

読み句の構造に基づく郷土かるた生成手法に関する研究

神澤 義人

近年、日本全国各地で地域活性化を目的とした活動が盛んに行われており、その一環として、郷土かるたの制作と普及がある。郷土かるたは、句頭が「あ」から「わ」で始まる読み句と、対応する絵札の組によって作られる、特定地域を題材としたかるたである。採用されている読み句は、いずれも対象地域の持つ特色や名所、縁のある人物について語感よく簡潔にまとめられている。本研究では、このような郷土かるたの読み句を地域固有の事物を示す名詞、特徴語について説明した短文の集合と捉える。郷土かるたには短歌や俳句の五、七、五のような、一定の音数規則は存在していない。本研究の目的は、例に挙げたような既存の郷土かるた読み句の文章構造を解析し、その結果を基にして、異なる地域を題材として新しい郷土かるたを生成することである。

解析対象とする既存の郷土かるたには、昭和22年に日本で初めて製作され、今でも高い普及率を誇る群馬県の上毛かるたを採用した。上毛かるたの各読み句について形態素解析を行った結果を基にして、かるた読み句のテンプレートファイルを作成した。かるたの読み句ごとに全体の音数、音の切れ目、読み句の内容、よみがな、また形態素ごとに形態素が持つ属性、語句、語句の音数を保存した。

読み句の生成は本研究で作成したテンプレートファイルから構造を取得して、そこへ新規語句を当てはめることで行なう。まず、新たに郷土かるたの読み句を生成したい地域と、その地域に属する、句の中心としたい固有名詞である特徴語を入力する。入力として与えられた特徴語と地域名から得られたweb検索結果から、上位に出現したスニペットを取得し、そこで頻出する語句を特徴語との関連がある語、関連語とみなす。その後、特徴語と関連語との間にある共起度を、シン普森係数を用いて算出し、上位に現れた関連語と特徴語とをテンプレートに沿って適用し、新規の読み句とした。

採用する関連語について、共起度の下限値を設ける、テンプレートの形態素が持つ音数と一致したもののみ採用する、というように、関連語について制約を設けた。その条件下での生成の結果、読み句として相応しくないような冗長な生成句や、特徴語との関連が薄い語句を選択せずに済み、結果として著しく生成句の絞り込みを行なうことができた。

今後は、テンプレートから参照する読み句の特徴について精緻化を行うとともに、語句の選出手法の改良、前後の品詞との共起を参照するなどの改良を行なうことにより、より自然的に特徴語について表現した読み句を生成できるよう、生成手法を発展させていく。

(指導教員 佐藤哲司)