけることが可能です。以下、少し私たちの具体的な研究内容に言及しつつ、この点を敷衍していきたいと思います。

3 テキスト／文書と制限言語

「言い換え」が具体的な言語表現が生成されたあとあるいは生成されるときに操作であることを確認し、したがってそれはいわゆる言語学における言語の捉え方よりもむしろ「言説」の研究において採用される視点に近い視点で言語表現を見るものであることを確認し、さらに「制限言語」もそのレベルで捉えることになったので、ただちに、次のような視点が出てきます。すなわち、言い換えや制限言語の規則は、これまで基本的に文あるいはそれよりも小さい単位において定義されてきたが、本来、それらの操作が具体的な言説に即して認識される類のものであるならば、制限言語および言い換えは、具体的な言説の基本的な単位であるテキスト／文書において定義されることが適切である、という視点です。

実際、私たちがテクニカルライティングと言ったりする行為やそこで課される規則、あるいは、他の人が書いた文章を添削すると言った行為は、言い換えの典型的な具体例を伴うとともに言い換えが具体的に発動する典型的な場であり、その中では、例えば段落全体を書き換えたり、順序を入れ替えるといった操作も行われること、また、文や句を対象として言い換えがなされるとしても、その操作は常にテキスト／文書を与える文脈が前提とされていることを観察するならば、ここで述べてきたような理論的検討を経ずとも、言い換えなり制限言語の適用なりといった操作が、基本的に現働的なテキスト／文書をその本質的な対象としていることは明らかです。

とはいえ、言い換えでも制限言語でも、事前に何ら

かのかたちで規則化しておくことを想定するならば、大きくても文の単位までに対象を制限しておかないと難しいという点も、もう一方では明らかです [2]。テキストや文書の言い換え規則や制限規則というのは、どうも文レベルまでで定義される規則のようにきちんと適用できない感があります。テキストの規則が、適格テキストと非テキストを区別できると主張することは非現実的であることと、それは対応しているようです。では、テキスト／文書のレベルを、制限言語／言い換えの問題に技術的にどのように取り入れるか、それが問題となります。

4 制限言語と文書構造

4.1 基本的な枠組み

私たちは現在、自治体文書の多言語化支援システムの研究開発を進めています。ここでは、制限言語の考えを、上のような枠組みのもとで、組み込んでいます。私たちがこの研究で採用しているアプローチは上述の視点に立つならば、極めてオーソドックスなものです。基本的には、限定された範囲の現働的なテキスト／文書の生成環境を構築し [3]、その中に、文書構造に依存したかたちで文を基本単位とした制限言語の規則を埋め込む、というものです [4]。

現在の研究では、自治体の手続き文書に対象を限定しています。このように範囲を絞るならば（もしかするともっと一般的なタイプを扱えるかもしれませんが）、実際に存在する文書あるいは存在しうる文書の基本的な構造をかなり具体的に指定することが可能となります。その構造を、私たちは DITA [5] を用いて定義しています。手続き文書は、DITA のトピック型のうち、タスク・トピックが中心となります。DITA タスク本文の各要素を自治体手続きに合わせて具体化したものが、表 1 です [6]。

このような文書構造を指定することで、自治体手続き型文書において「何をどのような順序・構成で書けばよいか」の型が与えられます。

次に、「何をどのような順序・構成で書けばよいか」を受けて「具体的にどのように言語表現を作ればよいか」という言語表現の形を指定していく必要が出てきます。上で述べたように、このレベルでの制限は、基本的に文までの単位に対してしか有効なかたちでは設定しにくいので、ここでいわゆる制限言語の規則を導入します。私たちが以前作成した一般的な制限言語規則 22 個については既に報告していますが [7]、現在は、自治体文書の言語表現を検討し、「法／モダリティ」、「構造」、「語彙」、「テキスト／表記」の 4 つのカテゴリ

表 1: DITA タスク本文の具体化

DITA(初期設定)	詳細機能要素
事前条件 (prereq)	個人条件 イベント条件 アイテム条件
背景情報 (context)	説明 (概要, 目的, 効力, 罰則, 関連概念)
手順 (steps)	必要なものを持参する 申請場所へ行く 様式を提出する (手数料を払う)
期待結果 (result)	得られる結果 (所要期間, 交付物, 連絡)
実行例 (example)	[不要]
タスク完了後の操作 (postreq)	関連手続きへの誘導

りにおいて、合計 55 個のコントロールすべき表現属性を設定しています。例えば、「法／モダリティ」の属性としては、命令形、禁止、可能、時制、態などがコントロールすべき表現として設定されています。

最後に、制限言語の規則の適用を、文書構造の機能要素によって条件付けます。すなわち、どの機能要素について書くときにどの制限言語規則が適用されるかを指定します。

文書構造の指定はいわゆる「言い換え」ではありませんが、仮にある特定の手続き文書のある特定の方法で書くことがある文書作成者の頭の中にあり、その文書作成者が私たちの指定する制限のもとで文書を書くときに頭の中の文書に対して変更を求められるかたちで制限が働いているとするならば、抽象的な意味でこれも「言い換え」と言うことはできるかもしれません。そしてそのレベルでの「言い換え」は、文より下の、制限言語として通常表現されるレベルでの「言い換え」と、人間の文書作成においては自然に繋がっています。ここで説明した枠組みは、従って、そうした人間の文書作成に見られる広い意味での「言い換え」（日常的な言葉では、編集とか推敲と言った方がよいかもしれません）を、文書レベルでの構造の制約と文レベルでの制限言語規則の適用により導くものであると見ることが出来ます。

4.2 可読性と多言語展開の改善

私たちが提案した枠組みは、基本的に、日本語文書の可読性の向上（主に、手続き文書に必要な情報の集合を明示的に定め、論理的な順序で提示する文書レベルでの枠組みにより）と、多言語展開の精度改善に資

すると考えています。少し説明を補足します。

例えば、

身分証明書を持参する。

という文を考えてみます。これを MT にかけてみると、Google translate では

I bring identification.

となります。この表現が、文書中の手続きを実行する手順の一つとして使われているときには、MT が“Please bring your identification”あるいは“Bring your identification”と訳してくれると嬉しいですし、表題であるならば“To bring identification”等と訳すことが望まれます。

ここで、手順の一つとして使われているという DITA が提供する情報をもとに、

身分証明書を持参しなさい。

に書き換えると、Google translate の出力は、

Please bring your identification.

となり、手順の一つとして完全に意味が通じるようになります。一方、表題としては、この表現を「身分証明書の持参」とすると出力は

Bring identification

となり、まあそれなりに意味が通じるようになります。

ところで、手順の一つを表現する日本語としては、「身分証明書を持参する。」が普通です。従って、前者の例では、日本語文書の側は言い換ええないものをそのまま、MT の入力としては言い換えられたものを用いるのが最もよいこととなります。

ここでは、二つの異なる場での「制限言語」の適用が問題になっています。

- 第一に、言い換えあるいは制限言語の適用は、その表現がどのような機能を担うかによって異なっていること。とりわけそれは、異なる言語への変換において明示的になること。
- 第二に、起点言語としての日本語の可読性を維持するために言い換えあるいは制限言語の適用が有効である場合と、翻訳のために有効である場合、その両方に有効である場合、一方が他方を阻害する場合があります。

実は、上で、私たちが設定した新しい制限言語の属性を「規則」ではなく、「コントロールすべき表現」と

したのは、これらの点とも関係しています。言い換えるとMTの出力は良くなるけれども日本語では元の表現を維持する方が望ましい場合には裏で言い換え規則を自動的に適用することが適切で、私たちは現在、その規則も設計しているところです。その際に、文の機能要素に応じた自然な範囲での日本語表現が制約のもとに書かれていれば背後での言い換えも定式化しやすいというメリットがあります。

ここまで話を進めると、文書執筆における制限言語の適用と、背後でMTの出力を改善するために適用する言い換えは、作業／処理ステップとしては分かれることとなります。しかしながら、それは見かけ上のことで、理論的な枠組みとしての、具体的なテキスト／文書の内容を基盤として「制限言語」あるいは「言い換え」を考える、ただし、技術的な「制限言語」「言い換え」の適用は文以下を基本単位とするのでその操作をテキスト／文書に結びつけるために文書の機能要素を具体的に絞った文書タイプにおいて定義し活用する、という点は背後で一貫しています。

5 テキスト／文書生成の環境と場

話が中途半端に技術的になりました。「言い換え」や「制限言語」が、「じっさいに述べられた」あるいは「述べられ」つつある「言語的まとまり」からなる、つねに有限で現働的に限定された言語表現を対象とする、という理論的な把握は、もうひとつ、応用においても重要な視点を含んでいます。すなわち、

「言い換え」や「制限言語」をめぐる処理の適用は、「なんでもありうる」潜在的に無限の言語表現を想定して、このような技術があるから、可能性としてはこのようなときに使える、誰かが使ってくれるだろう、というオープンなかたちで提供されるべきものではなく、つねに有限で限定された言語表現を対象とし、それが「述べられ」つつある場を現実のものとして設定し、そこに関与／貢献するかたちで提供されなくてはならない、従って、提供の環境の中で要素技術は考えられなくてはならない、

という点です。

そこで私たちの研究では、第4節で述べた技術を具体的に適用しつつ手続き文書の作成ができるようなオーサリング環境の開発も同時に進めています。現在は、直接、特定の自治体と共同して作業は行なってい

ませんが、それについてもまもなく開始する予定で技術側の準備を進めているところです。

謝辞

本研究の一部は、KDDI財団2014-2016年度調査・研究助成「自治体文書の多言語化支援システムの開発」の助成を受けて行われました。また、共同研究を進めるにあたり、日本学術振興会外国人招へい研究者（短期）2014年度事業の助成を受けています。

文献等

- [1] Foucault, M. (1968) “Réponse au Cercle d'épistémologie”, *Cahiers pour l'analyse*, 9: *Généalogie des sciences*, pp. 9–40. [石田英敬訳「科学の考古学について〈認識論サークル〉への回答」『ミシェル・フーコー思考集成 III 1968–1970』（東京：筑摩書房）]
- [2] 井佐原均ほか. (2012) 「企業の多言語情報発信を支援する取り組み：国際化をにらんだ産業文書の効率の作成へ向けて」言語処理学会第18回年次大会, pp.369-372.
- [3] Hartley, A. F. and Paris, C. (1997) “Multilingual document production: from support for translating to support for authoring,” *Machine Translation - Special Issue on New Tools for Human Translators*, 12, pp. 109–128.
- [4] Hartley, A. (2010) “Enabling multilingual applications of 'controlled language': the DITA framework,” *Asia-Pacific Association for Machine Translation Journal*, 48, pp. 15–18.
- [5] Carey, L., Schlotfeldt, M. and Bellamy, J. (2012) *DITA Best Practices: A Roadmap for Writing, Editing, and Architecting in DITA*, IBM Press.
- [6] 宮田玲ほか. (2015) 「機械翻訳の活用を見据えた文書構造と言語表現の対応づけ—自治体手続き型文書を対象とした予備的報告—」言語処理学会第21回年次大会, 予定.
- [7] Tatsumi, M. et al. (2013) “Towards acceptable quality machine translation without post-editing for municipal websites: an evaluation of Japanese controlled language rules,” *MT Summit XIV: QTLaunch-Pad Workshop on Human-Centric Machine Translation and Evaluation*, 2013.