7C－5

テキストのwriting-style依存構造と冠詞

電子技術総合研究所
内田ユリ子

1　はじめに

書き手の思考のカオスがひとまとまりのテキストの形に言語表現されるまでの過程ではさまざまの知識を使いながら情報を構造化していくと考えられる。人間の場合は、思考そのものが特定の言語に基付いて行われている部分が多く、深部の思考から表層のテキストにまとまっていく過程におけるさまざまな知識の適用も、順次的というよりも高次並列的と考えられる。これに対し機械によるテキスト生成では、書き手の思考内容に相当する非言語依存的な意味表現から出発し、それにたいして言語的な知識や常識に相当する世界知識を適用しながら目的言語によるテキストを生成していく方が、現実的である。

また、人間の書いたテキストの場合、同じ内容に付いて書かれたテキストでも書き手のwriting-styleに依存する表現の違いがある。機械による生成においても生成のスタイルに基付いて、テキストの表現構造を組み立てていく。writing-styleと言うと、書き手によって異なる捉えどころのないもののように考えられがちであるが、この過程も場当たり的な処理ではなく、秩序立って支配している法則のある過程と考えられる。すなわち、非writing-style依存的なテキストの表現構造に対して書き手と読み手の間に共有していることを前提とするwriting-styleに関する知識（以下ではwriting-style知識）を使いながら、writing-styleに依存した構造（以下writing-style依存構造）を表現していくと考えられる。writing-style知識も構造化した知識であり、蓄積され共有され、テキスト解析にも生成にも共通に使われる知識である。

筆者らはこれまでの研究で、言語知識や世界知識と共に、現実の題材に即した実際的な知識として複数パラグラフよりなるテキストのパラグラフ構成に関する知識をテキスト生成に取り入れてきたが、このような知識もwriting-style知識として構造化した知識の一部として位置付けられる。

本稿では、writing-style知識とwriting-style依存構造のテキスト生成過程における重要性を指摘すると共に、テキストの意味とwriting-styleとの区別を知識表現のレベルの違いとして把握し、writing-style依存構造がテキストに反映している具体例として、実際の英語テキストを組立ているユニット（パラグラフ、文）間の構造と冠詞の関係に付いて検討する。

2　知識表現のレベルと英語定冠詞

これまでの研究で、テキスト生成に際し、冠詞の選択には英語辞書と文法の知識はもとより、背景知識に依存する場合もあることを前提として冠詞生成規則を定めてきたが、ここでさらに、writing-style依存構造と冠詞の関係に付いて考察し、法則を把握し、生成規則に反映させる必要がある。

Writing-style Dependent Structures and Articles in a Text
Yuriko UCHIDA
Electrotechnical Laboratory

そこで、writing-style依存構造に依存しない場合との対比を知識表現との関係関係で整理しておく。テキスト生成システムにおける冠詞生成では、まず第一に冠詞と組合わせる名詞概念の意味を知識表現に結び付けることが必要であり、特に定冠詞の生成には、概念の唯一性の根拠を知識表現として把握することが重要であると考えられる。これまでに、唯一的概念の意味を文脈表現構造における表現として明瞭化し、冠詞生成規則の結び付け、テキスト生成規則に生かしてきた。

大局的には、読み手と書き手の共有世界で指示対象が唯一化されている場合、その表現形式として、定冠詞を生成する。その共有世界には、テキスト内の情報のみならず、概念世界の知識も含まれる。

定冠詞生成のための、概念の唯一性と知識表現のレベルとの関係は次のように整理される。

（ａ）テキストの内容だけから、背景知識を使わずに、唯一的概念であることが決まる場合。たとえば、形容詞最上級に対応する表現がこれに当たる。これはテキストの中間表現のレベルだけで判断できる。

（ｂ）唯一概念であることを、話題内容に依存せずに常識的に判断すべき場合は、背景知識の集約である概念辞書の項目ごとに記述しておき、生成時にそれを参照する。たとえば、「太陽」や「地球」の恒常的唯一性はこれにあたる。

（ｃ）条件付き唯一性を持ち、話題内容がそれを満たしていることを常識的に判断すべき場合は、概念辞書のそれらの項目に唯一化条件を記述しておき、生成時にそれを参照し、中間表現がそれを満たすかどうかを判定する。

（ｄ）知識表現には唯一性が指定されていないが、話題の中では唯一的であり（たとえば、話題になっている一人の人物）、それまでの生成過程から、読み手と書き手との間で唯一的に指示が一致するとみなされ、照応表現として定冠詞を生成する場合。

（ａ）から（ｃ）の場合は、書き手に依存しない。話題と背景知識が決まれば、誰が生成しても定冠詞が生成される。それに対して、（ｄ）は生成過程に依存する。

本稿で扱う定冠詞生成は、上記４つのどれにも当てはまらない場合である。照応的である点においては（ｄ）似にているが、話題の中で唯一的ではない場合について扱う。すなわち、テキスト中間表現の上でも概念辞書の上でも唯一性が指定されていないが、定冠詞を生成する場合である。

3　writing-style依存構造の英語冠詞における発現

この節では実例を観察し、英語冠詞とwriting-style依存構造の関係について述べる。

唯一的でない存在を表す名詞に付いてはテキストの中に同じ名詞が繰り返し出てきても「同一対象の繰り返し指示」ではないので、２回目以後も定冠詞をとる必要はない。このような場合、不定冠詞が繰り返される（タイプ１　[a-a]）。

しかし、指示が一致していないにも関わらず、定冠詞に切り替わる事もある（タイプ２　[a-the]）。そしてこのままtheを繰り返して行く例もあるが（タイプ３　[the-the]）、一方、また不定冠詞aに戻ることもある（タイプ４　[the-a]）。これらの冠詞切り替えは、意味に基付くものではなく、writing-style依存構造に関係している事を、以下に実例を示しつつ述べる。。

資料１は、高校生向きの教科書の一部であるが、パラグラフの構造およびパラグラフ内の文間構造と冠詞変化に付いて考察する。以下、（ｐｎ－ｓｎ）は第ｎパラグラフの第ｎ文を表すものとする。

ここで名詞句（cell membrane）に着目すると、これは(p3-s3)で初出で,以下、(p3-s5)、(p3-s6)、(p3-s7),(p3-s9)に存在する。ここで（cell membrane)は特定の対象ではない。そこで、テキストの中に同じ名詞が繰り返し出てきても、指示が一致しているわけではないので、その意味では定冠詞をとる必要はない。ところが、冠詞は（a-the-the-the-a）と変化しており、s-5,s-6,s-7におけるtheは、見かけ上は照応的である。そして、そのtheの連鎖はs-9で切れている。

次のパラグラフでの（cell membrane）の出現は,(p4-s1),(p4-s3),(p4-s4)で冠詞は(a-the-the)である。また、その次のパラグラフでは、(p5-s1),(p5-s2)で(a-a)である。

これらの冠詞切り替えは、記述対象に固有な意味構造に依存するものではなく 、テキストの組立構造に依存したものと考えられある。その組立構造と冠詞との関係には次のような法則が見いだされる。

テキストはパラグラフから組み立てられており、また一つのパラグラフは文化ら組み立てられているが、パラグラフ及び文をユニットと呼ぶと、ユニット間構造と冠詞連続には次のような法則がある。

テキストのユニットの中に同一名詞句が２回以上現れる場合、その名詞句が同一の対象物を繰り返し指示しているのでなければ、冠詞連のタイプ１は、その名詞句が含まれているユニットの並列性を示す（法則１）。タイプ２、３はユニット連の直列性を示す（法則２）。タイプ４は、それまで直列接続してきたユニット連の切断を示す（法則３）。パラグラフ切り替えに際して、しばしばタイプ４が現れるのはこの為である。

以上のように、英語冠詞（ゼロ冠詞を含む）の果たす役割は、その冠詞を含む名詞句全体の意味に関係しているばかりでなく、テキストのwriting-style依存構造を示す信号ともなっていることが実例から観察される。

４　おわりに

英語テキストを観察するとき、「なぜその場所にその冠詞が発現しているか？」理解できないことがある。native speakers が必ずしも明確にして安定した回答がを持っているとは限らず、筆者にとっては謎の自然現象の一部である。当然、日本語テキストを英語に翻訳しようとするとき、原テキストには存在しない冠詞をひねりだす為のアルゴリズムのかなりの部分は霧の中にある。機械翻訳を背景にして考える時、冠詞用例集の山に挑む事は、重要ではあるがそこで解決されるものは一部であり、また「意味」がすべてを解決する訳でもない。そこで冠詞現象をテキストのwriting-style依存表現構造と結び付けて検討したところ、霧の一部が晴れた。

５　参考文献

［１］　内田他：実際的な知識に基付く文脈表現構造からの英語テキスト生成、電子情報通信学会論文誌 D-2　No.9　pp.1472-1483,1989

［２］　武田修一：英語意味論の基礎的研究、リーベル出版（1987）

［３］　Joke　Dorrepaal：Discourse　Anaphora、COLLING-90　vol.2　pp.95-99,1990

P3 S1 A human cell contains cytoplasm, a jellylike material with a unique property--it is alive.
S2 Cytoplasm is a mixture of water, salts, and other materials.
S3 A thin, living cell membrane surrounds the cell.
S4 It is visible only with the aid of an electron microscope.
S5 The cell membrane is flexible and is a living, functioning part of the cell.
S6 The cell membrane controls the movement of materials into and out of the cell.
S7 The cell membrane is selecive.
S8 Only certain dissolved materials can pass through.
S9 In general, very large particles cannot pass through a cell membrane.
P4 S1 Dissolved materials move through a cell membrane by a process called diffusion.
S2 Diffusion is the movement of particles of a material from regions of higher concentration to regions of lower concentration of that material.
S3 Food and oxygen diffuse from the blood through the eell membrane into the cytoplasm of the cell.
S4 Waste materials diffuse in the opposite direction, from the cell into the blood through the cell membrane.
P5 S1 A special kind of diffusion, called osmosis, also occurs through a cell membrane.
S2 Osmosis is the diffusion of water through a membrane.
S3 Water moves into and out of a cell by osmosis.
P6 S1 Life functions occur in each body cell.
S2 The combined function of many cells keep you alive.
S3 Your body has different kinds of cells that do different jobs.
S4 There are muscle cells, blood cells, nerve cells, and other kinds of cells.
S5 Some cells contain structures that perform special jobs.
S6 For example, cells that line your throat have tiny, hairlike cilia.
S7 Cilia filter out dust from the air you breathe.

図１　冠詞がテキストのwriting-style依存構造を示す信号となっている例