『虎明本狂言集』コーパスの構造化　－仕様と事例の検討－

小林 正行（群馬大学 教育学部）
市村 太郎（国立国語研究所 コーパス開発センター）

Structuring the Corpus of *Toraakira-bon Kyogen*

Masayuki Kobayashi (Gunma University)
Taro Ichimura (National Institute for Japanese Language and Linguistics)

１．はじめに

国立国語研究所「通時コーパス」プロジェクトの一環として検討されている『虎明本狂言』の電子化について、資料の電子化に際し、いかなる要素を認定し、どのように構造化するのが適切かについて検討し、モデルを示す。

狂言テキストは演劇資料であり、台詞と卜書きから成る台本本文を中心とし、さらに舞台外の要素として注釈が付されることがある。底本である大塚光信編『大蔵虎明能狂言集翻刻註解』（2006，清文堂）は原資料に付された情報をよく残したまま活字化し、さらに原本にはない要素を付加している。

本発表では、多様なテキストの段階を持つ『虎明本狂言集』のタグセットや処理方針を示し、いくつかの例を提示する。

なお本発表では便宜上『大蔵虎明能狂言集　翻刻註解』を「底本」と呼ぶこととする。

２．『虎明本狂言集』コーパス化の意義

狂言は、中世から近世にかけての言語資料として重要な位置を占めている。登場人物が多彩で身分関係が明確であること、対話劇の形で進行し場面・状況が明確であることから、口語資料としての価値は極めて高い。

狂言資料の中でも『虎明本』は、寛永 19 年（1642）大蔵流十三世宗家大蔵弥太郎虎明の手による大蔵流の祖本である。本狂言 237 曲を収めており、狂言の類別や詞章の整備された台本として、質・量とも第一級の資料である。その詞章には、中世、室町時代の言葉を伝承していると見られる点、書写当時である近世初期の日常語の影響を受けたと思われる点、舞台言語として整理され固定化・類型化する兆候が見られる点がある。狂言史上の位置を踏まえ、他の台本との比較ということが不可欠であるが、注釈書や総索引が整備され、中世から近世の言語資料として広く利用されてきている。

しかし、刊行されている『大蔵虎明本狂言集総索引』は、狂言の類別に合わせた 8 分冊の形をとっており、単語認定の基準にばらつきがある。一定の基準でアノテーションされた形態論情報付コーパスの完成は、狂言の言語の研究だけにとどまらず、中世から近世初期にかけての言語研究に大きな成果をもたらす。

３．コーパスの設計方針

本研究では、コーパスの主な利用者として、言語研究者を想定する。そのため、言語的に重要な、文と短単位（ほぼ語に相当）が基本的な単位となる。

底本は、『大蔵虎明能狂言集　翻刻註解』上下巻を用いる。最新の活字本文であり、注記・ミセケチ等原本の情報を反映させることに配慮されており、また読みの指示など、詳細な注記がなされている。

本研究では、そのような底本の状況をできるだけ反映しつつ、単なる文字列の電子化ではなく、どこで得られた、どういう要素の、どのような性質を持つ語の表記体であるという情報が付された用例の一覧を、短時間で取り出せるようなコーパスを目指している。

そのため、底本内の各文書要素について XML を用いて記述し、国語研が作成した『太陽コーパス』の仕様や BCCWJ の仕様、『明六雑誌コーパス』の仕様を継承しながら、TEI P5 を参考に必要なタグを選択・追加し、構造化する。市村・河瀬・小木曽(2012)では、洒落本コーパスも含め、「近世口語テキスト」として、共通の基礎的な構造化案を示したが、本発表では、さらに実際の作業の過程で現れた問題を基に、タグ仕様を再設定する。

構造化されたデータには、さらに品詞情報や活用形等、形態素レベルで情報を付与する。なお、各演目はそれぞれ作品としては独立しているため、１演目を１テキストとする。

４．狂言テキストの構造とタグセット

狂言テキストは、台本本文を中心とし、その前後にはしばしば注釈が付される（図 1)。

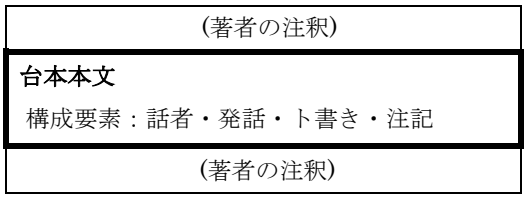

図 1　狂言テキストの構造概略

各々独立した演目ではあるが、全体として筆者は同一であり、形式や言語的状況は比較的安定している。

台本であるため、読み物とは異なり、序や後書がつかず注釈が多くなる。また当然台詞とト書きが中心となる。会話文に付記される話者の表示は、原著者によるものと、校注者によるものがあり、会話文の前後や合間にト書きが付される。

本コーパスと並行して「通時コーパス」プロジェクトでは『洒落本大成』のコーパスの設計も進められているが、文書の構造を比較すると、話者・会話文と割書きで主に構成される洒落本とはある程度の類似性があるといえる。そのため、洒落本大成コーパスの仕様との共通化を図り、基本的には共通のタグセットで表現する。

一方で、台詞やト書き、本文校訂や書き入れ・注釈という、舞台台本である狂言ならではの要素については、新たに要素を設定し、運用も改める。

以下、各要素について詳説する。(なお DTD については、同日発表の「洒落本コーパスの構造化」にある図を参照されたい。)

４．１文書の構造に関する要素

表 1　文書の構造に関するタグ（太線は階層上の大きな切れ目）

タグ（要素）	説明	属性
<text>	作品（演目）全体、作品のシリーズ・タイトル等を開始タグ内に記述	@textID（必須） @series シリーズ名（必須） @title　作品タイトル（必須） @yomi 作品名の読み（任意） @year　西暦成立年（必須） @year_w 和暦成立年（必須）
<front>	前付け部分 （狂言の場合は原則<titleBlock>のみ）	
<body>	主本文	

<article>	記事	@type（任意）
<titleBlock>	<article>レベルでのタイトル等の記述	
<p>	タイトルや注釈等を除く本文の塊	
<block>	<p>で記述された本文とは区別されるタイトル・注釈等のブロック要素	@type（必須）
<s>	文	
<SUW>	短単位	（多岐にわたるため省略）

文書構造に関する基本的な要素は洒落本と共通である。テキスト全体を表す<text>と、それを構成する<front><body>から成る。作品に関する情報は、属性値で<text>内に記述する。

さらに、内部は<titleBlock>と<article>に分割され、さらに本文に<p>、注釈に<block>を付す。これらはさらに<s>に分割され、文は形態論情報を記述した<SUW>に分割される。

なお作品ごとに序文・後書きが付されることはないため、実質台詞・ト書き・注釈による大きな<body>と、タイトルのみの<front>で構成される。また作品内に小見出し等はなく、洒落本等に比べればシンプルな文書構造といえる。

article 要素　前付・後付を除いた中心的本文は、小見出し等を伴う複数の要素から成ることがあり、このような階層の要素を表すものとして、<article>を用いる。狂言の各作品には小見出し等は見られず、実質テキストのタイトルを除くすべての部分が該当する。

p 要素　<article>内の本文の塊全体で付与する。視覚上、また内容上いわゆる段落を認定するのは困難である。本研究では「主たる本文かそれ以外か」に重点をおいている。

block 要素　視覚上また構成上、明らかに主本文の塊と区別される要素を表す。type 属性で、タイトル・著者・日付・注釈等の別を記述する。

狂言では、演目内的な会話文とト書きを「主本文の塊」と見、演目外的な、また追加的な情報を付加する注釈を「本文の塊と区別される要素」と見る。

titleBlock 要素　テキストのタイトル箇所に付与する。狂言単独で見たときには<block>タグのみでも事足りるが、共通仕様をめざす洒落本では内題が現れることがあり、<article>と同階層でマークアップされるため、それに合わせて本要素を付与する。

s 要素　すべてのテキストは文に分割される。ただしいわゆる「文」とは完全に同一ではなく、発話や割書の区切りでも切る。なお、<s>が<s>を含むような階層性は認めない。

SUW 要素　短単位（おおよそ語に相当）を表す。すべての文は短単位に分割される。本研究での基本的な単位である。語彙素・語形・書字形・活用型・活用形・発音形等語に関する多くの情報が、属性で記述される。開発中の「近世口語 UniDic」による解析結果を人手で修正して付与する。

<text textID="虎明本狂言_034_大名_入間川" series="虎明本狂言・大名狂言之類#34" title="入間川" year="1642" year_w="寛永 19"><front><titleBlock><block type="title"><s><pb n="170"/><lb/>入間川</s></block></titleBlock></front>
<body><article> <speech><s><speaker value="大名"/><lb/>「罷出たる者<kana>は</kana>、<hi rend="傍線">東</hi>国
<info text="「はるかおん国ともなのる"/>にかくれもなひ大名です、</s><s><ruby resp="annotator" rubyText="訴">そ
</ruby><ruby resp="annotator" rubyText="訟">せう</ruby>の事有て永々<lb/>在京仕る処に、<ruby resp="annotator"
rubyText="あん">安</ruby><ruby resp="annotator" rubyText="ど">堵</ruby>の<ruby resp="annotator" rubyText="み">御
</ruby><ruby resp="annotator" rubyText="げう">教</ruby><ruby resp="annotator" rubyText="しよ">書</ruby>をいた<odoriji
originalText="ゞ">だ</odoriji>き、殊にお<ruby resp="annotator" rubyText="いとま">暇</ruby>を下された程に、急でく<lb/>
だらふと存る、</s><s>太郎くわじやあるか</s></speech><speech><s><speaker value="太郎冠者"/>「お前に
</s></speech><speech><s><speaker value="大名"/>「いそひで<ruby resp="annotator" rubyText="立">た</ruby><lb/>て
</s></speech><speech><s><speaker value="太郎冠者"/>「御<ruby resp="annotator" rubyText="機">き</ruby><ruby
resp="annotator" rubyText="嫌">げん</ruby>が<ruby resp="annotator" rubyText="良">よ</ruby>う御ざある
</s></speech><speech><s><speaker value="大名"/>「その事よ、</s><s>訴<corr type="erratum" originalText="詔"
resp="annotator">訟</corr>こと<ruby resp="annotator" rubyText="悉">／＼</ruby>く<lb/><ruby resp="annotator" rubyText="
安">あん</ruby><ruby resp="annotator" rubyText="堵">ど</ruby>し、おいとまを下された<kana>は</kana>
</s></speech><speech><s><speaker value="太郎冠者"/>「やれ／＼それ<kana>は</kana>めでたひ事で<lb/>御ざる、</s>

図 2　作品冒頭部分の形式化例（上巻『入間川』p.170）

<speech><s><speaker value="入間"/>「扨<kana>は</kana><hi rend="傍線">いるま</hi>やうをのけ<lb/>てか
</s></speech><speech><s><speaker value="大名"/>「中／＼</s></speech><stage><s><add>二度云て三度め程に
</add></s></stage><speech><s><speaker value="入間"/>「<ruby resp="annotator" rubyText="存">ぞん</ruby>じも<ruby
resp="annotator" rubyText="寄">よ</ruby>らぬに、色々の物を<ruby resp="annotator" rubyText="貰">もら</ruby>ふて、うれ
<lb/>しうなひと申事<vMark>が</vMark>ござらふぞ、</s><s>身にあまつてかたじけなふ御ざる</s></speech><stage><s>
「と云て<lb/>いた<odoriji originalText="ゞ">だ</odoriji>く</s></stage><speech><s><speaker value="大名"/>「身にあまつて
<ruby resp="annotator" rubyText="かたじけない">忝</ruby>とおしやる<kana>は</kana>、うれしうなひといふ事じ<lb/>や
程に、こちへお<ruby resp="annotator" rubyText="返">かや</ruby>しやれ</s></speech><stage><s>「と云て皆とりかへす
</s></stage><speech><s><speaker value="入間"/>「あのたらしが、<pb n="176"/><lb/>やるまひぞ／＼
</s></speech><stage><s>「と云ておいいるなり</s><s>「太郎くわじや<kana>は</kana>、太刀を主にわたしてひ<lb/>つこ
む</s></stage> <block type="注釈"><s><lb/>「私<kana>に</kana>云、右つめのこと<kana>は</kana>何共がてんのゆき
かたき事也、</s><s>然共<hi rend="傍線">いるま</hi>やうをのけて<lb/>といふ<kana>は</kana>、のけいでと云事を、ま
こと<odoriji originalText="ゝ">と</odoriji>心得ていふたによつて取かへした<kana>は</kana>こと<kana>は</kana><lb/>
りなり<info originalPage=""/></s></block><block type="注釈"><s><lb/>一</s><s>いる<corr type="omission"
resp="annotator">ま</corr>やうをのけてと云<kana>は</kana>、のけ<add>い</add>でと云ことじやにと云へ
<vMark><kana>ば</kana></vMark>よくきこへ候へ共そ<lb/>れにて<kana>は</kana>人がしるによつていわぬがよき也
</s></block></article></body></text>

図 3　作品末尾の形式化例（上巻『入間川』pp.175-176）

キー	語彙素	出現形発音形	品詞	解析活用型	活用形
「	「		補助記号-括弧開		
罷	罷る	マカリ	動詞-一般	文語四段-ラ行	連用形-一般
出	出でる	イデ	動詞-一般	文語下二段-ダ行	連用形-一般
たる	たり	タル	助動詞	文語助動詞-タリ-完了	連体形-一般
者	者	モノ	名詞-普通名詞-一般		
は	は	ワ	助詞-係助詞		
、	、		補助記号-読点		
東国	東国	トーゴク	名詞-普通名詞-一般		
に	に	ニ	助詞-格助詞		
かくれ	隠れ	カクレ	名詞-普通名詞-一般		
も	も	モ	助詞-係助詞		
なひ	無い	ナイ	形容詞-非自立可能	形容詞	連体形-一般
大名	大名	ダイミョー	名詞-普通名詞-一般		
です	です	デス	助動詞	助動詞-デス	終止形-一般
、	、		補助記号-読点		
そせう	訴訟	ソショー	名詞-普通名詞-サ変可能		
の	の	ノ	助詞-格助詞		
事	事	コト	名詞-普通名詞-一般		
有	有る	アリ	動詞-非自立可能	文語ラ行変格	連用形-一般
て	て	テ	助詞-接続助詞		
永々	長々	ナガナガ	副詞		
在京	在京	ザイキョー	名詞-普通名詞-サ変可能		
仕る	仕る	ツカマツル	動詞-一般	文語四段-ラ行	連体形-一般
処	所	トコロ	名詞-普通名詞-副詞可能		
に	に	ニ	助詞-格助詞		
、	、		補助記号-読点		
安堵	安堵	アンド	名詞-普通名詞-サ変可能		
の	の	ノ	助詞-格助詞		
御	御	ミ	接頭辞		
教書	教書	ギョーショ	名詞-普通名詞-一般		
を	を	オ	助詞-格助詞		
いただき	頂く	イタダキ	動詞-非自立可能	文語四段-カ行	連用形-一般
、	、		補助記号-読点		
殊に	殊に	コトニ	副詞		
お	御	オ	接頭辞		
暇	暇	イトマ	名詞-普通名詞-一般		
を	を	オ	助詞-格助詞		
下さ	下す	クダサ	動詞-一般	文語四段-サ行	未然形-一般
れ	れる	レ	助動詞	助動詞-レル	連用形-一般
た	た	タ	助動詞	助動詞-タ	連体形-一般
程	程	ホド	名詞-普通名詞-副詞可能		
に	に	ニ	助詞-格助詞		
、	、		補助記号-読点		
急	急ぐ	イソイ	動詞-一般	文語四段-ガ行	連用形-イ音便
で	で	デ	助詞-格助詞		
くだらふ	下る	クダロー	動詞-一般	文語四段-ラ行	意志推量形
と	と	ト	助詞-格助詞		
存る	存ずる	ぞんずる	動詞-一般	文語サ行変格	連体形-一般
、	、		補助記号-読点		
太郎	太郎	タロー	名詞-普通名詞-一般		
くわじや	冠者	カジャ	名詞-普通名詞-一般		
ある	有る	アル	動詞-非自立可能	文語ラ行変格	連体形-一般
か	か	カ	助詞-終助詞		
「	「		補助記号-括弧開		
お	御	オ	接頭辞		
前	前	マエ	名詞-普通名詞-副詞可能		
に	に	ニ	助詞-格助詞		

図 4　短単位解析済みデータの例（一部項目省略・上巻『入間川』p.170）

４．２　文・語の機能に関する要素

表 2　文・語の機能に関する要素

タグ（要素）	説明	属性	
<speech>	会話	@source（任意） @type（任意）	
<quotation>	①単純な発話以外の引用要素 ②ト書き内の台詞指示等	@source（任意） @type（任意）	
<stage>	ト書き		↑文以上
<speaker> <speaker/>	話者 （校注者付記の場合は空要素）	@value（任意）	↓文未満
<delivery>	発話等のスタイルの表示		
<verse>	韻文		

speech 要素　1 回的な会話文の連続を表す。<speaker>を発話の内部に認定し、一体として扱う。会話文内に話者が示されていない場合には@source 属性で話者を可能な限り記述する。また、底本では、紙面上（）付で校注者により話者が示されことも多く、それについては空要素とし、@value 属性で話者を記述する。

quotation 要素　手紙や和歌等、単純な会話文以外の引用要素を表す。@type 属性でどのような種の引用かを、@source 属性で出典を記述する。

　また、しばしば現れるト書き内の台詞指示等は、本来階層は異なるが、本要素で記述する。その場合、基本的に話者表示がないため、@source 属性で話者を記述する。

stage 要素　本文内的な要素としてト書きを表す。狂言は舞台演劇であり、台詞とト書きが比較的明確に分かれるのが特徴である。ト書きは時に内容としては本文外的な挿入的なものあり、この点注釈と重なるのだが、会話と会話を割って、または会話に付属して述べている点において、本文の塊の外側に付される注釈ほどの独立性はない（つまり階層的には別の次元のもの）と見る。そのため、内容が注釈的なト書きであっても、会話に割って入る以上は、本文内的な要素であって、<block>扱いはしない。

speaker 要素　会話文に付属する、小書き等で記される話者の表示である。底本では原作者による話者表示のほか、校注者が補った（　）付の話者がある。これらは、原作者の表示と区別するため空要素とし、@value 属性内に記述する。

delivery 要素　台詞の内部には、話者だけでなくその台詞のスタイルを小書き等で記してある場合がある。狂言においては、「舞がけり」など、散文資料の発話に比べ細かく台詞指示がなされており、重要な要素である。

verse 要素　韻文は、歌・舞等について、文未満の単位で付与する。

４．３　語・文字単位で外形・機能等を表す要素

『虎明本狂言集』は、筆者による本文修正や、テキストの追加・削除の指示が頻繁に行われるという点で特徴的である。

また、底本には『洒落本大成』とは異なり、校注者による誤りの指摘（ママ注）や、校注者が追加した振り仮名等があって、原資料の筆者の指示と、校注者による情報との 2 段階で記述し分ける必要がある。

そのため、「洒落本コーパス」の仕様に比べ、本文校訂に関する記述が詳細である。

表 3　語・文字単位で外形等を表す要素

タグ(要素)	説明	属性	
<hi>	文字列（語）に対する装飾	@rend（必須）	↑短単位以上
<lRuby>	左ルビ	@rubyText（必須） @rubyBase（任意） @resp（任意）	↓短単位未満
<ruby >	ルビ	@rubyText（必須） @rubyBase（任意） @resp（任意）	
<odoriji>	踊り字を開いた文字	@originalText（必須）	
<gap/>	抹消・破損等で判読できない文字の存在（空要素）		
<corr> <corr/>	本文修正	@type（必須） @originalText（任意） @resp（任意）	
<unclear>	推読された文字	@originalText（任意） @type（任意）	
<vMark>	濁点付仮名に変換した箇所		
<g>	外字	@type（必須） @ref（任意）	
<kana>	片仮名を平仮名に変換した箇所		
<add>	著者によって追加されたテキスト		
<kanbun> <kanbun/>	漢文（返読）	@type（任意）返読前 返読後 @originalText（任意） @id（任意）	

hi 要素　傍線が付される、小書きされるなど、外形的特徴を持った文字列（語）を表す。狂言では固有名詞に傍線が引かれるケースがあるが、必ずしも機能は一定ではない。

ruby 要素　文字列の右側に付され、文字・文字列の読み等を表す振り仮名等を指す。@rubyText 属性内にルビ文字列が記述される。右側漢字傍記も含む。

凡例によると、原資料に付されている振り仮名・漢字傍記(A)については＜＞が付されており、校注者によって新たに付されたもの(B)には何も付されていない旨の記述がある。そのため、@resp 属性で校注者により付与されたものを区別する。

(A)　<ruby rubyText="〈ソサノヲ〉">素盞烏</ruby>（上巻『ゑびす大黒』p.6）

(B)　<ruby resp="annotator" rubyText="戯">ざれ</ruby>事（上巻『連歌毗沙門』p.10）

lRuby 要素　文字列に沿って小書きされる文字は、右側の振り仮名だけでなく、左側に付されることもある。rubyText 属性内にルビ文字列が記述される。

corr 要素　本文テキスト修正箇所であり、文字単位で付す。狂言の場合、本文テキストの正誤にかかわる指示としては、ミセケチ等による筆者の校訂箇所と、ママ注によって校注者が誤りを指摘している箇所の 2 種があり、特徴的なものと言える。これらは区別すべきものであり、また原文を確認できる形にすることが重要である。一方で、形態論情報を付すことを考慮すると、本文としては「きれいな本文」であることが望ましい。

まず@type 属性で誤字（erratum）・衍字（excess）・脱落（omission）の別を付し、本文

は修正後の形とするが、@originalText 属性で元のテキストを記述する。また@resp 属性で筆者（writer）の指示によるものか、校注者（annotator）の指摘によるものかを記述する。

校注者の指摘するママ注には、いかなる誤りかが頭注に明記されず、推測困難な場合がある。そのような箇所は修正せず、@type 属性で「修正なし」と記述する。

```
<s>今日<ruby resp="annotator" rubyText="最">さい</ruby><ruby resp="annotator" rubyText="上">じやう</ruby>吉<corr originalText="日日" type="excess" resp="annotator">日</corr>でござる<lb/>により、聟殿のおいでなされうずるとのおこ<corr type="omission" resp="annotator">と</corr>じや</s>
```

図 5 ママ注の形式化例（上巻『鶏聟』p.353　下線は筆者）

vMark 要素　底本にはなく、電子化に際して新たに濁点を付与した箇所に付与する。ただし踊り字箇所はもとのテキストを属性値に記録するため、タグ付け対象とはしない。

add 要素　筆者による傍記や「〇」符号等によって、挿入指示がなされた本文に付与する。文字単位から複数文単位まで多岐にわたり、強調表示も含む。文を超える単位で挿入指示がなされる場合は文単位で付与し、短単位未満の場合は、文字単位で付与する。

```
<speech><s><speaker value="えびす"/><delivery>かたり</delivery></s><s>「夫<hi rend="傍線">ゑびす</hi>三郎殿といつ<vMark><kana>ぱ</kana></vMark>、<hi rend="傍線"><ruby resp="annotator" rubyText="伊弉諾">いざな<vMark>ぎ</vMark></ruby></hi><hi rend="傍線"><ruby resp="annotator" rubyText="伊弉冉">いざな<kana>み</kana></ruby></hi>の<ruby resp="annotator" rubyText="尊">みこと</ruby>、あ<info originalPage=""/>ま<lb/>の岩くらの<ruby resp="annotator" rubyText="苔">こけ</ruby><ruby resp="annotator" rubyText="蓆">むしろ</ruby>にて、<hi rend="傍線">男</hi><hi rend="傍線">女</hi>の<ruby resp="annotator" rubyText="語">かた</ruby>らひをなし、日神月神、<ruby resp="annotator" rubyText="蛭">ひる</ruby><ruby resp="annotator" rubyText="子">こ</ruby><lb/><ruby resp="annotator" rubyText="素盞烏">そさのお</ruby>の御子をまうけ給ふ、<hi rend="傍線">ひるこ</hi>と<kana>は</kana>某が事、</s><s><info text="○"/><add><info text="○"/><kana><hi rend="傍線">天照太神</hi>より三番めのをと／＼成<kana>れ</kana><vMark>ば</vMark>とて、<hi rend="傍線">西の宮</hi>の<hi rend="傍線">ゑびす三郎</hi>殿といは<odoriji originalText="ゝ">は</odoriji>れ</kana></add>うぢ<info text="氏〈ウジ〉"/><corr originalText="し" type="erratum" resp="writer">す</corr><corr originalText="ゆ" type="excess" resp="writer"/><ruby rubyText="〈図性〉">じやう</ruby><info text="種〈シユ〉 姓〈ジヤウ〉"/><ruby resp="annotator" rubyText="誰">たれ</ruby><lb/>にか<ruby resp="annotator" rubyText="劣">おと</ruby>りたまふべき、</s><s>なんぼういみじき<ruby resp="annotator" rubyText="位">くらい</ruby>にて<kana>は</kana>なきか、よく／＼<lb/><ruby resp="annotator" rubyText="信">しん</ruby><ruby resp="annotator" rubyText="仰">がう</ruby>せよ、</s><s><ruby resp="annotator" rubyText="楽">たのし</ruby>うなさうずるぞ<lb/></s></speech>
```

図 6 複雑な注記・本文訂正等の形式化例（上巻『ゑびす大黒』p.4　下線は筆者）

４．４位置情報と本文外情報

表 4　底本テキストの位置情報を表すタグ

タグ(要素)	説明	属性
<pb/>	ページ開始（空要素）	@n（必須）
<cb/>	段開始（空要素）	@n（必須）
<lb/>	行開始（空要素）	
<info/>	本文外情報（空要素）	@originalPage（任意）@text（任意） @type（任意）　@originalText（任意）

info 要素　本文外の情報を空要素**<info/>**で表す。底本には影印の改ページが付されており、その位置情報を@originalPage 属性で記述する。また注などの傍記が本文脇に付されることがあるが、本文外に相当する傍記・注記等は@text 属性で記述する。

５．コーパス化に向けての課題

５．１　本文認定と読み順の確定

『虎明本狂言集』には、筆者・校注者による本文に関わる多くの校訂や注記の情報があり、複雑な状況を呈している箇所もある。上欄や本文末に付された挿入指示については、場所が指定されていない場合、内容によって適切な挿入箇所を定めなければならない。

また、同じ傍記であっても、本文に追加する要素、本文を訂正する要素、注記と多様で、また必ずしも校注者による言及があるわけではなく、個別に検討・判断する必要がある。

何が本文で何が本文でないか、また、どの順で読むべきか等は、基本的なことではあるが、本文を決めなければならないコーパスにおいて大きな課題である。

５．２　解釈の問題

舞台台本であるため会話の切れ目が比較的わかりやすく、また底本には校注者の詳細な注が付されているため、近世散文資料等に比べれば文認定は容易である。しかし「。」によって文が区切られているわけではなく、また間接引用か直接引用かがはっきりせず、文認定が困難な箇所は存在する。文認定や現代語訳が行われていない資料を扱う際の共通の課題である。

また、２で述べたように、言語的な状況は中世語に近いとされるため、濁点付与に関しては、タグを付与するとはいえ慎重を期する必要がある。例えば、現代では濁音で発音されるものでも、清音で読んでおくべきものがしばしばみられ（「かがやく」―「かかやく」など）、これらは『日葡辞書』等の記載を参照するなどし、個別に検討する必要がある。

６．おわりに

『虎明本狂言集』では、主に版本が主体である洒落本とは異なり、筆者の本文に対する校訂や補遺・補入が随所に見られる。底本においてもそれがよく反映されており、また校注者によって追加された要素も多く見られる。今後、本文校訂に関する多様な要素を持つ資料を対象とするにあたって、文書構造としてどのレベルまで想定し、記述するのかが課題となる。本研究で言えば、傍記等の中での振り仮名等を構造化するのは現状では困難であり、このようなものの扱いをどうするのかは今後の課題である。

また、さまざまなレベルで出現する補遺・補入の類の扱いは、階層構造を前提とする XML を用いる以上、資料ごとに検討され続けなくてはならない課題であろう。

日本語史資料として、狂言はもちろん、浄瑠璃・歌舞伎等の舞台資料は極めて重要である。本研究での検討は、「日本語歴史コーパス」構築に向けて、これら舞台作品を含めた仕様を作る上での足掛かりになると考える。

文　献

市村 太郎、河瀬 彰宏、小木曽 智信(2012)『近世口語テキストの構造化とその課題』情報処理学会研究報告　人文科学とコンピュータ研究会報告(CH96) pp.1-8

大塚光信編(2006)『大蔵虎明能狂言集　翻刻註解』清文堂

北原保雄、村上昭子、鬼山信行、小川栄一、山崎誠、吉見孝夫、土屋博映、大倉浩編(1983-1989)『大蔵虎明本狂言集総索引』1-8　武蔵野書院

近藤明日子、田中牧郎「『明六雑誌コーパス』の仕様」『国立国語研究所共同研究報告 12-03 近代語コーパス設計のための文献言語研究 成果報告書』pp.118-143　国立国語研究所

近藤泰弘(2012)「日本語通時コーパスの設計について」『国語研プロジェクトレビュー』3 pp.84-92 国立国語研究所

田中牧郎(2005)「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」『国立国語研究所報 122　雑誌『太陽』による確立期現代語の研究　『太陽コーパス』研究論文集』pp.1-48　博文館新社

田中牧郎、小木曽智信(2000)「総合雑誌『太陽』の本文の様態と電子化テキスト」『日本語科学』8　pp.141-152　国立国語研究所

安永尚志(1998)『国文学研究とコンピュータ』勉誠社

山口昌也、高田智和、北村雅則、間淵洋子、大島一、小林正行、西部みちる(2011)『特定領域研究「日本語コーパス」平成 22 年度研究成果報告『現代日本語書き言葉均衡コーパス』における電子化フォーマット ver.2.2』　文部科学省 科学研究費 特定領域研究 「日本語コーパス」データ班

関連 URL

「Text Encoding Initiative」（ガイドライン P5 日本語版）
http://docsci.infon.org/stack/P5JA/index-toc.html