

役者評判記計量テキスト分析のためのノート

—— 現状と課題 ——

戸塚 史織（立命館大学大学院文学研究科）

E-mail it0875sf@ed.ritsumeit.ac.jp

要旨

歌舞伎研究に不可欠な資料である評判記の新たな研究の可能性として、計量テキスト分析を考えている。この分析では形態素解析の過程が一つの難関になることが予想されるが、評判記を対象とした既存形態素解析辞書の有効性や問題の発生箇所については不明瞭であり、実践に踏み切ることが難しいのが現状である。本稿では、既存形態素解析辞書による評判記解析結果の比較分析を通し、評判記形態素解析の現状と課題について検討する。

abstract

Yakusha hyobanki (Commentary on Actors) is one of the indispensable materials for *Kabuki* studies. For its further research use, this paper examines the possibility of its quantitative text analyses. As its preparatory stage, the process of morphological analyses can present difficulty. Moreover, we do not know so much about how effective the existing morphological analysis dictionaries can be for analyzing *Hyobanki* and which parts of its text present problems for the dictionaries. Therefore, in order to put them into practice, this paper will compare and analyze the results of analyses using the dictionaries, and discuss the current status and future issues of morphological analyses of *Hyobanki*.

1. はじめに 評判記と評判記研究

歌舞伎の上演資料の一つに、評判記がある。現代では演劇雑誌に当たるもので、各役者の経歴・技芸・人気等を示すとともに、台本の伝存しない狂言の内容を知る有効な資料となる。現存最古の評判記は万治3（1660）年刊の『野郎虫』であり、元禄年間にほぼ定型¹⁾が完成すると、それは幕末²⁾まで続いた。

このような評判記は、かつては限られた研究者の間でのみ珍重される資料でしかなかったが、昭和中期以降資料として積極的に活用されるようになった。今日では歌舞伎研究に不可欠な資料として認識されており、例えば評文のみならず挿絵にも注目

した研究が行われるなど、多様な展開を見せている。このような研究の発展には、評判記を翻刻した『歌舞伎評判記集成』³⁾の貢献が大きい。翻刻され、研究利用が容易になったからこそ、新たな視点での研究が可能になったものと推測される。現在『歌舞伎評判記集成』は第三期の刊行に入っており、今後更なる活用の展開が望まれる。筆者は新たな研究の可能性の一つとして、計量テキスト分析を考えている。

2. 計量テキスト分析の可能性

2.1. 計量テキスト分析と形態素解析

計量テキスト分析⁴⁾は大量のデータから偏りのない客観的な印象を導き出せることが利点の一つであり、信頼性・客観性の向上とデータ探索に有用である。またこの分析によって得られた結果を元に、質的側面に関する新たな発見が得られる場合もある。本手法は人文・社会系の研究・実務の現場において幅広く活用され始めており、文学研究の世界では著者推定や真贋判定、成立年の推定などの手段として利用した例もある⁵⁾。

計量テキスト分析は【図1】のように行われる。

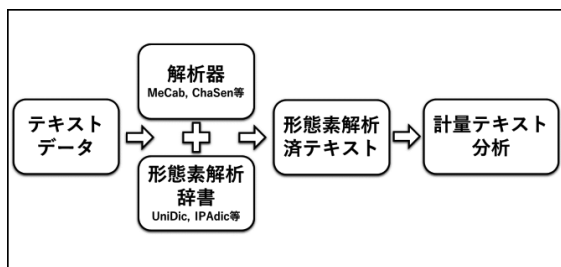


図1: 計量テキスト分析過程図

日本語文書の場合特徴的なのが形態素解析⁶⁾の過程である。テキスト内の単語計量が基本となる計量テキスト分析では単語の分割が必須であり、かつその分割単位は(1)単位の斉一性⁷⁾と(2)見出しの同一性⁸⁾を確保している必要がある。日本語はこのような分割を行うことが難しい⁹⁾ため、形態素解析の過程が一つの難関となる。

近年は形態素解析器としてMeCab¹⁰⁾やChaSen¹¹⁾などのフリーソフトウェアが普及し、形態素解析辞書もIPAdic¹²⁾や複数のUniDic¹³⁾などが作成され、現代日本語の形態素解析の問題はひとまず解決されたように見受けられる。しかしながら、古文を対象にした形態素解析には未だ多くの課題が残されたままである。日本語の文章は時代によって変遷しているため、分析対象に適した形態素解析辞書を用いなければ正確な解析は不可能である。しかしながら古文の単語認定は現代文より難しい。例えば市村¹⁴⁾は「第一に仮名遣いや送り仮名等の多様性の問題がある。(中略)これに加えて、片仮名平仮名の

混在、踊り字の使用、濁点無表記箇所が存在などが、表記パターンを増大させることとなる」と近世口語資料の表記と語形の多様性から生じる問題を指摘している。

2.2. 評判記類似資料の形態素解析実践

評判記も形態素解析が難しい古文であるため、計量テキスト分析による研究ではこの過程が障壁となることが予想される。評判記に類似する資料は、過去どれほど形態素解析に取り組まれているのだろうか。

まず時代に注目し、近世期資料を対象にした形態素解析の先行事例を調査すると、国立国語研究所の古文用UniDicの中に「近世口語（洒落本）UniDic」及び「中世口語（狂言）UniDic」がある。前者は近世後期の口語実態を探る資料として18世紀後半から19世紀の戯作の一つである洒落本を、後者は中世から近世にかけての言語資料として寛永19（1642）年の大蔵流の祖本『虎明本』を対象に電子化と形態論情報の付与を行い、それを学習用コーパスにして作られた形態素解析辞書である。解析精度には未だ課題が残るとされるが、近世期の形態素解析辞書はこれら以外に存在しない¹⁵⁾。同研究所は『日本語歴史コーパス江戸時代篇』¹⁶⁾も構築している。これは洒落本、人情本、近松浄瑠璃のテキストデータに形態論情報を付与したもので、この情報はコーパス検索アプリケーション「中納言」¹⁷⁾を通し簡単に確認することができる。またインターネット公開はしていないものの『新編西鶴全集データベース』¹⁸⁾もある。上阪¹⁹⁾は、本データベースを「西鶴作と考えられる120作品の文章を電子化及び形態素解析したデータベースで、本文と自立語は『新編西鶴全集』巻1～5の本文編・自立語索引編として出版されている」と説明している。

次にジャンルに着目し、歌舞伎関係の実践事例を調査した。まず既に紹介した『日本語歴史コーパス江戸時代篇』の内「Ⅲ. 近松浄瑠璃」には、後に歌舞伎に移入された作品も多数含まれている。『新編西鶴全集データベース』にも天和3（1683）年刊とされる評判記「難波の良は伊勢の白粉」が含まれ、形態論情報が付与されているようである。その

他、元禄6（1693）年上演の歌舞伎「好色伝授」の絵入狂言本の索引『好色伝授：本文・総索引・研究』²⁰⁾がある。本索引は自立語・付属語共に品詞を検討しており、形態論情報の付与が行われたと推定される。

評判記に関わる事例に限定すると『歌舞伎評判記集成第一期芸能索引』²¹⁾は趣向や特殊な演技演出などに関わる用語を抽出したもので、単語への分割という意味では形態素解析に繋がる要素を持つ。また永井²²⁾らは、中古和文 UniDic の形態素解析結果を基に、評判記テキストデータから人名や別名などの人物表現を抽出する手法を提案している。しかしながらこれらの事例は形態素解析そのものを行った事例ではない。評判記を対象とした実践は直接的なものを確認出来なかった。

以上の調査により、近世期資料及び歌舞伎資料を対象とした形態素解析の実践は少なく、評判記の形態素解析に特化した形態素解析辞書や手法は存在しないことが確認出来た。

2.3. 調査の目的

計量テキスト分析の手法は信頼性・客観性の向上とデータ探索の有用性に加え、新たな研究の糸口を得られる可能性があり、期待ができる研究手法である。しかしながら評判記を対象にする場合、形態素解析の過程が一つの大きな課題となる。先行事例の調査から評判記に特化した形態素解析辞書や手法は存在しないことがわかっており、個人で評判記に特化した形態素解析辞書を作成することも時間的に非常に困難である。

そこで考えられるのが、既存の形態素解析辞書を評判記の解析に用いることである。しかしそれらの辞書による評判記の解析について調査した事例はなく、専用の辞書でないためにその有効性や問題の発生箇所については不明瞭である。それ故、実際の利用に踏み切ることが難しいのが現状である。

そこで本稿では既存の形態素解析辞書による評判記の解析結果を比較分析することで、解析に生じる問題を明らかにすると共に、現存する辞書の中で有効なものを探る。これにより評判記形態素解析の現状と課題を明確にし、計量テキスト分析研究の

一助としたい。

3. 形態素解析辞書の検証

3.1. 対象と方法

評判記の計量テキスト分析に先立ち、その形態素解析に利用できる辞書の現状を把握し、最も有効な辞書を見極めるため、形態素解析辞書の検証を行う。古文を対象とした形態素解析辞書は先に紹介した「近世口語（洒落本）UniDic」・「中世口語（狂言）UniDic」を含む、国立国語研究所が作成した7種のUniDic²³⁾があるのみである。

今回はこの内①「近世口語（洒落本）UniDic」、②「中世口語（狂言）UniDic」、③「中古和文 UniDic」の3種を検証対象とする。①②は評判記が属する近世資料を対象とした形態素解析辞書として作成されていることがその理由である。しかしこれらは解析精度に課題が残ることが指摘されている。そこで古文用 UniDic の中では高い精度を誇り、永井らの研究でも利用された③も検証対象とした。

形態素解析の対象は評判記の定型を作ったとされる元禄12（1699）年刊『役者口三味線（京）』とし、形態素解析支援アプリケーション Web 茶まめでの解析結果を比較する手法を採る²⁴⁾。今回の検証は評判記の形態素解析により有効なものを探ることが目的であるため、解析結果の語彙素もしくは語彙素読み不一致箇所に着目し、比較する辞書のどちらで正解数が多いかを確認する方法を採用した。また具体的な解析課題を明らかにするため誤解析の分析も行った。誤解析例は、先行研究²⁵⁾にならい A. 境界認定、B. 品詞認定、C. 語彙素認定、D. 発音形認定の四段階に分け表に記したが、近世期の日本語の文法的な問題の多さから品詞認定については特徴的な例を表記するに留めている。

3.2. 近世口語（洒落本）UniDic と中古和文 UniDic

まず①「近世口語（洒落本）UniDic」と③「中古

和文 UniDic」を比較する。Web 茶まめでの解析結果 2500 語まで²⁶⁾を対象としたところ、語彙素もしくは語彙素読み不一致箇所は 620 件だった。内訳は、①のみ正解したものが 476 件、③のみ正解したものが 38 件、①③とも誤解析したものが 91 件、要検討としたものが 15 件である。今回の解析結果では①の方が、圧倒的に正確数が多い。

解析の誤りを詳しく確認するため、誤解析例を収集した【表 1】を作成した。

まず A. 境界認定段階でも①の方が正しく解析することが多い。全体的に誤解析しやすいのは市村が主張した近世資料の表記と語形の問題に関わる語（表 1: A-11~13, 20~23）と歌舞伎の用語（表 1: A-3, 14~18, 24~27）である。また A 段階の誤解析は辞書を問わず殆どが過分割の例であり、過結合の誤りは無きに等しい。ここで要検討としたのは「第一」を「第一」と 1 語にするか「第 | 一」と 2 語にするかという問題である。ここに揺れが生じると単位の斉一性を確保することが出来なくなる。

B. 品詞認定段階では、助詞・助動詞の識別が特徴的であった。特に③は A 段階で過分割した時それを助詞・助動詞と解析することが多い（表 1: A-14 「き」、A-15 「や | は」）一方、B 段階では「也」（表 1: B-8）などを助詞・助動詞と認識できないことが少なくない。また①③とも濁音の欠落（表 1: B-19）など表記の問題で解析を誤る例があった。ただし品詞については、例えば完了「た」（表 1: B-28）は現代文法では「た」、中古の文法では「たり」が正しいため、近世資料でこのような例をどう判別し、見出しの同一性を確保するかという課題が生じる。

C. 語彙素認定段階でも①の方が多くの場合正しく解析している。誤解析しやすいのは A 段階と同じく表記・語形に関わる語、歌舞伎の用語に加え、同音異字や同形異音の語である。同音異字は「度」と「旅」（表 1: C-4）のように同じ発音だが別の字のもので、同形異音は「ケンブツ」と「ミモノ」（表 1: C-12）のように同じ「見物」という表記だが別の読みを持つものを指す。日本語には同音異字、同形異音の語が多く、これらは文脈から正しい表記、読みを解釈しなければならない場合が多い。例えば「情」を「ジョウ」「ナサケ」（表 1: C-28）どちらと

するかという問題については慎重な内容解釈が必要になるため、今回は要検討とした。機械的な解析でこれらの判別を可能にするためには、専門家らの判断を仰ぎつつ学習させていく必要があるだろう。

最後に D. 発音形認定段階の誤解析例には、文脈に応じ「ども」と読むべき「共」を「とも」と読んでいる例（表 1: D-8）がある。ただし D 段階は A~C 段階の認定が全て正しいことが前提であるため数としては少なくなる。①と③を比較すると、①の方が D 段階でも正しく解析しやすいようである。

以上の検証から、解析精度が高いとされる③よりも①の方が圧倒的に正解数は多く、誤解析例の検討でも多くの段階で比較的正しく解析していることがわかった。従って①と③では①「近世口語（洒落本）UniDic」が評判記の形態素解析により有効であると考えられる。

3.3. 近世口語（洒落本）UniDic と中世口語（狂言）UniDic

続いて①「近世口語（洒落本）UniDic」と②「中世口語（狂言）UniDic」を比較する。正解数が拮抗したため、①と③の比較より検証対象語彙数を増やし、Web 茶まめでの解析結果 5000 語²⁷⁾までを対象にした。

語彙素もしくは語彙素読み不一致箇所は 368 件だった。内訳は、①のみ正解したものが 111 件、②のみ正解したものが 93 件、①②とも誤解析したものが 114 件、要検討としたものが 50 件である。①②の比較という点では、大きな差ではないものの①の方が正解数を上回っている。ただし①②とも誤解析していた件数は①の正解数と拮抗しており、これを無視することは出来ない。

誤解析の例を詳しく確認するため【表 2】を作成した。

まず A. 境界認定段階では①②とも同程度の誤解析が確認出来た。やはり表記・語形に関する語と歌舞伎用語の解析を誤ることが多い。強いて言えば①の方が評判記に良く登場する語（表 2: A-9~12）の解析に強く、②は近世の表記（表 2: A-4~5）に強いようである。また①②とも誤解析だっ

た語の多くはこのA段階で解析を誤っている。殆どが過分割による誤りで、解析しにくい表記・語形に関わる語や歌舞伎用語の誤解析の影響を受けた誤りもある。例えば「山下流」(表2:A-33)は役者人名と「〇〇流」という歌舞伎用語の解析に失敗した結果、上接の「其上」も正しく解析出来なかったと思われる。加えて「やつし」という歌舞伎用語に注目すると、この語は①②ともに正解例と誤解例の両方があり、見出しの同一性が確保されていないことを示している。

B. 品詞認定段階では、各辞書の解析の癖による誤解析が確認出来た。例えば「にて」(表2:B-2, 9)について、その正誤に関わらず①では断定「なり」+ 接続助詞「て」と解析しやすく、②では格助詞「にて」と判別しやすい。それぞれが基にした学習用コーパスに多く登場した文法が優先されているためだろうか。また助詞・助動詞については「な」(表2:B-38)のように現代文法と中古の文法で判断が分かれ、近世資料で何を正解とするべきか判断が難しい例がここでも確認出来、要検討の多くを占めた。

C. 語彙素認定段階でも①②の正解数の間に大きな差はないが、①の方が歌舞伎用語(表2:C-14~17)の解析に多く成功しており、「あんたる」(表2:C-13)の撥音便を正しく解析するなど表記・語形の問題にも僅かながら対応できているようである。

以上の検証から、大きな差ではないものの②より①の方が正解数は多く、誤解析例から歌舞伎に関わる語の解析に成功しやすいことがわかった。この点から①と②では①「近世口語(洒落本) UniDic」の方が有用性は高いと言えそうである。

3.4. 考察

以上の検証により①「近世口語(洒落本) UniDic」、②「中世口語(狂言) UniDic」、③「中古和文 UniDic」の3種のUniDicでは、①を活用するのが評判記の形態素解析に最も有効であることが判明した。

ただし現在の①の解析では計量テキスト分析に用いるに十分な精度を確保出来ていないといえない。解析には未だ多くの課題が残っており、特に表

記・語形の問題に関わる語や歌舞伎用語での誤解析が多いことが判明した。また①③比較時と①②比較時の①の解析結果は一致しない箇所が確認出来、同じ文章でも解析毎に違う解析結果が出ることもあるようで問題である。解析結果が一致しない箇所の多くは、誤解析しやすい語であった。これらの解析における問題は、専用辞書でないために見出し語登録が不十分なこと、またその見出し語に基づいた学習用コーパスによる学習が圧倒的に不足していることから生じていると考えられる。

最も有効な解決手法は、誤解析が多かった語を見出し語として登録しUniDicの更新をすることだが、これを実行することは簡単ではない。個人で行うことが出来る解決手法としては、十分な解析が出来なかった語についてMeCabのユーザー辞書²⁸⁾を作成することやKH Coderの強制抽出機能²⁹⁾を利用することなどが考えられる。しかしながらこれらの機能を用いて語彙指定を行う負担は決して小さいものではなく、かつこれらの機能では計量テキスト分析で重要となる単位の斉一性と見出しの同一性が確保された語彙認定は難しい。

4. 評判記形態素解析辞書の作成

今回の検証結果を踏まえると、評判記を基にした形態素解析辞書が作成されることが望ましい。そこで、評判記を基にした形態素解析辞書作成の可能性と、作成上予想される課題について考察しておきたい。

4.1. 評判記形態素解析辞書の可能性

評判記を基に作成した形態素解析辞書は、評判記を対象とする分析に留まらず、歌舞伎研究における分析に幅広く有用である可能性が高い。

評判記は万治3年(1660)刊の『野郎虫』から江戸末期までほぼ抜けることなく連々と現存しており、かつその刊行は元禄年間に完成した定型を保ち継続的に行われた。台本などの他資料は現存資料の年代幅は狭く、作者の偏りや純粋な近世語を用

いない演目の存在などのため均質性もそれほど高くない。従って、どの資料より広範な年代を包含し均質性を保持する評判記を基にすることで、どの時代にも対応し得る安定した形態素解析辞書が作成できる見込みが高い。

加えて、評判記には歌舞伎の基本的な用語が網羅され、他資料の特徴も有している。研究に用いられる基本的な語は歌舞伎の舞台を包括的に記述の対象としているからこそ記されるものが多く、台本などには登場しない用語も多い。さらに狂言内容についての記述は、重要な台詞や演出の様子を直接抜粋するような形で掲載することもあり、台本などに共通する部分が確かに内包されていると言える。この点で、評判記を基にすれば歌舞伎関連のどの資料でも有効な形態素解析辞書を作成できる可能性が高いと言えよう。

さらに、歌舞伎研究に留まらず近世資料の形態素解析辞書という視点から見ても、評判記を基に形態素解析辞書を作成することは有効であると考えられる。歌舞伎という江戸期の大衆演劇を記述の対象とした評判記には、その特性から江戸時代を通した絵画や風俗、近世語、江戸期文化などあらゆる分野の情報を遍く保持しており、近世研究において非常に有用な資料であることは明らかである。特に近世語の研究においては広範な年代を包含し均質性を保持していることに加え、京・江戸・大坂それぞれの版があり多様な人物による合評形式を採っている点から、地域性と身分の違いを反映した近世口語の実態を探ることが出来る可能性があり、資料としての重要性は極めて高いと言えるのではなかろうか。

4.2. 評判記形態素解析辞書作成の課題

UniDic の作成³⁰⁾を例に挙げると、その作成には【図2】のように解析に用いる見出し語一覧となる(1)辞書データと、(1)との整合性を保つ形態論情報が付与されたテキストデータである(2)学習用コーパスが必要である。

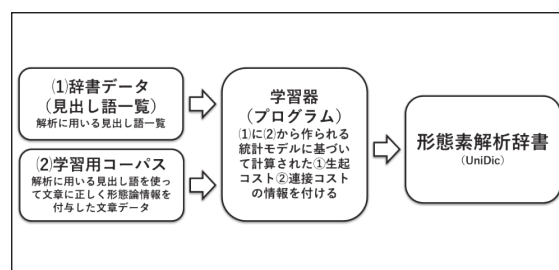


図2: 形態素解析辞書作成過程図

これを踏まえ、評判記形態素解析辞書の作成にあたり各過程で想定される課題について考えておきたい。

まず(1)について、歌舞伎には専門的な用語が多く見出し語一覧作成の負担が大きいという課題がある。この解決策として『歌舞伎評判記集成第一期芸能索引』³¹⁾を活用することを提案したい。本書は趣向や特殊な演技演出などに関わる用語が専門家の熟慮の末に抽出されており、見出しとして採用するに適切な資料である。また先に紹介した形態素解析実践例からも、語彙認定における同一性の確保のための様々な情報を得られるのではないかと思う。

次に(2)について考えられる課題として、テキストデータの用意とタグ仕様の設定が考えられる。まずテキストデータの用意についてだが、小木曽ら³²⁾は学習コーパス量について5千語程度で90%の精度、5万語程度で95%の精度に到達し、10万語程度で精度の向上は鈍化することを指摘している。今回分析対象とした評判記『役者口三味線(京)』はWeb茶まめの解析結果によれば約1万9千の語彙数であり、『役者口三味線』三都分の評判記を学習用コーパスとするだけでもかなりの語彙数が確保出来る。またタグ仕様については、市村³³⁾や小林³⁴⁾が洒落本、『虎明本狂言集』で設定したタグセットが大方活かせるように見受けられる。

5. おわりに

評判記の計量テキスト分析では、形態素解析の過程が第一の難関になると考えられる。本稿ではこの過程に着目し、既存の形態素解析辞書による解

析結果の比較分析を行うことで、評判記の解析に生じる問題を明らかにすると共に、最も有効な辞書を検討した。

結果として、評判記の形態素解析には「近世口語（洒落本）UniDic」が最も有効であると考えられるものの、その解析精度は不十分であることが確認出来た。単語分割の段階から誤解析が多いため、基礎的な確認のための抽出語検索をはじめ、コロケーション統計や共起ネットワークのような計量テキスト分析を行うことは難しそうである。また特に誤解析が多いのは表記・語形の問題に関わる語や歌舞伎用語であり、この問題の要因は、専用辞書でないために辞書作成段階での見出し語登録とその見出し語に基づいた学習量が不足していることが大きいと考えられる。

これを受け、今後の課題として二つの方策を検討している。第一は長期的な方策であり、第4章でも述べたように評判記を基にした形態素解析辞書を作成することである。これは歌舞伎研究のみならず、近世研究に幅広く有効な辞書となる可能性があり、多様な研究展開が見込まれる。しかしながら形態素解析辞書の作成は一朝一夕に行えるものではない。

そこで第二に短期的な方策として、形態素解析辞書に多くを頼らない、計量テキスト分析の手法について検討することを考えている。今回の形態素解析辞書の検証は、同時代の資料を基にした形態素解析辞書でもジャンルが異なれば十分な解析が難しいことを示している。形態素解析辞書の作成はかなりの時間を要する困難な作業であり、評判記に限らず多様なジャンルで辞書の開発が望まれる状態であることを考えると、それを待つのは現実的ではない。そうであるならば、専門の形態素解析辞書がなくても可能な分析手法について検討しておくことが必要である。現在考えられる具体的な例として、計量テキスト分析用ソフトウェア KH Coder に、文字列そのものから語を抽出できる機能³⁵⁾がある。この方法の場合、分析者の理論仮説や問題意識によって抽出される語が限定されてしまうという難点はあるものの、調査対象が明確であれば、むしろ分析者の理論仮説に基づきながらテキストデータの多様な側面に自由に焦点を絞ることができ利点にもな

る。このような機能についてもその現状と課題を明らかにすることは、評判記、延いては多様なジャンルにおける計量テキスト分析研究を推し進めることに繋がるだろう。

〔謝辞〕

本研究資料として、赤間研究室のテキストアーカイブを活用しました。研究にあたり御指導、御協力頂いた皆様に感謝の意を表します。

〔注釈〕

* URL は全て 2021 年 2 月 7 日最終確認

- 1) 定型は黒表紙小型横本で京・江戸・大坂の役者評を各一巻とする三都三巻三冊の編成になっている。本文では立役・敵役・若女方などの役柄別に、技芸の巧拙や人気の高下による序列に従って配列された役者の芸評を述べており、上上吉・上上・上・中ノ上上・中ノ上・中という位付を採用している。評判は合評形式で賞賛・非難をとりまぜ、読者の幅広い思いを先取りすると同時に公平を期す形式になっている。(赤間亮『図説江戸の演劇書』早稲田大学坪内博士記念演劇博物館、2003 年・松崎仁「役者評判記」『歌舞伎事典』平凡社、1983 年)
- 2) 慶応 2 年まで刊行が続いた。明治に入ると形態を変え「俳優評判記」として大坂、東京で復活し、それは後の新聞劇評との橋渡しを果たした。(注 1 に同じ)
- 3) 歌舞伎評判記研究会編『歌舞伎評判記集成』第一期全 11 巻 (1972 ~ 1977 刊) に万治 3 (1660) 年 ~ 享保 21 (1736) 年の全 112 点が収録。第二期全 11 巻 (1987 ~ 1995 刊) に元文 2 (1737) 年 ~ 明和 9 (1772) 年の全 93 点が収録されている。2018 年より第三期刊行中。
- 4) 「計量的分析手法を用いてテキスト型データを整理または分析し、内容分析を行う方法である。計量テキスト分析の実践においては、コンピュータの適切な利用が望ましい。」と定義される。テキストマイニングの語で認識されることも多い。(樋口耕一「第一章 内容分析から計量テキスト分析へ」『社会調査のための計量テキスト分析: 内容分析の継承と発展を目指して』ナカニシヤ出版、2014 年)
- 5) 樋口耕一『社会調査のための計量テキスト分析: 内容分析の継承と発展を目指して』ナカニシヤ出版、2014 年・上坂彩香「日本語テキストを対象とした計量的研究の現状」『研究報告人文科学とコンピュータ 2016-CH-112 (6)』情報処理学会、2016 年などに詳しい。
- 6) コーパス中のテキストを単語に分割し、それぞれに対

- して見出し語・品詞・語種などの形態論情報を与えること(伝康晴、小木曾智信、小椋秀樹ほか「コーパス日本語学のための言語資源—形態素解析用電子化辞書の開発とその応用」『日本語科学22』国書刊行会、2007年)
- 7) ある同一の単位設計の元で単位が均質であること。例えば「幾何学」は1語だが、「心理 | 学」は2語というような不揃いな単位認定が行われないよう対処されていること。
 - 8) 異表記や異形態の扱いが正しく行われること。例えば「表す」「表わす」のような送り仮名の違いや「攪乱」「攪乱」のような新旧字体の違いに惑わされず、同じ語とみなさるよう対処されていること。
 - 9) 例えば英語の場合、語と語の間にはスペースがあるため明確な語の区切りを簡単に確認することが出来る。しかしながら日本語にはそのような語の区切りがなく、また読みや表記も多様であるため、分割の難易度が高い。形態素解析辞書 UniDic の例を見ると、単語分割における(1)単位の斉一性のために3種の単位を、(2)見出しの同一性のために4つのレベルで見出しを定義している。このような複雑な設計を要する事実からも、その難しさを窺い知ることができる。
 - 10) MeCab: Yet Another Part-of-Speech and Morphological Analyzer <https://taku910.github.io/mecab/>
 - 11) ChaSen -- 形態素解析器 <https://chasen-legacy.osdn.jp/>
 - 12) IPAdic legacy <https://ja.osdn.net/projects/ipadic/>
 - 13) UniDic <https://unidic.ninjal.ac.jp/>
 - 14) 市村太郎、小木曾智信「文書構造を利用した近世期洒落本の形態素解析」『言語処理学会第22回年次大会発表論文集』言語処理学会、2016年、p. 108.
 - 15) 小木曾智信、市村太郎、鴻野知曉「近世口語資料の形態素解析の試み」『第4回コーパス日本語学ワークショップ予稿集』国立国語研究所、2013年。
 - 16) 国立国語研究所日本語歴史コーパス江戸時代編 https://pj.ninjal.ac.jp/corpus_center/chj/edo.html
 - 17) コーパス検索アプリケーション中納言 <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiQ8Nir9NfuAhUIK6YKHb0NBG0QFjAAegQIARAC&url=https%3A%2F%2Fchunagon.ninjal.ac.jp%2F&usg=AOvVaw0SOCWkqTyTqOpMHVKZAol->
 - 18) 人文系データベース協議会新編西鶴全集データベース <https://www.jinbun-db.com/database/archives/515>
 - 19) 上阪彩香「ジャンルの観点から西鶴作品の文章の数量分析」『じんもんこん2015論文集』情報処理学会・人文科学とコンピューター研究会、2015年、p. 16.
 - 20) 坂梨隆三、小木曾智信ほか『好色伝授:本文・総索引・研究』笠間書院、2000年。
 - 21) 土田衛ほか『歌舞伎評判記集成第一期芸能索引』和泉書院、1994年。
 - 22) 永井規善、前田亮ほか「役者評判記からの人物表現抽出手法の提案」『じんもんこん2014論文集』情報処理学会・人文科学とコンピューター研究会、2014年。
 - 23) 古文用 UniDicS https://unidic.ninjal.ac.jp/download_all#unidic_chj
 - 24) Web 茶まめはサーバー内の形態素解析エンジン MeCab と9種類の UniDic、1種類の IPAdic を用いて形態素解析を行い、その結果を容易に出力できる形態素解析支援アプリケーションである。辞書を2つまで選択し形態素解析を行い、その結果を比較して見ることも可能である。(堤智昭、小木曾智信「歴史的資料を対象とした複数の UniDic 辞書による形態素解析支援ツール『Web 茶まめ』」人文学とコンピューターとシンポジウム、2015年・川口寛治、薦田龍輝、堤智昭「形態素解析ソフトウェア『Web 茶まめ』の改良とWebAPIの試作」『言語資源活用ワークショップ発表論文集1』2017年)
 - 25) 小木曾智信、小町守ほか「歴史的日本語資料を対象とした形態素解析」『自然言語処理20(5)』言語処理学会、2013年など。
 - 26) 『歌舞伎評判記集成第一期』第2巻 p. 182 立役之部 坂田藤十郎条からp. 186 竹嶋幸左衛門条「にた役者も、まひとりあるか云て」まで。
 - 27) 『歌舞伎評判記集成第一期』第2巻 p. 182 立役之部 坂田藤十郎条からp. 190 敵役之部 小野山宇治右衛門条「但シきりやうよすぎ、近年げいの上る」まで。
 - 28) MeCab 単語の追加方法 <http://taku910.github.io/mecab/dic.html>
 - 29) 樋口耕一「資料 A KH Coder リファレンス・マニュアル」『社会調査のための計量テキスト分析:内容分析の継承と発展を目指して』ナカニシヤ出版、2014年、pp. 132-133.
 - 30) 小木曾智信、小椋秀樹ほか「中古和文を対象とした形態素解析辞書の開発」『情報処理学会研究報告 人文科学とコンピューター研究会報告85』情報処理学会、2010年・小木曾智信「中古仮名文学作品の形態素解析」『日本語の研究9(4)』日本語学会、2013年・注25などに詳しい。
 - 31) 注21に同じ。
 - 32) 注25に同じ。
 - 33) 市村太郎、河瀬彰宏、小木曾智信「近世口語テキストの構造化とその課題」『研究報告 人文科学とコンピューター (CH) 2012-CH-96 (1)』2012年。
 - 34) 小林正行、市村太郎「『虎明本狂言集』コーパスの構造化—仕様と事例の検討—」『第3回コーパス日本語学ワークショップ』2013年。
 - 35) 注29に同じ、pp. 121-122.

誤解析例表一覧

- * A, B, C, D, それぞれの左カラムに評判記の表記と正解を示した。「()」内に表記したのが正解である。
- * A, B, C, D, それぞれの右カラムに誤解析結果を示す。
- * 「」は単語境界を表す。
- * 「[]」は該当の語ではないものの、該当単語の分析に影響を与えたと考えられる語を参考として示した場合に用いている。
- * 「・」は複数の解析結果を示す場合の区切りを表す。
- * 両方とも誤解析だったものは、基本的には「・」で区切った。先に表記するのが近世口語 UniDic の、後に記すのが中古和文もしくは中世口語 UniDic の誤解析結果である。ただし複数種の誤解析結果や各辞書が異なる箇所でも同じように間違えている場合もあるため、変則的に複数の誤解析例を示したこともある。
- * 歌舞伎に強く関係する語は太字で表記した。

表1：近世口語（洒落本）UniDic と中古和文 UniDic 誤解析例

	A：境界認定			B：品詞認定		C：語彙素認定		D：発音形認定	
近世口語× (中古和文◎)	1	にせて(似せる)で	に 為る て	「世間」で	なり	「惣領」めいて(めく)で	召す て		
	2	見あかれず(見る)趣く(れる)す	散れる(アカレル) す			よい(良い)	酔う		
	3	やつし(贅す)	奴 為る			見とれ(見惚れる)	見取れる		
	4					「此」たび(度)	旅		
	5					面度(二 度(フタタビ))	面度(リヨウド)		
	6					四(ばん)め(四 番 目)	四 番 奴		
	7					大根(オオネ)	ダイコン		
中古和文× (近世口語◎)	8	あいざつ(あいざつ)	秋沙 つ	「成まじき事」也	(なり)	尤可也(尤も べし なり)	咎める べし や	「ござれ」共(ども)	とも
	9	所有(所 有る)	所有(シヨウウ)			今(迄)(今)	今日	古今「無類」(コゴン)	コキン
	10	云かけ(言い掛ける)	言う 掛ける			「申さるる」通に(通り) に	道(ミチ) に		
	11	ぶ得手 不 得手	はぶ びる 得手			「訴訟」人(ニン)	ヒト		
	12	しうしんじや(執心 じや)	周 シンジ や			見物(ケンブツ)	ミモノ		
	13	きりやうの(器量 の)	きり 様 の			太夫(タユウ)	タイフ		
	14	「一りう」だし(仕出す)	仕 出す・き 出す			前評あしき(前 評 悪い)	前(マエ) 測る 悪い		
	15	役者は(役者 は)	訳す や は			後家がた(後家 方)	後家 形		
	16	かほみせの(顔見世 の)	顔 店 の			あてこと(当て 事)	貴 事		
	17	大じん(大臣)	大 陣			くるは(廓)	狂う		
	18	どのし(どの)の(芝居)	殿 為る ば 居			けいせいかたる(傾城 買) たり	傾城 変える たり		
両方×	19	風儀(風儀)	カゼ 義・カゼ ヨシヤ	「大やうに」そ(たち)たる」(ぞ)	そう・そ	「巻頭(こ)おく(置く)	奥・起さる		
	20	さなから(宛ら)	さ 無い・然 半ら			「けいせい」本間」ごと(事)	毎		
	21	一くはん(一貫)	苦 さん・九 版			間(狂言)(アイ)	アイダ・マ		
	22	名高き(名高い)	ナダカ 機・黄			御家(オ イイエ)	ゴ カ・オオン イエ		
	23	一チ(一)	イチ チ			美(に)(シツ)	ケニ・ミ		
	24	大しん(大臣)	大心・大深・大 為る ず			小ざつ(小ざつ)	小 ざつ・小座		
	25	しぐみ(仕組)	き 組			けいせいかい(傾城 買)	傾城 戒・階		
	26	一りう「しだし」(一流)	一 般・一人 丑						
	27	ねおい(根生)	根 甥・音 おい						
	28	第一	第一・第 一	「打ぞろふ」た	た・たり	情	ジョウ・ナサケ	正月	シヨウガツ・ムツキ
	29	一年	一年・一 年			御事	オ・オオン・オン		
	30	一所	ヒトコロロ・ヒト トコロ						

表2：近世口語(洒落本)UniDicと中世口語UniDic(言狂)UniDic誤訳解析例

	A：境界認定	B：品詞認定	C：語彙素認定	D：発音形認定
近世口語× (中世口語◎)	1 似せる(似る せる) に 為る	(大やうに)そ(たちたる)(ぞ) そう	見とれ(見惚れる) 見取る	下男(ゲナン) シモオトコ
	2 おとしと(御 年 と) 落とし と	(あまでらあみがさ)にて(にて) なり て	(あたつたに)より(因る) 寄る	
	3 一画年(一画 年) 一 画年		(此)たび(度) 旅	
	4 役目(役目 高 嵌まる たり)	役目 高 嵌まる たり	ほんさま(坊 様) 盆 様	
	5 一チ(一) イチ チ	イチ チ	(あんばい)よう(良い) 酔う	
	6 やつし(簀す) 奴 為る	奴 為る	思慮(分別)(思慮) 慮り	
	7 替名(替り 名) 代わり 名		大根(オオネ) ダイコン	
	8		御家(オ イエ) ゴ カ	
	9 でけた(出来る た) で 蹴る た	(げいしゃ)だし(にて(なり)て) にて	かる(すぎで)(軽い) 借る	一「あで」(ヒト) イチ
	10 よの(よこむ 良く 飲み込む) 浴 の 見込む	(町中)で(で) なり て	(近年に)なき(無い) 鳴く・泣く	古今「むるい」(コゴン) コキン
	11 (狂言)めかず(めく ず) 奴 数		人「のにくまぬ」(ヒト) ニン	
	12 (男)つきよき(付き 良い) 月夜 気		外「(の)じつごとし」(ホカ) ガイ	
	13 知る(知)り 知る 片		あん「(たる)事」(有る) 編む	
	14 きかし(しよ 聞かす ます) 聞く します よ		女郎(ジョロウ) メロウ	
	15 (面白)所有(所 有)り あらゆる		くるは(郎) 狂う	
	16 やつし(簀す) やる き・奴 詩		太夫(タユウ) タイフ	
	17		(後家)がた(方) 形	
両方×	18 ざなから(宛ら) 然 ない・最中 等	かぎらず(限る ず) 限る せる・候	かんがへおほ(せ 考え 果せる) アケル(アクル)・サヤ	
	19 ざつぱり(ざつぱり) ざ つ 張り・針		明ル(アケル) アケル・サヤ	
	20 (紙子)一くはん(一貫) 苦 さん・九 版		(ゆびを)所(せり)(所す) 習む・疲らす	
	21 かいぎやう(戒行) 会議 様		ま「(ひと)り」(もう) 真・舞	
	22 御覧(やる 御覧する ある) 御覧する 遣る・やる		大手のもん(大手) の 門 オオテ の 物・者	
	23 申(申す) 申す ス・申す です		(しばい)しり(知り) 尻・後	
	24 (先)口(口跡) 先口 セキ・関		ぬれごと(濡れる 事) 濡れる ごとし 濡れ 毎	
	25 しやみせん(三味線) 三味 線(三味 前)			
	26 ねおい(根生) 根 勢 追う 音 勢 追う			
	27 (盆)がはり(替り) が 張り・が 針			
	28 しぐみ(仕組) き・為る 組			
	29 したし(仕出し) 為る たす・き 出し			
	30 へんげ(変化) 変 化・辺 気			
	31 小ざつ(小ざつ) 小 ざつ・少し ざつ			
	32 一流(一流) 一 竜・一 粒			
	33 (其上)山下流(山下 流) カミヤマ 下流・山 下流			
	34 はりまや(播磨屋) 張り 馬屋			
35 やつしごと(簀す 事) 奴 仕事				
36 あげやへ(揚屋 へ) 上げる やる・へ 上げる 八通				
37 しょさ(所作) 為る さ・書 さ				
要検討	38 第一	第一・一・第一	情	今年
	39		上つ「たと」	アガル・ノボル
	40	(ありそふ)な	尤	尤も・最も
				コトシ・コンネン
				イチ・ヒトツ