意味カテゴリを用いたサ変動詞同音異義語誤り検定方式

5P－2

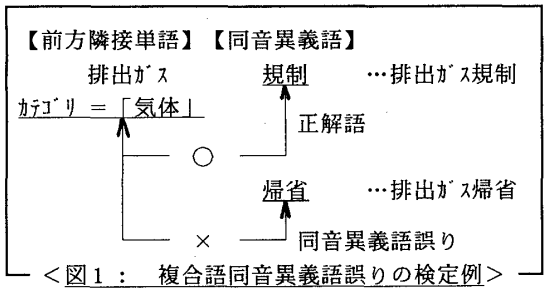
*奥　雅博　　**大橋　隆弘

*ＮＴＴ情報通信処理研究所　　**早稲田大学

1．はじめに

複合語に含まれる同音異義語誤りを検出する方式として、同音異義語とこれに隣接する単語の意味カテゴリとの連接関係を検定する方式を提案し、有効であることを示した[1]。この方式に従った複合語同音異義語誤り検定の例を図1に示す。検定対象となる同音異義語ごとの意味カテゴリとの連接可否情報は、意味連接辞書に記述している。

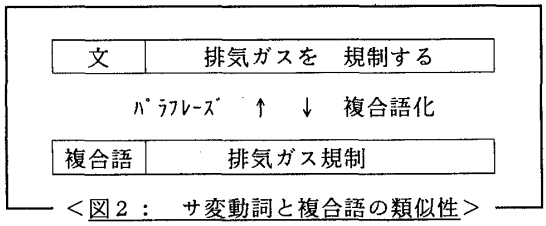
本稿では、この方式をサ変動詞同音異義語誤りの検出に適用した結果について述べる。



＜図1：　複合語同音異義語誤りの検定例＞

2．複合語同音異義語誤りとの類似性

複合語は複数の名詞が助詞を介さずに連接した表現であり、構成名詞間には種々の関係（並列関係、格関係、修飾関係など）が存在する。特にサ変名詞は造語力が強く、格関係によって他の名詞と連接して複合語を構成しやすい。従って、サ変動詞とこれを修飾する格要素とを結合することによって、1つの複合語を作ることができる（図2）。



＜図2：　サ変動詞と複合語の類似性＞

この考え方に基づくと、サ変動詞と格要素を結び付けて、あたかも複合語であるかのように扱うことによって、複合語同音異義語誤りの検出方式をサ変動詞同音異義語誤りの検出にそのまま利用できる。

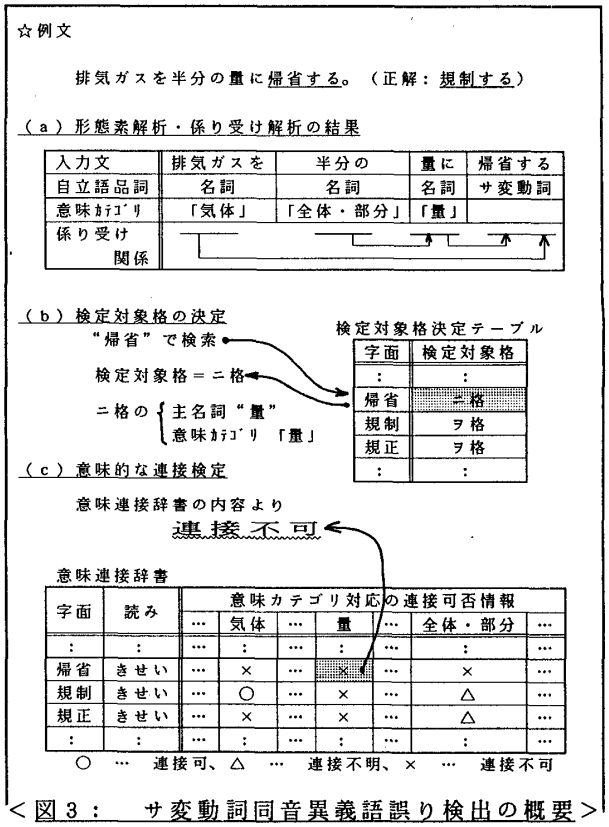
3．サ変動詞同音異義語誤り検出方式

3．1　方式の概要

本稿では上記の類似性に基づいて、複合語同音異義語誤り検定で用いた意味連接辞書（同音異義語の字面とこれに隣接しうる単語の持つ意味カテゴリとの連接可否を記述）を、サ変動詞同音異義語誤りの検定に用いる方式を提案する。

本方式の特徴は、入力文中からサ変動詞とこれに関係の深い1つの格要素（以下、検定対象格と呼ぶ）を取り出し、両者を結合することによってあたかも複合語であるかのように扱うことにある。検定対象格はサ変動詞同音異義語の字面ごとに異なる。ここでは字面ごとの検定対象格を検定対象格決定テーブルに記述している（検定対象格の決定方法については3．2節で述べる）。

図3に「排出ガスを半分の量に帰省する（正解：規制する）」の検定例を示す。



＜図3：　サ変動詞同音異義語誤り検出の概要＞

A Detection Method for Homonym Errors using Semantic Relation

*Masahiro OKU, **Takahiro OHASHI

*NTT Communications and Information Processing Laboratories

**Waseda University

まず、入力文に対して形態素解析、係り受け解析を行い、意味カテゴリの付与、同音異義語“帰省する”を修飾する格要素の決定を行う（図3（a））。次に、“帰省する”の検定対象格を決めるために、“帰省”で検定対象格決定テーブルを検索し、「検定対象格＝ニ格」を得る（図3（b））。入力文のニ格の名詞は、“量”であり、その意味カテゴリは「量」である。意味連接辞書中の“帰省”に対する意味カテゴリ＝「量」の連接可否は連接不可（×）であるので、“帰省する”を同音異義語誤りであると検定する（図3（c））。

3．2　検定対象格の決定

本方式では、どの格要素を検定対象格として選択するかが検定精度を決める1つの要因となる。複合語同音異義語誤りとの類似性に着目していることから、検定対象格として、以下の2つの条件を満足する格要素を選択することとした。

① 複合語において結び付きやすい格であること。
② 文中に対象となる同音異義語とともによく現れる格であること。

検定対象格の決定にあたっては、まず、複合語をパラフレーズすることによって①を満足する格要素を求め、次に、同音異義語をサ変動詞として含む文中においてこの格要素が②を満足することを確認するという方法を採用した。このようにして、各同音異義語に対して検定対象格を求め、検定対象格決定テーブルを作成した。

4．評価実験

本方式の有効性を確認するために評価実験を行った。実験には以下のデータを用いた。

・検定対象とするサ変動詞同音異義語

表1に示す7種の異なる読みを持つ19語

表1：　評価実験に用いた同音異義語

読　み	字　面
かこう	加工　下降
きせい	規制　寄生　帰省　規正
きょうそう	競争　競走
こうたい	後退　交代　交替
こうひょう	講評　公表
しょうきゃく	焼却　消却　償却
たいこう	対抗　対校　対向

・意味連接辞書

複合語同音異義語誤り検定に用いた意味連接辞書（新聞記事90日分から検定対象語を含む複合語を抽出し、検定対象語に隣接する単語の意味カテゴリを調査することによって作成）。

・検定対象格決定ルール

意味連接辞書の作成に用いた複合語をパレフレーズし、3．2節の手順に従って作成。

・実験データ

① 正解語データ　…　704文
新聞記事90日分から表1の同音異義語をサ変動詞として含む単文を抽出。

② 誤り語データ　…　1217文
①の正解語データ中の同音異義語をその誤り語に置き換えた単文。

誤り検出方式の評価は、正解語を正しいと指摘できる能力と、誤り語を誤りとして検出できる能力との2つの観点から行う必要がある。本稿では以下の2つの値によって評価を行う。

$$\text{正解指摘率} = \frac{\text{連接可と判定した語数}}{\text{正解語データ中の正解語数}} \times 100[\%]$$

$$\text{誤り検出率} = \frac{\text{連接不可と判定した語数}}{\text{誤り語データ中の誤り語数}} \times 100[\%]$$

正解指摘率は正解語データから、誤り検出率は誤り語データからそれぞれ算出するが、このとき、誤り検出の網羅性を重視する立場をとる（多少正解語を誤りであると検定しても誤り語を見逃さないという立場）。従って、意味連接辞書中で連接可（○）以外のものはすべて連接不可として扱う。

4．3　結果と考察

評価実験の結果を表2に示す。誤り検出率は非常に高い値を得ているのに対して正解指摘率は、6割弱とあまり高くない。これは両者の算出にあたって、誤り検出の網羅性を重視したため、前者は上限値を、後者は下限値を示しているからだと考えられる。従って、誤りを検出するという推敲支援の立場からは、本方式はサ変動詞同音異義語誤り検出に関して有効であると考えられる。

表2：　評価実験結果

正解指摘率［%］	誤り検出率［%］
58．2	99．8

5．おわりに

本稿ではサ変動詞同音異義語誤りを複合語同音異義語誤りとの類似性を用いて自動的に検出する方式について述べた。新聞記事を対象に行った評価実験の結果、正解指摘率＝58．2%、誤り検出率＝99．8%を得、複合語同音異義語誤りとサ変動詞同音異義語誤りとを1つの情報（意味連接辞書）を用いて検定する見通しを得た。

今後は、正解指摘率の向上を目指して、意味連接辞書の情報の充実、検定対象格の決定方法の充実を図る。

【参考文献】

［1］奥：「意味カテゴリを用いた複合語の同音異義語誤り検定方式」
第38回情処全国大会2J-7(1989)