

# 検証計画書テキストにおける属人性に関わる部分の検出

日本アイ・ビー・エム株式会社

荻野 紫穂, 竹内 広宜, 伊川 洋平, 秋本 仁志, 中田 武男, 坂本 佳史

{shiho, hironori, yikawa, fallbook, nakada, sakay}@jp.ibm.com

## 1. 概要

製品開発工程において、テスト工程は最終段階に当たる。テスト工程で発見できなかった欠陥は製品の一部として市場に出回って製品や企業のイメージを大きく損ね、欠陥に対応するためのコストを嵩ませる。テストの仕様を定義する検証計画書やテスト仕様書は、市場に欠陥が流れるのを食い止める、最後のフィルタを定義していると言っても過言ではない。これらの文書中で問題になりやすい記述については、書くべき項目とその内容の示唆<sup>1)</sup>や、インスペクションの手法<sup>2)</sup>などによって、間接的に喚起されている。

開発のテストより前の工程、特に実装に携わった人員がテストを実施する場合には、テスト仕様書が曖昧に記述されていても、テスト実施者が背景知識を利用してうまく情報を補完できることも多い。

一方で、設計・実装が開発期間の多くを占める場合には、テスト工程において、それまで工程に参加していなかった新たな人員がテスト実施者として投入されることがある。こうした場合、テスト実施者は、それまでの工程で生じた経緯や用語を知らないため、テスト仕様書の記述の曖昧さを、背景知識で補完できない。テスト実施者が、曖昧な記述を自分で勝手に補完した場合は、正しいテストが行なわれない可能性がある。テスト実施者が、テスト仕様書の曖昧な記述を、「曖昧である」と正しく判断した場合には、その記述の問い合わせのために、テスト期間が増大する可能性がある。つまり、曖昧な記述のテスト仕様書に従ったテストが、テスト設計者・作成者の意図通りに行なわれるかどうかは、記述を解釈するテスト実施者の属人性に委ねられる。

上記を考えると、検証計画書やテスト仕様書は、それらの文書とそこで参照される文書群の内容を読むだけで、テストの設計者・作成者の意図がテスト実施者に誤解なく伝わるように、できる限り属人性に依存せずに記述されていることが望ましい。しかし、実際には、設計段階や実装の初期段階で仕様が決定していなかったり、仕様の変更が予想されたりすることから、テスト仕様書の作成は、テスト工程の開始のすぐ前に行なわれることも多い。仕様の変更は実装工程にも影響を及ぼすが、そうした場合には、実装がまず重視される<sup>3)</sup>。テストに関しては、その実施自体に重点が置かれ、テスト仕様書の記述には、あまり焦点が当たっていない。

我々の研究は、開発関連文書のテキストに自然言語処理技術を適用して、文書の品質やその影響を見積もり、開発工程における問題発生を抑えることを目標とする。本稿では、テスト仕様書の記述が曖昧であるかどうかを判断するために、テスト記述に特徴的な語と品詞に着目して、テスト記述の中で、解釈がテスト実施者の属人性に依

存している可能性が高い部分を調査した。また、開発工程を熟知している読み手とあまり知らない読み手とによるテスト記述の理解の差にどのような傾向があるかについても観察した。

## 2. テスト関連文書群の記述と対象範囲

### 2.1. テスト関連文書群とその分析

欠陥の原因が実装以前にある場合、後工程への影響を考えると、いわゆるテスト工程ではなく、開発工程のできる限り早期に発見できることが望ましい<sup>4)</sup>。0 章で述べた通り、実装などの前工程が増大した場合などには、実装やテストの実施自体が重視されるため、テスト仕様書を記述する手間は歓迎されない。こうした手間を削減し、同時に、上流工程と下流工程での情報の非一貫性を防ぐ目的で、上流工程で作成された文書から情報を抽出する試みも出ている<sup>5)6)</sup>。これらは有望な技術ではあるが、まだ特定の工程に特化した定義を必要とする段階にある。テスト関連文書は、現在のところ、人手による作成がほとんどであり、テスト関連文書群の作成やその記述に関する注意事項の多くは、人手による作成を前提に書かれている<sup>7)</sup>。こうした注意事項の多くは、IEEE Std 829-1998<sup>8)9)</sup>で定義された文書群と、そこに含まれる項目をテンプレートとして利用することを推奨している。

1)には、テスト関連の文書群と、そこに含まれるべき項目と、項目の中で記述すべき事項が定義されている。本稿では、テスト項目の仕様の記述について、その解釈が読み手の属人性に依存する部分を抽出するために、1)から対象となる文書群の記述項目を選んで、その記述に関する示唆に反する部分を、実際のテスト記述の中から単純な観点を使得って抽出することを考える。

### 2.2. 本稿で対象とする範囲

1)は、テスト関連文書群として以下の文書種を挙げている。

- Test Plan
- Test Design Specification
- Test Case Specification
- Test Procedure Specification
- Test Item Transmittal Report
- Test Log
- Test Incident Report
- Test Summary Report

Test Plan から Test Item Transmittal Report までがテスト実施以前に作成され、テストにおいて参照される。テスト

<sup>i</sup> IEEE Std 829-1998 の revision として 829-2008 が発表されているが、ここでは(6)などの引用に従って 829-1998 を挙げる。

実施からその後にかけて、Test Log から Test Summary Report などがまとめられる。このうち、テストの実施に関する具体的な値や手順の解説は、Test Case Specification と Test Procedure Specification に記述される。

実際の現場で作成されるテスト関連文書群は、上記の1)の文書種に完全に合致しているわけではない。各文書種の項目や役割を分解し、個々の開発プロジェクトの採用手法や経緯などによって組み合わせ、文書を作成することも多い。本稿では、実際のテスト仕様書から、Test Case Specification 中の項目である Output specifications と、Test Procedure Specification 中の項目 Procedure steps のサブ項目である Proceed とに該当する部分に着目して、読み手の属人性との関係を調査した。

## 3. テスト記述の処理

### 3.1. 従来手法

テスト関連文書群自体や、その記述に曖昧さがあることに付随する問題点は、様々な文献で指摘されている。問題の解決策の案として、7)を始めとする様々な文献が、IEEE Std 829-1998<sup>1)</sup>で記述される文書群と、そこに定義されている項目をテンプレートとして利用することを推奨しているのは、前章でも述べた通りである。

2)は、インスペクションで使われる手法を用いて、文書を吟味する過程を示している。これらには、一般にどのような内容を記述すべきかに関する示唆や、どのような記述に注意して文書を吟味すべきかに関する個々の例は示されているが、テスト記述に生じやすい曖昧さを半自動的に抽出する具体的な手法は示されない。この背景として、テスト関連文書群は書式も記述も非常に多岐に渡ることから、一般的で、かつ、具体的な手法を示すのが困難である、という面がまず挙げられる。同時に、安易に具体的な表現を多く例示しすぎると、そこに書かれた表現に特化して、そうした表現を使わないことだけに力を注ぐといった、表面的な記述だけの訂正を招きやすい面も大きいと思われる。

6)は、設計文書の書式をモデル化して書式の違いを吸収した上で、定義される属性名とその値とを抽出する。6)の関連研究の章で述べられている通り、自然言語処理技術を要素技術として組み込むことが可能だが、具体的にどのような自然言語の記述に着目するかは6)の主旨ではないため、触れられていない。5)は、ユースケース記述を処理してモデル化するための要素技術と、その精度を報告している。この技術は非常に有用だが、現在のところ、ユースケースの記述に特化してアクターやアクションなどの情報を抽出処理しているため、テスト記述に適用した場合の精度等は不明である。

1 章でも述べた通り、テスト仕様書の記述自体には、あまり焦点が当たっていない。本稿では、テスト仕様書の記述に特徴的な語と品詞を中心に観察し、テスト仕様の記述と、その解釈の属人性への依存との関連性の発見を図った。

### 3.2. 抽出の観点

1)では、Test Case Specification 中の項目である Output specifications と、Test Procedure Specification 中の項目 Procedure steps のサブ項目である Proceed とについて、それぞれ、

- Output Specifications

Provide the exact value (with tolerances where appropriate) for each required output or feature

- Proceed

Describe any actions necessary during execution of the procedure

という、項目内容の記述に関する示唆がある。本稿では、これらについて、“exact value が記述されているか”、“any actions が記述されているか”について、表現が曖昧な部分を調査するものとした。

調査において着目する点については、できるだけ単純で、様々なテスト仕様の記述で共用性が高いものを採用した。実際に採用した尺度は、以下の通りである。

- exact value が記述されているか

- 「確認する」という動詞の有無とそのヲ格

「確認する」という動詞を使って、テストで観察すべき対象を明確にしているかどうかと、その動詞に対して、「○○であるか(どうか)」「○○であること」などの exact value が与えられているかどうかを調査する。これは、「変数の値が 0 以上 5 未満であることを確認する」には exact value が記述されているが、「変数の値を確認する」では観察対象の値が exact ではなく、解釈が読み手の属人性に依存することによる。「期待結果」などの項目設定がある場合は、その項目の存在を「確認する」という動詞の存在と同様に扱い、項目範囲に「○○ていること」「○○であること」などの述語句があることを、ヲ格に値の範囲が示されている状態と同様に扱う。

- 「正しい」などの形容詞・形容動詞の有無

「正しい値」「正しく動いている」などの記述は、読み手によって解釈の違いを招きやすい。こうした表現は、8)で挙げた、IEEE 830-1998<sup>9)</sup>の検証可能性 (verifiability) の違反に当たる。

- any actions が記述されているか

- テスト項目記述の動詞数

any actions を記述するためには、actions のための動詞の記述が必要であると言う予想から、テスト項目の動詞数をカウントした。

これらの観察のための日本語の構文解析には、IBM Cognos® Content Analytics (CCA)<sup>10)</sup>の日本語処理部を使用した<sup>ii)</sup>。

<sup>ii)</sup> IBM, Cognos は、International Business Machines Corp. の商標です。

## 4. 実際の検証計画書とテスト仕様書での調査

### 4.1. データと被験者

データには、実際のハードウェアの検証計画書1文書と、ソフトウェアのテスト仕様書1文書から、テスト項目に関する記述を人手で抜き出したものを使った。データの概要を表1に示す。

被験者には、ハードウェア・ソフトウェアそれぞれに関して、ドメイン知識のある者となない者を選び、ドメイン知識がある者については、データである検証計画書・テスト仕様書の対象プロジェクトのテストや技術提供の現場で関わっていたか(TEC)、プロジェクトのそれ以外の場(マネジメント等)で関わっていたか(PRJ)、プロジェクトにはかかわっていないがドメイン知識が豊富か(DOM)に分類した。被験者の内訳を表2に示す。ソフトウェアのテスト仕様書については、データのプロジェクトに属する被験者がいなかったため、ドメイン知識のあるなしだけに関して分類した。

被験者には、「テスト記述と被験者の背景知識などの属人性との関係を調べる」という目的を伝え、

- テスト記述中に知らない単語がある場合、「名詞・動詞の意味くらいは解るが、それ以上の詳細知識はない」という仮定でテスト記述を読む
- 「自分には解るが他の人が読んだら解らない」という観点ではなく、「記述に曖昧な部分もあるが、自分の背景知識を使って情報を補って理解できる」という観点でテスト記述を読む

という前提条件を伝えて、各テスト項目の記述に対して、以下の4段階のスコア付けをお願いした。

1. 全く情報が足りなくてテストできない。問い合わせるのも手間がかかりすぎると思われる。
2. 記述を読んで、テストの不明点を5点未満問い合わせないとできないが、多分5点未満問い合わせればできる。
3. 記述を読んでテストできると思うが、間違える可能性も捨てきれない。
4. 記述を読んで、テスト項目の記述者の思ったテストができる。

これらのスコアと、読み手の属人性との関連について、次節で述べる。

## 4.2. 属人性と尺度との関係

### 4.2.1. 確認系動詞との関係

ソフトウェアのテスト仕様書は、9つ全ての項目が「～を確認する」に関する基準を満たしていたため、ここではハードウェアの検証計画書についてだけ、観察する。表3に、「～を確認する」とそれに相当する表現の有無と、「確認する」のヲ格に「○○が××であることを」などの具体的な表現が入っているかどうかごとに、被験者のスコアの平均値を示す。

全体的に、確認系動詞がある・そのヲ格が具体的に定義される、という詳細化のステップに従って、スコアの平均値は微増する傾向にある。その傾向は、ドメイン知識のない被験者グループ(GEN)で最も高く、対象プロジェクト

の一般知識を持つグループ(PRJ)がそれに次ぐ。また、どの被験者グループも、「～を確認」に相当する表現を持つ項目に付与したスコアの平均は、各グループの項目全体に対する平均スコアを超えている。この結果から、「何がどんな値(状態)であることを確認するか」と記述することは、読み手の理解に影響を与えると推測できる。ただし、その影響は、プロジェクト開発に携わった人員に対してよりも、背景知識の少ない人員に対して大きい。

### 4.2.2. 検証可能性違反との関係

検証可能性違反の有無と平均スコアとの関係を表4に示す。検証可能性違反の有無は、プロジェクトの現場で関わっていたグループ(TEC)のスコアには影響していない。これに対し、PRJとGENとでは、検証可能性違反のない項目の平均スコアは、全体の平均スコアよりもやや高い。

「正しく書き込まれているか」「正しく発生するか」などの記述は、ドメイン知識や、開発経緯を共有している場合、あまり問題にならない。むしろ「正しく」と一言書くだけで余分な記述を避けられるメリットを持つ。

一方で、そうした知識や経緯が共有されない場合、「正しく」などの検証可能性違反の表現は、理解を妨げる要因となることがある。今までプロジェクトの開発者がテストしていた部分を外部に任せる場合や、開発依頼者が開発経緯を詳細には共有しない場合も、どういったテストが行なわれるかに関する考えの相互疎通を妨げる可能性がある。

### 4.2.3. 動詞の数との関係

動詞の数と平均スコアとの関係を表5に示す。ハードウェアの検証計画書において、動詞が1つもないテスト仕様書の記述項目(例:「○○○テスト」のように、テストの名前だけが挙げられているものなど)は、全てのグループにおいて、動詞が1つ以上あるテスト仕様の記述項目よりも、平均スコアが低い。動詞の数が増えるに従って、平均スコアが上がる傾向があるが、特にGENにおいて、その傾向が強く、他のグループでは微増傾向に留まる。テスト記述に動詞が多いソフトウェアのテスト仕様書も、ハードウェアの検証計画書と同様の傾向が見られる。動作記述がある程度以上整っていることは、読み手の背景知識に関わらず必要と思われる。特に、GENグループの被験者に近い、知識の浅いテスターが読むテスト仕様書には、テストステップのアクション記述が丁寧であることが重要である。

一方で、GENグループにおいて、動詞数の増加に対する平均スコアの増加が他のグループに比べて大きいことは、「専門知識のあるテスターが『動詞が増えても』不明点は依然として存在する』と解釈するテスト仕様を、知識の浅いテスターは、詳細なアクション記述があるというだけで『テストを実施できる』と思い込みやすい」という側面にも繋がるため、注意が必要である。

### 4.2.4. 全体の傾向

背景知識の有無という面では、TECとPRJが「プロジ



エクトに関する知識とドメイン知識と一般知識」、DOM が「ドメイン知識と一般知識」、GEN が「一般知識のみ」を持っていると考えられる。ここから、TEC と PRJ がやや似た値を示し、GEN がその反対に位置して、DOM が TEC-PRJ と GEN との中間にある、とも想像ができる。しかし、今回の実験結果のスコアの平均値を見ると、TEC がかなり強く独立した傾向を示し、PRJ と DOM が似た傾向を示す、という結果が得られた。

今回は、被験者数も対象プロジェクト数も少ないため、これ以上の確認はできないが、他のデータについても同じような傾向が得られるとしたら、テストの仕様記述の解釈の属人的な背景知識は、プロジェクトそのものの知識よりも、現場で培われて累積されていく知識と、それ以外の知識の差が一番大きい、ということになる。本稿の最初に述べた通り、テストの局面で投入される人員は、DOM または PRJ に近い属人性を持つと考えられるため、こうした新たな投入人員のためのテストの仕様記述は、それまでずっとプロジェクトに携わっていた人員が解釈できる仕様記述よりも具体性が必要とされる可能性が高い。また、開発依頼者は、プロジェクトの知識はあっても開発経緯を共有していない、PRJ に近い状態であることも多い。テスト仕様の記述者と開発依頼者との間で文書を交換する場合、TEC と PRJ のような理解の差が存在することを念頭に置いて記述するほうが、テストに仕様に関する誤解を防げる可能性がある。こうした点を明らかにするには、追加調査が必要となる。

## 5. まとめ

本稿では、テスト仕様の記述に着目して、単純な尺度を用いて、記述の解釈が読み手であるテスト実施者の属人性に依存する傾向を捉える調査の結果を報告した。

本稿では、テスト記述は人手で抽出したが、6)などと組み合わせることによって、文書の書式の差を吸収することも考えられる。また、5)のような技術や 8)と合わせて、開発プロセス上の文書成果物の処理を統合的に行なう方向も考えられる。

本稿の調査の被験者は偏りがあったため、ドメイン依存の部分と属人性が切り分けられない部分があった。こうしたことも考慮しながら、開発関連文書に対する自然言語処理技術の適用によって、開発プロセスをサポートする手法を研究していく予定である。

## 参考文献

- 1) IEEE Standard for Software Test Documentation
- 2) 鈴木 三紀夫: 三色ボールペンで読む仕様書, ソフトウェアテスト入門, pp. 138-150, 技術評論社 (2009)
- 3) 大西 健児: テストはなぜ必要なのか, ソフトウェア・テスト PRESS, pp. 2-5, 技術評論社 (2005)
- 4) ローレンス・H・パトナム, ウェア・マイヤーズ: テストでの欠陥抽出, 始めて学ぶソフトウェアメトリクス (山浦 恒央訳), pp. 195-196, 日経 BP 社 (2005)
- 5) A linguistic analysis engine for natural language use case description and its application to dependability analysis in industrial use cases, 39th Annual IEEE/IFIP International Conference on Dependable Systems &

Networks (DSN 2009), pp. 327-336 (2009)

6) 岩間 太, 中村大賀: 設計文書に対する静的形式検査のための型システム, 日本ソフトウェア科学会ソフトウェア工学の基礎研究会 2009 (FOSE2009), レクチャーノート ソフトウェア工学の基礎XVI, pp251-258 近代科学社 (2009)

7) 大西 健児: テスト仕様書をどう書くか テストケースに何を挙げるか, ソフトウェア・テスト PRESS, vol. 6, pp. 2-66, 技術評論社 (2008)

8) 荻野 紫穂, 竹内 広宜, 中田 武男: 開発プロセス上の文書成果物の分析における自然言語処理の適用, 第 15 回言語処理学会 (2009)

9) IEEE Recommended Practice for Software Requirements Specification

10) <http://www-06.ibm.com/software/jp/data/search/textmining.htm>

	HW	SW
項目数	87	9
平均スコア	2.7	3.5

表 1 データ概要

被験者の背景知識など	HW	SW
対象プロジェクトテスト・技術 (TEC)	4	0
対象プロジェクト一般 (PRJ)	2	0
ドメイン知識 (DOM)	1	3
一般知識のみ (GEN)	3	7
合計	10	10

表 2 被験者の概要

確認系動詞	HW				
	例数	TEC	PRJ	DOM	GEN
なし	18	3.2	1.9	2.1	1.6
確認系あり	23	3.3	2.3	2.1	2.3
～ことを確認	45	3.4	2.2	2.3	2.7

表 3 確認系動詞に関する平均スコア

検証可能性違反	HW				
	例数	TEC	PRJ	DOM	GEN
あり	15	3.4	2.1	2.7	2.2
なし	71	3.4	2.2	2.1	2.4

表 4 検証可能性違反に関する平均スコア

動詞数	HW					SW		
	TEC	PRJ	DOM	GEN	例数	DOM	GEN	例数
0	3.2	1.9	2.1	1.7	16	0	0	0
1-3	3.4	2.2	2.3	2.4	39	0	0	0
4-6	3.4	2.1	2.5	2.4	15	4.0	3.1	3
7-9	3.3	2.0	2.0	3.7	12	3.8	3.3	3
10-	3.4	2.4	1.6	2.8	5	4.0	3.4	3
計	3.4	2.1	2.2	2.4	87	4.0	3.2	9

表 5 動詞数に関する平均スコア