

ソフトウェア故障記述文の形式化による意味解析法*

2Q-02

矢田佳子† 佐藤匡正‡

島根大学大学院総合理工学研究科§

1. 序論

ソフトウェア開発における障害は、障害票によって、その状況や原因が記録されている。この記録を分類し、検索することで過去の障害の経験が活用され、新たな設計へ反映される。しかし、障害の記述方法は多様である。例えば「データへの設定内容誤り」と「データに設定する内容が正しくない」は同一の事例を示す記述だが、使用される語句や表現方法が異なる。これでは、分類や検索に有効な鍵語を決定することが難しい。また、鍵語となる語句が予め限定され、選択した鍵語と記述内容が一致しないことがある。

そこで自然に記述された文から鍵語を取り出すことで幅をもたせる。鍵語は記述中の重要な部分である。しかし、記述のゆらぎから重要部分に一定の基準がない。ゆらぎを吸収するために文を形式化する。

本稿では、故障記述文の形式を表現する方法を案出し、特定の語句の抽出への有効性を確かめる。

2. 形式化の方法

1) 記述上の特性

故障記述文の記述項目を大別すると、①現象の状況、②原因、③措置である。①は、「ある状況下で生じた現象」の記述であり、②は、「現象の発生原因と結果」の記述である。これらの記述には、仕様を基準とする論理性があると予想される。つまり、何らかの前提条件とそれに対する結果を表している。前提と結果に着目した形式化を行う。また、記述文中で使用されている単語の種類は限定されている。そこで、その特徴を表すように形式化の

表現規則を考案する。

記述文は文成分として1つ以上の述語をもつ。述語を1つだけもつ文を単文とする。文成分の修飾関係は述語によって決まる。複数の単文が文を構成しているのに対して、省略されている文がある。これを句表現とする。

2) 表記方法

記述上の特性を表すような表現方法を考案する。形式化するための表現規則を表1に示す。また、文中の名詞は簡略化し、原則として動詞の活用は無視する。

表1 形式化規則

		表記法			表記法
1	主語(S)+述語(V)	S~V	6	否定	Λ
2	目的語(~を)(DO)	(DO)	7	及び・加えて	+
3	間接目的語(~に)(IO)	(IO)	8	前提条件	—
4	存在する(~がある)	~*	9	省略項目	?
5	可能である(~できる)	~#	10	~の	.

3) 述語に着目した修飾関係決定法

形式を記述する場合、各文成分の修飾関係を決定する方法である。述語から他の文成分との関係が決まることから述語に着目して修飾関係を決定する。

4) 句表現の単文変換法

句表現では、述語に着目して修飾関係を決定することができない。そこで、予め述語部分を補って単文に変換してから修飾関係を決定する。

3. 記述の方法

1) 分解

記述文を分解し、単文単位で修飾関係を考察する。述語を主体として修飾関係が視認できるように表現する。また、前処理として句表現は、単文に変換しておく。

例えば「 α が β を出力」という文を考える。これは句表現であるので、変換すると「 α が β を出力する」になる。そして、述語「出力する」に着目して修飾関係を決定する。「出力する」は文成分として主語、目的語をもつ。これを、表1に従って記号で表記すると「 $\beta | \alpha \sim$ 出力する」となる。

*Semantic Analysis Methods by Formalizing Software Failure Description

†Yata Yoshiko

‡Satou Tadamasu

§Interdisciplinary Graduate School of Science and Engineering, Shimane University

2) 前提条件の識別

記号化した単文を、修飾関係から前提と結果に分離する。前提は、他の単語を修飾する節や他の文を修飾する節であり、後者は「～において」、「～の時に」等の接続表現を用いて表される。結果は前提によって修飾される単語又は文である。

3) 組立て

前提とそれに対する結果を関連付ける。前提の上部にその結果を記述し、両者の修飾関係を表す。区切り記号として"|"を用いる。

4) 例示

次の2つの例文を形式化し、その結果を図1に示す。

[1]入力処理(P)において、αコード(C)に対して入力(O)を行おうとしたところ、αコード(C)が表示されなかった。

[2]エラーメッセージ(M)を表示して異常終了した。

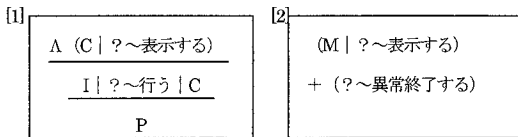


図1 形式化による表現

4. 分析

形式化された記述文から特定の語句を抽出し、分析する。

4.1 誤りを示す記述の分析

記述文を、誤りを表す記述部分によって3つに分類する。分類項目は、A.前提、B.結果、C.その他とする。分類した結果は、Bが9割以上、Cが1割未満である。これより、記述文において、誤りを示す記述は結果部分に含まれると言える。また、Cに分類される文は、故障記述文として不完全で、故障の状況を正確に記述していない。例えば、図1の[2]では、記述文中に前提が存在しない。形式化を行ったことで、こうした記述上の不備が発見できる。

4.2 結果部分の分析

結果部分に着目し、各文成分の関係について分析を行う。形式「O | S~V」、「S~V」(S:主語 V:述語 O:目的語)を考える。前者をSVOP型、後者をSVP型とし、各成分に対応する語句を抽出する。それぞれの型でのS、Oの出現率を図2に示す。

1) 全体の傾向

SVOP型ではSは省略されているが、8割の文がOをもつ。SVP型では7割の文がSをもつ。これより次のことが重要だと言える。

- ・SVP型：SとVの関係
- ・SVOP型：OとVの関係

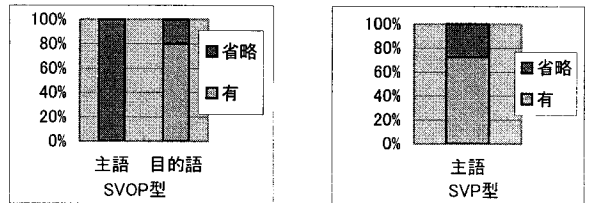


図2 主語(S)・目的語(O)の出現率

2) 誤りの性質

SVOP型ではOとVの関係、SVP型ではSとVの関係に着目し、それぞれの関係がどのような性質の誤りを表しているかで分類する。分類項目は次の3つである。

- A.本来実現すべき機能が実現できない、誤っている
- B.当初の機能の変更
- C.誤りの発生、存在

各型での性質の出現率を表2に示す。

表2 誤り性質の出現率

性質	出現率(%)	性質	出現率(%)
A	29%	A	80%
B	0%	B	20%
C	67%	C	0%
その他	4%		

SVP型

SVOP型

A: SVOP型の誤りの性質の8割を占める。OとVに対応する語句を抽出すると、両者の関係は本来実現すべき機能が実現できないことを示す。

C: SVP型でのみ出現し、6割を占める。SとVに対応する語句を抽出すると、両者の関係は誤りが発生したことや存在を示す。

5. 結論

提案した方法で形式化を行い、特定の形式に着目して語句を抽出し、分析を行った結果、SVOP型でOVの関係、SVP型でSVの関係が重要であることがわかった。形式化が特定の語句の抽出に有効であることを確かめた。