

5 S-9

科学技術文書のためのテクニカルライティング -並列表現と修飾表現が存在する場合-

木村 真理子, 平川 秀樹, 天野 真家
 (株) 東芝 総合研究所

1. はじめに

テクニカルライティング(以下TWと呼ぶ)の目的は、理解しやすい文で情報を正しく伝搬することである。その手段としては、ある種の制限が加えられた言語(制限言語)が有効である。[1] その制限言語が持つべき性質を以下に示す。[2]

- 記述力があること
- 曖昧性がないこと
- 読み手に理解しやすいこと
- 書き手に負担がかからないこと

機械翻訳の前編集におけるメタ記号(並列句・挿入句の範囲を指定する)の挿入など、現在行われている(広義の)TWにはc)の性質に該当しない場合もある。しかし、この種のアプローチは、その不自然さの為、人間の読者を対象にした科学技術文書のTWとしては不適格であると考えられる。

上記のTWの枠組みに加えて、実際にTWを行う際に要求されるのがTW規則の明確性・整合性である。一般に自然言語の全ての場合についてTWの規則を規定するのは困難である為、TWの規則の記述は概して指針的であり、行うべき処理が明確にされていない。また、規則同士が無矛盾であるという保証が明示的にはなかった。例えば、「、」には複数の用法がある為、複数の規則によって打たれるべき位置が指定された場合、これらが一致するとは限らない。すなわち、任意の概念を表そうとする時、使われるべき表現を決定する方法を確立する必要がある。

我々は以上の点を踏まえて、1)言語の持つべき性質を備えた制限言語に基づいた、2)アルゴリズム的に表現を導くことができる、TWを目指している。本報告では、並列表現と修飾表現が共起する場合のTWについて述べる。

2. テクニカルライティング

TWによる表層の生成について、その際使われる制限言語の性質も含めて説明する。図1に一般的なTWのモデルを示す。

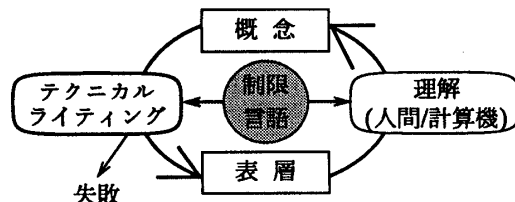


図1 テクニカルライティングのモデル

TWは概念を制限言語によって表層に変換する。人間や計算機は、生成された表層を制限言語の知識に基づいて理解する。この時、理解とは表層をその元

となった概念へマップすることである。また、日本語の枠内でのTWは失敗することがある(注)。

1. で挙げられた言語の性質は、制限言語に対しては以下のように適用される。

- 記述力がある $\Rightarrow$ 広いカバレッジを持つこと
「カバレッジ」 $\equiv$ 実文で表現されている概念の何%が表現できるか
- 曖昧性がない $\Rightarrow$ 表層を一意にマップできるようなアルゴリズムが存在すること
- 理解しやすい $\Rightarrow$ 自然な表層が生成できること
「自然さ」 $\equiv$ 実文の何%が制限言語を使って生成された表現と一致するか
- 負担がかからない $\Rightarrow$ 明確なアルゴリズムが存在すること

3. 複雑な並列表現のためのテクニカルライティング

3.1. 複雑な並列表現

実文において、並列表現と修飾表現は頻繁に共起する。その際に発生する曖昧性(並列句の係り先、並列句に係る修飾句の係り先)を解消する為には並列表現と修飾表現の両方を相互的に考慮せねばならない。しかし、並列表現と修飾表現との関係が確立していない為、正しい解釈を得ることは人間にも機械にも困難である場合が多い。

3.2. 接続詞・修飾句の優先順位

複雑な並列表現のためのTWは、並列接続詞と修飾句の相対的な優先順位を制限言語の規則として利用する。この為、3.1.で述べた曖昧性についてはb)の性質を満足することとする。また、c)の性質を考慮し、優先順位は実文に即して決定した。表1にその結果を示す。

並列接続詞	修飾句
「と、」 「、」「、および」 「(呼応の)と」 「ならびに」 「および」「と」 「・」	「～の、」「～する、」 「～の」「～する」

表1 並列接続詞・修飾句の優先順位

この結果は[3]が報告した並列接続語間の優先関係の調査結果に合致している。ここで特徴的であるのは、以下の二点である。

- 優先順位決定においては、同じ語であっても句点の有無(「と」と「と、」「および」と「

(注) 例えば、並列要素が n 個ある時、その解釈の可能性はカタラン数個ある。しかし、句点を使って表せる表現は 2^{n-1} 個である。この為、句点の付与では一意に解釈を表すことは不可能である。

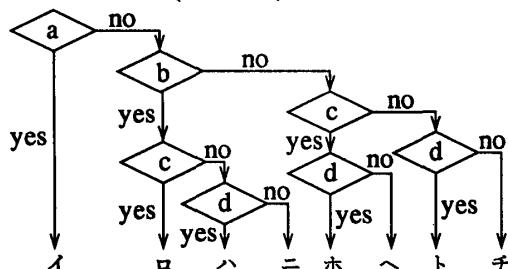
および」)によって異なる要素として扱ったこと
理由: 句点には、並列接続詞としての役割と、その直前に位置する修飾句の係り先を制限する役割とがある為

- 呼応の「と」(「～と～」のように、助詞「と」と「と」が呼応している場合の、後方の「と」)を一般の接続詞「と」とは区別したこと
- 理由: その並列要素が明確である(「と」と「と」に挟まれた名詞句)為 [3]

3.3. アルゴリズム

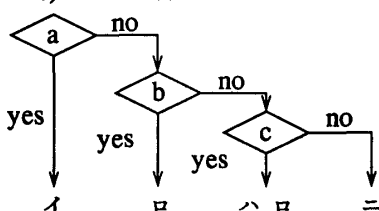
以下に並列接続詞と修飾語の表層表現を選択するためのTWのアルゴリズムと、それから導かれる表現の代表的なものを示す。我々は、並列要素の複雑さによって並列表現を3レベルに分けることにより、a)やd)の性質の実現を試みた。

i) 並列要素が単純(名詞のみ)である場合



- a. 上位レベルの並列句に含まれているか
- b. 並列句全体を修飾している修飾句はあるか
- c. 並列句全体として修飾しているか
- d. 並列要素は二項か
- イ: N・N・N
- ロ: ～するN・N・Nの
- ハ: ～するN(と/および)Nが
- ニ: ～する、N、N(、および/、)Nが
- ホ: N(と/および)Nの
- ヘ: N・N・Nの
- ト: N(と/および)Nが
- チ: N、N(、および/、)Nが

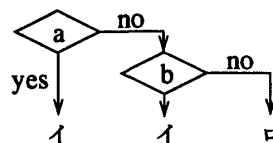
ii) 並列要素が中程度複雑(用言を含まない複数語から成る名詞句)である場合



- a. 上位レベルの並列句に含まれているか
- b. 並列句全体を修飾している修飾句はあるか
- c. 並列句全体として修飾しているか
- イ: N・N・N
- ロ: 挿入句化
- ハ: ～(と人)～((呼応の)と人)の
- ニ: ～(と人、人)～((呼応の)と人、人 および/ならびに)～(と/∅/∅/∅)が

iii) 並列句が複雑(用言を含み二語以上から成る修飾

句を持つ名詞句 / ii)より複雑な並列句)である場合



- a. 並列句全体を修飾している修飾句はあるか
- b. 上位レベルの複雑な並列句に含まれているか
- イ: 挿入句化
- ロ: ～と、～と

次に、例を挙げて説明する。

例) ストロークの深さと重さの変化

この場合、並列要素は「深さ」と「重さ」である為、使用するアルゴリズムはi)である。これらの並列句はより大きな並列の範囲に含まれていない(a)。しかし、「ストロークの」に修飾されていて(b)、「変化」を修飾している(c)。従って、アルゴリズムから得られた表現(ロ)は以下ようになる。

例') ストロークの深さ・重さの変化

3.4. 調査

アルゴリズムの評価として、1) 技術論文の実文約17,000文から接続詞「と」「および」「あるいは」を含んだ並列表現を抽出し、2) 1)で抽出された表現547文に対して、2.で定義した「カバレッジ」と「自然さ」を調査する。結果を下記に示す。

並列接続詞	失敗	記述力	一致	自然さ	合計
「と」	10	97	190	57	334*
「および」	7	96	74	39	188*
「あるいは」	0	100	9	36	25
	17(件)	97(%)	273(件)	50(%)	547

「カバレッジ」= ((表現の合計数)-(失敗))/100

「自然さ」= (アルゴリズムによって生成された表層と実文の表現とが一致した場合の数)/100

表2 調査結果 (*:部分)

「カバレッジ」は97%であり、「自然さ」は50%であった。

4. むすび

テクニカルライティングの在り方を検討し、並列表現と修飾表現が共存する場合について、表層表現を選択する為のアルゴリズムを提案した。その際使われる文法規則として、接続詞・修飾句の優先順位を実文の表現から見出した。

今後の課題としては以下の事項が考えられる。

- 1) 実文調査に基づく改良、調査時に明らかになった問題点の解決
- 2) 機械化
- 3) 他の言語表現への対応を目指した拡張

参考文献:

- [1] 長尾,田中,辻井:制限文法にもとづく文章作成援助システム,自然言語処理研究会,44-5,1984.
- [2] 吉田:日本語の規格化と制限日本語の設計,第1回「大学と科学」公開シンポジウム「日本語の特性と機械翻訳」,1987.
- [3] 長尾,辻井,田中,石川:科学技術論文における並列句とその解析,自然言語処理研究会,36-4,1983.