言い換え処理を適用した文書検索

“言い換える”技術で賢く検索する

人間が発する自然言語(文や発話)では，ほぼ同じ意味内容を伝えようとするのにも様々な表現形態があります。この様々な表現どうしを“言い換え”(paraphrase)と呼びます。言い換え処理は，高度に知的な内容が含まれることから，難しい課題とされてきました。しかし，最近では言い換え処理技術の研究が盛んになり，種々の観点から地道な研究が続けられています。

東芝ソリューション(株)は，言い換え処理技術と，これを利用した検索処理の研究開発に取り組んでいます。言い換え表現を生成する技術と表現の類似度を判定する技術により，従来よりも意味の理解に踏み込んだ文書検索の実現を目指しています。

“言い換え”とは

“言い換える”という単語は，一般的には，ある表現を別の表現で“言い直す”という意味で用いられることも多いですが，ここで，“言い換え”とはもう少し広く，人間(や機械)が，ほぼ同じ意味内容を伝えようとするときに発する様々な表現の間にある関係を指します。これは別のことばで“換言”，“パラフレーズ”などと呼ばれることもあり，**図1**のような言語現象が知られています。例えば，能動態の表現と受動態の表現の関係(図1の①)や，語い(用語)の置換え(同③)も言い換えにあたります。

文書検索においては，言い換えに対処することで検索漏れを防いだり，検索結果を絞り込んだりすることができます。

言い換え処理技術

言い換え処理技術の研究には，**図2**に示すように，二つの方向性があります。一つは，ある表現を入力として，それを別の表現に言い換える技術，すなわち，言い換え表現の生成技術の研究(図2(a))です。また，もう一つは，文書などに含まれている表現の中から，ある条件の下，言い換えと見なせる表現を見つける技術，すなわち，言い換えの発見技術の研究(図2(b))です。

(a)においては，言い換え規則などを用いて言い換え表現を生成し，生成された表現が言い換えとして妥当であるかを評価する，というアプローチが多く用いられています。

一方，(b)においては，二つの表現の間の共通性を用いて言い換えの候補を発見し，候補となる表現間の対応関係を調べることで言い換えであるか否かを判断する，というアプローチが多く用いられています。

言い換え処理を利用した文書検索

東芝ソリューション(株)は，この二つのアプローチを融合して，言い換え処理を文書検索に適用する研究に取り組んでいます。当社が開発している文書検索は，次の3段階の処理で構成されています(**図3**)。

(1) 言い換え候補の発見(図3の①)
検索のための入力文(検索要求文)を構成する単語と類似する単語から成る文を検索し，これを言い換え候補とします。

(2) 言い換え表現の生成(図3の②)
言い換え候補となった文に対し，言い換え規則(**図4**)を適用して言い換え表現を生成します。
このとき，入力文を制限条件として用い，言い換え表現が多数生成されることを防いでいます。
また，言い換え処理を再帰的に実行することにより，言い換え規則を簡素化しています。

(3) 比較処理(図3の③)
言い換え候補から生成された表現と入力文を比較し，言い換え表現であるか否かを判断します。

言い換え処理を利用することにより，従来よりも意味の理解に踏み込んだ文書検索を行うことができます(**図5**)。

図5の入力例1では文書(A)が，入力例2では文書(B)が，本来望ましい検索結果です。従来の検索では，入力例1に対し，文書(A)，文書(B)が共に検索されますが，言い換え処理を利用した検索では，文書(A)だけが検索されます。文書(A)から生成された言い換え表現と入力例1とは同じ意味を伝える表現であると判断され，文書(B)から生成された言い換え表現と入力例1とは同じ意味ではないと判断されるからです。

一方，入力例2に対しては，文書(B)から生成された言い換え表現と入力例2とが同じ意味と判断されるので，文書(B)が検索されます。

言い換え処理の適用分野

言い換え処理には，文書検索をはじめとする情報検索のほか，次の例に示すような幅広い適用分野があります。

(1) 人間の読解を支援するためのわかりやすい表現への言い換え

(2) 限られた表示スペースのための短い表現への言い換え

(3) 人間が書いた文と機械が処理しやすい表現の結付け

言い換え処理は，解析的観点に加えて，生成的観点で言語を処理することに特徴があります。言い換えの研究は，様々な個性豊かな表現はどのように生成されるか，また，その様々な表現をどう理解するか，という視点につながっています。今後とも，言い換え処理の利用により，高機能・高精度な文書検索の実現を目指していきたいと考えています。

齋藤　佳美
東芝ソリューション(株)
IT技術研究所研究主務

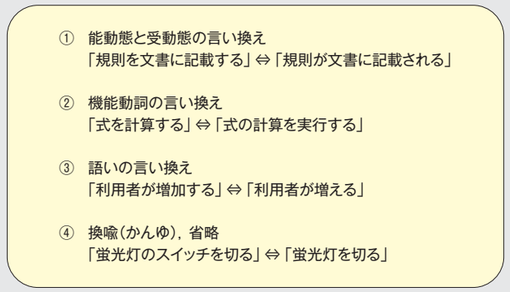

図1. 言い換えの例 — 様々な言い換えの現象があります。

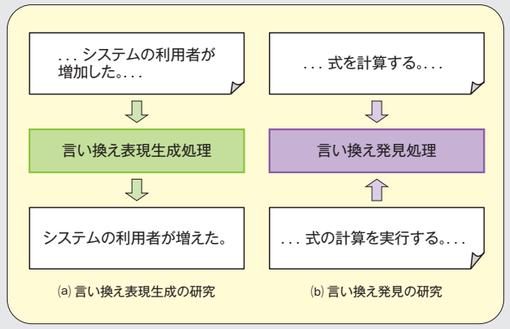

図2. 言い換え処理技術 — 言い換え処理技術には表現生成と発見の二つの方向性があります。

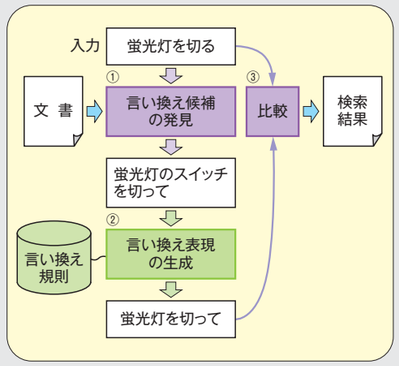

図3. 言い換え処理を利用した文書検索 — 候補の発見，表現の生成，及び比較の3段階の処理で文書を検索します。

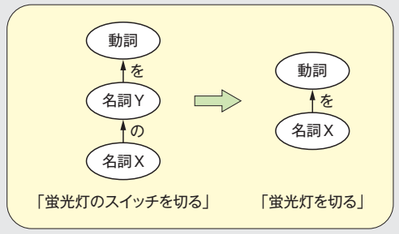

図4. 言い換え規則の例 — 規則を適用して言い換え表現を生成します。

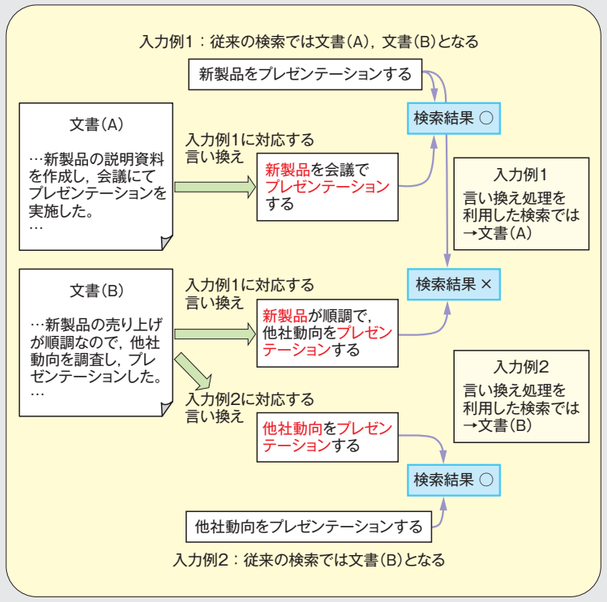

図5. 文書検索結果の例 — 従来に比べ，意味の理解に踏み込んだ検索を行います。