



数理の窓

「意味のある」文章のルール

人工知能は“意味のある文章”を書けるのか？

1つ目のハードルは、文法的に正しい文になるかだ。1950年代に言語学者のチョムスキーは、すべての文を生成するための変形操作やルールの研究を始めた。これは生成文法と呼ばれ、大きく句構造規則と変形規則でできている¹⁾。

変形規則には、平叙文から疑問文を作る「Wh-移動」や、能動文『SVO』から受動文『O be V+ed by S』を作る「受動変形」などがある。規則は英語や日本語など言語毎に異なり、50以上あるとも言われる。

一方、句構造規則は、単語の並びの規定と、記号から記号への“書換えルール”群だ。例えば

『猫が犬を好きになる』は、以下の構造になる。

- ①<文>→<名詞>+ “が” +<動詞句>
- ②<動詞句>→<名詞>+ “を” +<動詞>
- ③<名詞>→ “猫”、<名詞>→ “犬”
<動詞>→ “好きになる”

文は<名詞>などの“抽象記号”に書換えられ、さらに猫などの“具体記号”になったとみるのだ。

チョムスキーは、言語の表現力に応じて、書換えルールの複雑さが異なることを明らかにした。例えば、鳥の鳴き声は、直前の幾つかの音によってのみ、次の音が決まるため、音の繋がりルールは単純だ。しかし、人間の言語は、履歴の考慮、離れた単語が依存しあう「長距離依存」、名詞+動詞で句を

形成、句が集り大きな句を作る、埋込・入れ子の再帰構造など複雑だ。例えば『犬を好きになる猫が、踊る』を作るには、主語部分のルールとして

名詞句→<名詞>+ “を” +<動詞>+<名詞>が必要になる。人間の言語はかなり自由な書換えルールになっているのだ²⁾。ただし、完全に自由にすると、『猫が犬を踊る』も生成するので、一工夫必要だ。とはいえ、ルールを守れば、正しい文法の文を生成することは可能だ。

では文法が正しければ、人工知能に“意味のある文”は生成可能なのだろうか？ 単語と単語の繋がりから文の意味が通っている確率や“感情スコア”を計算する研究は進んでいる。また、文同士の関連を分析し、どの文や段落が重要か判定するツールもある。このように階層毎にある“暗黙ルール”の可視化は進んでいるものの、現段階では、これを守れば、意味が通る文章になるというスキームはない。

ここまでで文章を書く難しさが分かった。この文章自体があらゆる“ルール”に則っている。文法的にはおそらく正しいので、読者にとって「意味をなした」か願うだけで。(外園 康智)

- 1) 生成文法の理論は次々と変化し、この分け方は現在主流ではない。多くの理論の中でどれが正しいかの一致した見解はない。
- 2) チョムスキー階層で人間の言語は「文脈自由文法」以上となる。