

公開シンポジウム

京都大学人文科学研究所
共同研究プロジェクト

情報の構造と メタデータ

情報処理技術は
漢字文献から
どのような情報を
抽出できるか

—人文情報学の
基礎を築く—

マンガにおける異本研究 (安岡孝一)

TEIテキスト・モデルの今昔 (Christian Wittern)

漢字文献における電子的翻刻の課題
——或いは翻刻者の使命—— (白須裕之)

CiNiiのメタデータ・デザイン (大向一輝)

全国共同利用・共同研究拠点
「人文学諸領域の複合的共同研究国際拠点」
2012.2.24

目次

この冊子の説明	... p. 1
シンポジウムについて	... p. 2

マンガにおける異本研究／安岡孝一（京都大学人文科学研究所）	... p. 3
TEIテキスト・モデルの今昔／Christian Wittern（京都大学人文科学研究所）	... p. 21
漢字文献における電子的翻刻の課題——或いは翻刻者の使命——／白須裕之	... p. 37
CiNiiのメタデータ・デザイン／大向一輝（国立情報学研究所）	... p. 51

この冊子の説明

この冊子は、京都大学人文科学研究所共同研究プロジェクト：「情報処理技術は漢字文献からどのような情報を抽出できるか——人文情報学の基礎を築く」主催の公開シンポジウム「情報の構造とメタデータ」（2012年2月24日開催）の予稿集である。

シンポジウムについて

日時と場所

2012年2月24日（金）, 13:00-17:50
京都大学人文科学研究所本館101セミナー室

録画による記録

人文: USTREAM (<http://ustream.tv/channel/zinbun/>)

プログラム

第1部

趣旨説明

13:00-13:10

マンガにおける異本研究／安岡孝一（京都大学人文科学研究所）

13:10-13:50

TEIテキスト・モデルの今昔／Christian Wittern（京都大学人文科学研究所）

13:50-14:30

第2部

漢字文献における電子的翻刻の課題について／白須裕之

14:50-15:30

CiNiiのメタデータ・デザイン／大向一輝（国立情報学研究所）

15:30-16:30

第3部

パネルディスカッション／コメントと問題提起: 永崎研宣（人文情報学研究所）

16:50-17:50

マンガにおける異本研究

安岡孝一*

1 はじめに

日本のマンガの多くは、週刊誌や月刊誌等に連載されたのち、単行本としてまとめられる、という出版形態をとっている。しかも、単行本にまとめる際に、フキダシやコマ割り、あるいはストーリーそのものを改変することが、かなり頻繁におこなわれている。すなわち、初出と単行本とが、いわゆる異本の関係にあり、それらの異同を調査することが、マンガ研究の一部をなす、ということである。

本稿では、そのような異同を記述するにあたり、マンガの情報構造を、できるかぎり簡便に反映する方法を探る。具体的には、手塚治虫『三つ目がとおる グリーブの秘密編』の異本を例に、マンガにおける異同について実際に調査をおこない、その異同を記述する方法について考察する。

2 『三つ目がとおる グリーブの秘密編』

手塚治虫は、ほとんど全ての作品においてリライトをおこなっており、したがって、作品の数だけ異本が存在すると言っても過言ではない。1974年から1978年にかけて週刊少年マガジン(講談社)に連載された『三つ目がとおる』も例外ではなく、その中でも「グリーブの秘密編」は、広範囲なリライトが二度に渡っておこなわれた作品である。本稿では、あえて、この『三つ目がとおる グリーブの秘密編』を、異同調査の例として取り上げることにする。

2.1 改題

週刊少年マガジン 1975年6月1日号～8月24日号初出。

講談社コミックス第325巻[KC325]『三つ目がとおる ③』(講談社、1976年3月1日第1刷発行)で単行本化の際、8月17日号・8月24日号掲載分のプロットがカットされた。また、1月19/26日号～3月30日号掲載分の各コマを再構成して「第1章」が作られ、同時に「第2章」～「第5章」の章立てもおこなわれた。さらに、手塚治虫全集108巻[MT108]『三つ目がとおる ⑧』(講談社、1980年5月20日第1刷発行)に収録の際、8月17日号・8月24日号掲載分が復活し、同時に「グリーブの秘密編」というサブタイトルが付けられた。また、フキダシの総ルビは、ほぼ全て削除された。

なお、その後のKCスペシャル253巻[KCSP253]『三つ目がとおる 第三集』(講談社、1986年9月6日第1刷発行)や、あるいはそれ以後の単行本における「グリーブの秘密編」は、管見ではMT108がそのまま踏襲されており、異同はないように思われる。ただし、講談社プラチナコミック63巻[KPC63]『三つ目がとおる グリーブの秘密編』(講談社、2003年5月28日第1刷発行)では、MT108の冒頭pp.8-30が削除されており、残りの部分だけが収録されている。

*京都大学人文科学研究所附属東アジア人文情報学研究センター

手塚治虫の 新伝奇ロマン

★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆
★☆☆☆☆

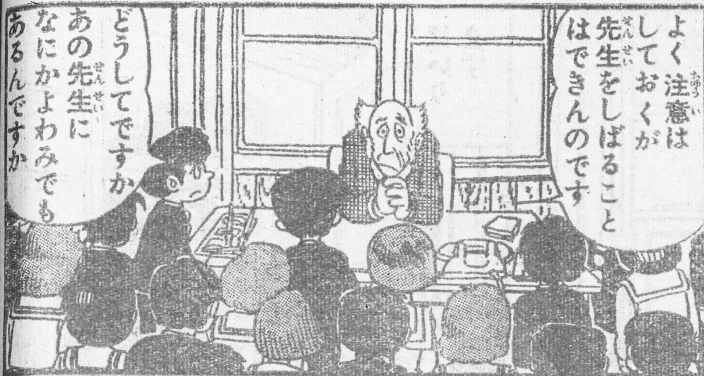
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★

十日後八時から、東京・渋谷公会堂でTBSテレビ「8時だヨ！全員集合」に出演。



図1: 週刊少年マガジン 1975年6月1日号での上底先生の正体

三つ目がとおる



どうしてですか
あの先生に
なにかよわみでも
あるんですか

よく注意は
しておくが
先生をしぼることは
できませんのです



まっつてくれ
いろいろ
事情が
あつてね



そんな
よく
学園マンガに
あるような
ものじゃ
ないんだ



いた
い
ところを
つくねえ

たとえば
うちの学校の
理事長の
しんせきとか



はつきり
いって
ください

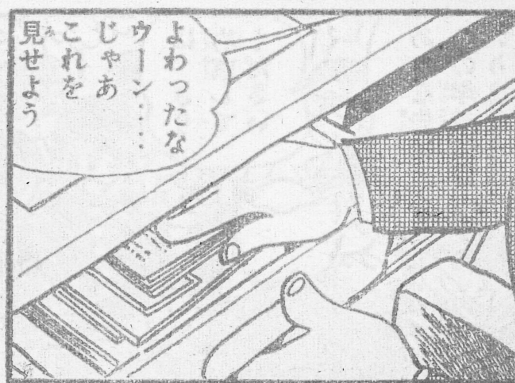
でないと
ぼくたち
ストをしますよ



ちがうたら
そういう
ものとは
関係ない



じゃあ
なんとか
組の
ボスの



よわったな
ウーン……
じゃあ
これを
見せよう

講談社コミックス

紅の挑戦者①⑥

猛烈キックがうなる大迫力の激闘漫画!

絶賛発売中! 各350円



図 3: MT108 での上底先生の正体

2.2 主な登場人物

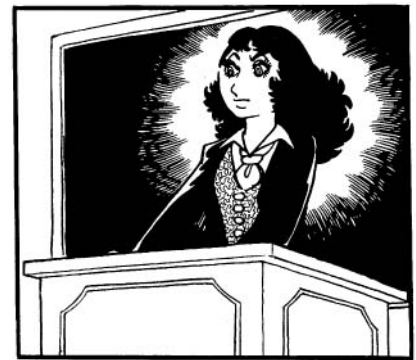
写楽保介 本作の主人公。古代三つ目族の生き残りで、第三の目を持つ。現代人類を遥かに超える高い知能を有し、脳ミソをトコロテンにする機械を作ったり、人類には解読不能な古代文字を読んだりする。第三の目をバンソウコで塞がれると、小学一年生なみの中学二年生。



和登千代子 本作のヒロイン。写楽の同級生でコイビト。お寺の娘で、住職の父親に、写楽との付き合いを禁じられている。写楽のバンソウコを、自由にはがしたり貼ったりする腕を持つ。上底先生によって、写楽と共に、ナバホ山ちかくの「グリーブ」へ連れて行かれる。



上底先生 写楽と和登の学園に赴任してきた女教師。初出では1975年5月25日号～8月10日号に登場。ICPOの捜査官であり、CIAのエージェントでもあるという設定(図1)。「グリーブ」の暴走に巻き込まれて死亡。KC325では設定が変わっており、「全ピキ連」の闘士ということになっている(図2)。また、KC325では死亡せず、写楽・和登と共に小型船で日本に帰る。一方、MT108でも「全ピキ連」の闘士という設定(図3)だが、最期は「グリーブ」の暴走に巻き込まれて行方不明となっている。



ブラック・ホーン 上底先生の夫で、ナバホ族の出身。「グリーブ」の研究者。初出では1975年7月6日号～8月17日号に登場、「グリーブ」の暴走で怪我を負い、マクドナルド・ハンパーカーに撃たれて死亡。KC325では「グリーブ」の暴走で負った怪我がもとで死亡。MT108ではポーク・ストロガノフに撃たれて死亡。



マクドナルド・ハンパーカー 初出では1975年8月10日号～8月24日号に登場。CIA部長で、写楽と和登をCIA本部に連れて行く。KC325では登場しない。MT108では登場しているものの、名前がポーク・ストロガノフに変わっている。



3 マンガにおける異同の記述

3.1 異同のレベル

一口に異同と言っても、マンガにおける異同には、様々な種類のものが存在する。本稿ではそれらの異同を、多少乱暴だが、以下の3レベルに分類することにする。

1. プロットの異同

マンガのプロット、すなわち話の筋そのものが変わっている場合。一般には、コマの異同やフキダシの異同を伴う。

2. コマの異同

コマの位置や、コマに書かれている内容が変わっている場合。

3. フキダシの異同

コマ中の人物や絵は変更されずに、フキダシの内容だけがかわっている場合。

図1(週刊少年マガジン 1975年6月1日号 pp.104-105)と図2(KC325 p.62)および図3(MT108 p.62)の異同を例に、少し考察してみよう。上底先生の横暴をやめさせるべく、生徒たちが校長に直談判するシーンなのだが、ここで明かされる上底先生の正体は、図1と図2・3では全く異なっている。すなわち、図1では、生徒たちに正体が明かされる(ICPOの捜査官)ものの、読者にはわからない。一方、図2・3では、「全ピキ連」の闘士という正体が明かされる。これが「プロットの異同」である。次に、図1と図2・3の各コマをチェックすると、図1のp.104の下半分にある5つのコマと、図1のp.105の上半分の大コマは、図2・3では使われていない。そして、図1の残りの7コマが、図2・3の1ページ(p.62)に詰め込まれている。これが「コマの異同」である。また、図2・3のうち下3段の5コマは、それぞれ図1に同一の絵柄のコマがあるものの、フキダシの内容が完全に変わっている。これが「フキダシの異同」である。

これらの異同のレベルのうち、「プロットの異同」については、自然言語で記述するしかなく、機械可読な記述は非常に難しいと考えられる。一方、「コマの異同」と「フキダシの異同」は、機械可読な記述が可能だと考えられるが、その際、著作権法に抵触しないような記述が必要である。端的には、コマの画像そのものを直接記述せずに、異同を記述する必要がある。そのような記述が可能かどうか以下で考察する。

3.2 各コマを特定するための記述

図1の各コマを仮に「1975.6.1-104A」～「1975.6.1-105D」、図2の各コマを仮に「KC325-62A」～「KC325-62G」、図3の各コマを仮に「MT108-62A」～「MT108-62G」と名づけた場合、これらのコマを特定するためには、どのような記述が必要かを考えてみる。

● 1975.6.1-104A

フキダシ右上=校長 “ま	コマ上= 1975.6.1-104天
まってくれ”	コマ右= 1975.6.1-104 右ハシラ
フキダシ左上=校長 “いろいろ	コマ左= 1975.6.1-104B
事情が	コマ下= 1975.6.1-104C
あってね”	

● KC325-62A

フキダシ右上=校長 “ま	コマ上= KC325-62天
まってくれ”	コマ右= KC325-62 右ハシラ
フキダシ左上=校長 “いろいろ	コマ左= KC325-62B
事情が	コマ下= KC325-62C
あってね”	

● MT108-62A

フキダシ右上＝校長 “ま
待ってくれ”
フキダシ左上＝校長 “いろいろ
事情が
あってね”

コマ上＝ MT108-62 天
コマ右＝ MT108-62 右ハシラ
コマ左＝ MT108-62B
コマ下＝ MT108-62C

● 1975.6.1-104B

フキダシ右上＝校長 “よく注意は
しておくが
先生をしばること
はできんです”
フキダシ左上＝男子生徒 “どうしてですか
あの先生に
なにかよわみでも
あるんですか”

コマ上＝ 1975.6.1-104 天
コマ右＝ 1975.6.1-104A
コマ左＝ 1975.6.1-104 左ノド
コマ下＝ 1975.6.1-104C
コマ下＝ 1975.6.1-104D

● KC325-62B

フキダシ右上＝校長 “よく注意は
しておくが
先生をしばること
はできんです”
フキダシ左上＝男子生徒 “どうしてですか
あの先生に
なにかよわみでも
あるんですか”

コマ上＝ KC325-62 天
コマ右＝ KC325-62A
コマ左＝ KC325-62 左ノド
コマ下＝ KC325-62C
コマ下＝ KC325-62D

● MT108-62B

フキダシ右上＝校長 “よく注意は
しておくが
先生をしばること
はできんです”
フキダシ左上＝男子生徒 “どうしてですか
あの先生に
何か弱みでも
あるんですか”

コマ上＝ MT108-62 天
コマ右＝ MT108-62A
コマ左＝ MT108-62 左ノド
コマ下＝ MT108-62C
コマ下＝ MT108-62D

● 1975.6.1-104C

フキダシ右上＝校長 “ いたい
ところを
つくねえ ”

フキダシ左上＝男子生徒 “ たとえば
うちの学校の
理事長の
しんせきとか
.... ”

コマ上＝ 1975.6.1-104A
コマ上＝ 1975.6.1-104B
コマ右＝ 1975.6.1-104 右ハシラ
コマ左＝ 1975.6.1-104D
コマ下＝ 1975.6.1-104E
コマ下＝ 1975.6.1-104F

● KC325-62C

フキダシ右上＝校長 “ あの人は
全ピキ連
なのです ”

フキダシ左上＝男子生徒 “ なんですか
その全ピキ
連って.... ”

コマ上＝ KC325-62A
コマ上＝ KC325-62B
コマ右＝ KC325-62 右ハシラ
コマ左＝ KC325-62D
コマ下＝ KC325-62E
コマ下＝ KC325-62F

● MT108-62C

フキダシ右上＝校長 “ あの人は
全ピキ連
なのです ”

フキダシ左上＝男子生徒 “ なんですか
その全ピキ
連って.... ”

コマ上＝ MT108-62A
コマ上＝ MT108-62B
コマ右＝ MT108-62 右ハシラ
コマ左＝ MT108-62D
コマ下＝ MT108-62E
コマ下＝ MT108-62F

● 1975.6.1-104D

フキダシ右上＝校長 “ そんな
よく
学園マンガに
あるような
ものじゃ
ないんだ ”

コマ上＝ 1975.6.1-104B
コマ右＝ 1975.6.1-104C
コマ左＝ 1975.6.1-104 左ノド
コマ下＝ 1975.6.1-104F
コマ下＝ 1975.6.1-104G

● KC325-62D

フキダシ右上＝校長 “ 「全女性
ピンカラ
キリマデ
連盟」
というので
す！ ”

コマ上＝ KC325-62B
コマ右＝ KC325-62C
コマ左＝ KC325-62 左ノド
コマ下＝ KC325-62F

● MT108-62D

フキダシ右上＝校長 “ 「全女性
ピンカラ
キリマデ
連盟」
というので
す！ ”

コマ上＝ MT108-62B
コマ右＝ MT108-62C
コマ左＝ MT108-62 左ノド
コマ下＝ MT108-62F

● 1975.6.1-104E

フキダシ右上＝男子生徒 “ じゃあ
なんとか
組の
ボスの
…… ”

コマ上＝ 1975.6.1-104C
コマ右＝ 1975.6.1-104 右ハシラ
コマ左＝ 1975.6.1-104F
コマ下＝ 1975.6.1-104H

● 1975.6.1-104F

フキダシ右上＝校長 “ ちがうったら
そういう
ものとは
関係ない ”

コマ上＝ 1975.6.1-104C
コマ上＝ 1975.6.1-104D
コマ右＝ 1975.6.1-104E
コマ左＝ 1975.6.1-104G
コマ下＝ 1975.6.1-104H
コマ下＝ 1975.6.1-104I

● 1975.6.1-104G

フキダシ右上＝男子生徒 “ はっきり
いって
ください ”
フキダシ左上＝男子生徒 “ でないと
ぼくたち
ストをしますよ ”

コマ上＝ 1975.6.1-104D
コマ右＝ 1975.6.1-104F
コマ左＝ 1975.6.1-104 左ノド
コマ下＝ 1975.6.1-104I

● 1975.6.1-104H

フキダシ右上＝校長 “ よわったな
ウン……
じゃあ
これを
見せよう ”

コマ上＝ 1975.6.1-104E
コマ上＝ 1975.6.1-104F
コマ右＝ 1975.6.1-104 右ハシラ
コマ左＝ 1975.6.1-104I
コマ下＝ 1975.6.1-104 地

● 1975.6.1-104I

コマ上＝ 1975.6.1-104F
コマ上＝ 1975.6.1-104G
コマ右＝ 1975.6.1-104H
コマ左＝ 1975.6.1-104 左ノド
コマ下＝ 1975.6.1-104 地

● 1975.6.1-105A

フキダシ左上＝“ あっ ”

フキダシ左＝男子生徒 “ あっ ”

コマ上＝ 1975.6.1-105 天

コマ右＝ 1975.6.1-105 右ノド

コマ左＝ 1975.6.1-105 左ハシラ

コマ下＝ 1975.6.1-105B

コマ下＝ 1975.6.1-105C

● 1975.6.1-105B

フキダシ左上＝校長 “ …… こういうわけだ

事情^{じじょう}はいえんが

あの先生^{せんせい}は この学校^{がっこう}では

すぎにふるまえるんだ…… ”

コマ上＝ 1975.6.1-105A

コマ右＝ 1975.6.1-105 右ノド

コマ左＝ 1975.6.1-105C

コマ左＝ 1975.6.1-105D

コマ下＝ 1975.6.1-105 地

● KC325-62E

フキダシ左上＝校長 “ あの^{ひと}人をおこらせたり

すると 全世界^{ぜんせかい}の

五百万人^{まんにん}の女^{おんな}が政府^{せいふ}へ

どなりこんで この学校^{がっこう}を

ぶつつぶしてしまいます！ ”

コマ上＝ KC325-62C

コマ右＝ KC325-62 右ハシラ

コマ左＝ KC325-62F

コマ左＝ KC325-62G

コマ下＝ KC325-62 地

● MT108-62E

フキダシ左上＝校長 “ あの^{ひと}人をおこらせたり

すると 全世界^{ぜんせかい}の

五百万人^{まんにん}の女^{おんな}が政府^{せいふ}へ

どなりこんで この学校^{がっこう}を

ぶつつぶしてしまいます！ ”

コマ上＝ MT108-62C

コマ右＝ MT108-62 右ハシラ

コマ左＝ MT108-62F

コマ左＝ MT108-62G

コマ下＝ MT108-62 地

● 1975.6.1-105C

フキダシ右上＝女子生徒 “ でも

あんまり

だわ ”

フキダシ左上＝男子生徒 “ よりによって

この学校^{がっこう}へ

くるなんて

ひどいや ”

コマ上＝ 1975.6.1-105A

コマ右＝ 1975.6.1-105B

コマ左＝ 1975.6.1-105 左ハシラ

コマ下＝ 1975.6.1-105D

● KC325-62F

フキダシ右上＝女子生徒 “ じゃあ

しかたが

ないわね ”

フキダシ左上＝男子生徒 “ ^{おんな}女^{おとこ}ってのは

いま男^{おとこ}の

十倍^{じゅうばい}も

つよいって

からなア ”

コマ上＝ KC325-62C

コマ上＝ KC325-62D

コマ右＝ KC325-62E

コマ左＝ KC325-62 左ノド

コマ下＝ KC325-62G

● MT108-62F

フキダシ右上=女子生徒“じゃあ
しかたが
ないわねえ”
フキダシ左上=男子生徒“女ってのは
いま男の
十倍も
強いつて
からなア”

コマ上= MT108-62C
コマ上= MT108-62D
コマ右= MT108-62E
コマ左= MT108-62 左ノド
コマ下= MT108-62G

● 1975.6.1-105D

フキダシ右上=和登千代子“あたし
学校を
やめたく
なったわ”

コマ上= 1975.6.1-105C
コマ右= 1975.6.1-105B
コマ左= 1975.6.1-105 左ハシラ
コマ下= 1975.6.1-105 地

● KC325-62G

フキダシ右上=和登千代子“ボク
十倍も
つよくない
よ 一倍半
ぐらいよ”

コマ上= KC325-62F
コマ右= KC325-62E
コマ左= KC325-62 左ノド
コマ下= KC325-62 地

● MT108-62G

フキダシ右上=和登千代子“ボク
十倍も
強くないよ
一倍半
ぐらいよ”

コマ上= MT108-62F
コマ右= MT108-62E
コマ左= MT108-62 左ノド
コマ下= MT108-62 地

各コマのフキダシと相対位置を記述しただけのものだが、各コマを特定するという点では、これにト書きを加えればほぼ十分だと考えられる。仮に、複数ページに渡るコマであっても、「コマ上」や「コマ下」が複数ページに渡るというだけで、記述には困らない。あるいは、矩形でないコマであっても、上下左右には何がしか存在しているはずなので、それを記述すればよい。極端な話、コマと呼べるものが存在しない場合でも、ページを区分している単位があるのなら、それを順不同に記述すればいいだろう。

ただし、上の例では、フキダシの話者を記述に含めたが、話者は無くてもよいような気がする。特に、1975.6.1-104Hのフキダシの話者は、ここでは「校長」としているが、それは前後関係が無ければ特定できない。あるいは、絵柄上は同一人物であっても、初出では「マクドナルド・ハンパーガー」で、MT108では「ポーク・ストロガノフ」になってしまうという問題もある。そういうややこしさを考えると、あるいは話者を記述しない方が良いのかもしれない。

また、上の例では「コマ左」と「コマ下」を記述したが、通常の矩形のコマだけの場合は、「コマ上」と「コマ右」だけで十分である。ただし、タチキリ(コマの端がページの端より外

にある)の場合には、たとえば「コマ下= bbb-pp タチキリ」のように記述すべきだと考えられることから、とりあえずは「コマ上」「コマ右」「コマ左」「コマ下」を全て記述しておくことにした。

3.3 機械可読な異同の記述

各コマの記述をもとに、「コマの異同」と「フキダシの異同」をどのように記述するのだが、これは端的に言えば、異同を記述するのではなく、コマの絵柄上の同一性を記述すればよい。すなわち、前節の例で言えば

1975.6.1-104A \cong KC325-62A \cong MT108-62A
1975.6.1-104B \cong KC325-62B \cong MT108-62B
1975.6.1-104C \cong KC325-62C \cong MT108-62C
1975.6.1-104D \cong KC325-62D \cong MT108-62D
1975.6.1-105B \cong KC325-62E \cong MT108-62E
1975.6.1-105C \cong KC325-62F \cong MT108-62F
1975.6.1-105D \cong KC325-62G \cong MT108-62G

という記述が、それにあたる。これにより、同一だと認められなかったコマ、すなわち 1975.6.1-104E、1975.6.1-104F、1975.6.1-104G、1975.6.1-104H、1975.6.1-104I、1975.6.1-105A に「コマの異同」があることがわかる。

これに加え、1975.6.1-104C の下にあるコマと、KC325-62C の下にあるコマは、同一ではない。1975.6.1-104D の下にあるコマと、KC325-62D の下にあるコマも、同一ではない。また、1975.6.1-105B の上にあるコマと、KC325-62E の上にあるコマは、同一ではない。1975.6.1-105C の上にあるコマと、KC325-62F の上にあるコマも、同一ではない。したがって、KC325-62C と KC325-62D の下、KC325-62E と KC325-62F の上にある「隙間」に、「コマの異同」が潜んでいることが理解できる。MT108 についても同様である。

一方、絵柄上の同一性が認められるコマどうしであっても、前節の記述におけるフキダシに相違があれば「フキダシの異同」があることがわかる。たとえば、1975.6.1-104C と KC325-62C は同一のコマだが、フキダシの記述は全く異なっており、「フキダシの異同」があるといえる。1975.6.1-104C と MT108-62C との間にも「フキダシの異同」がある。ただ、KC325-62C と MT108-62C との間は、ルビの有無だけの違いであり、これを「フキダシの異同」と呼ぶべきかどうか悩ましい。さらにヤヤコシイのが KC324-62F と MT108-62F の差であり、かな漢字の変更や語尾のちょっとした変化を「フキダシの異同」とするかどうかは、かなり難しいところである。なお、これらの細かい差異を全て「フキダシの異同」とみなした際には、フキダシも含めて全く同一のコマは、以下のように記述できる。

1975.6.1-104A \equiv KC325-62A
1975.6.1-104B \equiv KC325-62B

3.4 コマの絵柄上の同一性

前節では「コマの絵柄上の同一性」というものが、ア priori に与えられるかのような説明をおこなったが、この点に関してはいくつか考慮すべき事項がある。

多色刷の場合

初出で四色刷あるいは二色刷であったにもかかわらず、単行本で一色刷となっている場合は、かなりある。これらのコマに関しては、あえて同一だとみなして「≒」で繋いでおくべきだろう。

コマに加筆がおこなわれている場合

初出になかった人物や事物あるいは書き文字が、コマの中に書き加えられている場合。あるいは、スクリーン・トーンやベタが直されている場合。これらのコマを同一だとみなすことはできないが、何がしかの関係づけをおこないたいのも事実である。とりあえず、これらのコマについては、たとえば「〜」のような関係演算子でコマどうしを繋いでおく、というのも一案だろう。

コマに変形が加えられている場合

コマの形が変えられている場合。あるいはコマを回転したり、鏡像を用いた場合。このようなコマについても、とりあえず「≈」のような関係演算子でコマどうしを繋いでおく、というのも一案だろう。

タチキリの変更による場合

タチキリのコマは、週刊誌と単行本で余白部分の取り方がかなり異なるため、コマの端が変わってしまったり、あるいはタチキリでなくなったりする場合もある。程度にもよるが、変更が小さい場合はあえて「≒」で、大きい場合は「≈」で、コマどうしを繋いでおくのも一案だろう。

4 おわりに

マンガにおける「コマの異同」と「フキダシの異同」を、記述する方法について考察した。具体的には、各コマの相対位置記述とフキダシの記述をおこない、さらに「コマの絵柄の同一性」を記述することによって、「コマの異同」と「フキダシの異同」の双方を、同時に記述できることを示した。

また、この手法で『三つ目がとおる グリーブの秘密編』の各コマの記述をおこない、週刊少年マガジン(初出)と講談社コミックス [KC325] と手塚治虫全集 [MT108] との間の異同を、実際に確認した。異同を確認した結果、コマの対応関係はかなり複雑なものとなってしまったため、読みやすいようページ毎の対応に直して、次ページ以降に示す。

週刊少年マガジン		KC325	MT108
第1章 先生も狂った!			
1975.1.19/26号	pp.10-17	pp.8-15	pp.8-15
	p.18 と p.24	p.16	p.16
	p.25	p.17	p.17
1975.2.23号	pp.112-116	pp.18-22	pp.18-22
1975.1.19/26号	pp.26-27	pp.23-24	pp.23-24
1975.2.2/9号	p.59	p.25	p.25
	p.60 と p.61	p.26	p.26
	p.63	p.27	p.27
	-	pp.28-30	pp.28-30
1975.3.23号	pp.123-124	pp.31-32	pp.31-32
	p.126 と p.129 と p.128	p.33	p.33
	p.130	p.34	p.34
1975.3.30号	pp.90-96	pp.35-41	pp.35-41
1975.6.8号	pp.78-84	pp.42-48	pp.42-48
第2章 悪魔のような女			
	-	p.50	p.50
1975.6.1号	pp.92-93	pp.51-52	pp.51-52
	p.94	p.53	p.53
	p.95		
	p.96	p.54	p.54
1975.6.15号	p.127 と p.128	p.55	p.55
	p.129	p.56	p.56
1975.6.1号	p.97		
	pp.98-99	pp.57-58	pp.57-58
	p.100		-
	pp.101-103	pp.59-61	pp.59-61
	p.104 と p.105	p.62	p.62
	pp.106-110		-
1975.6.8号	p.87-89	pp.63-65	pp.63-65
	p.90		-
	p.91	p.66	p.66
	p.92	p.67	p.67
1975.6.15号	p.121	p.68	p.68
	p.122		
	pp.123-127	pp.69-73	pp.69-73
	pp.131-134	pp.74-77	pp.74-77

週刊少年マガジン		KC325	MT108
1975.6.22号	p.124	-	
	p.125とp.126	p.78	p.78
	pp.127-128	pp.79-80	pp.79-80
	p.129	p.81	p.81
	p.130	p.82	p.82
	p.131		
	p.132	p.83	p.83
	p.133	p.84	p.84
	1975.6.8号		
p.85とp.86			
	p.96	p.85	p.85
1975.6.15号	p.117	p.86	p.86
1975.6.8号	p.94	p.87	p.87
	p.95	p.88	p.88
1975.6.15号	p.118		
	p.119	p.89	p.89
1975.6.8号	p.95		
1975.6.15号	p.129	p.90	p.90
1975.6.22号	p.133		
第3章 死の谷			
1975.6.22号	pp.134-135	pp.92-93	pp.92-93
	p.136	-	
	p.137	p.94	p.94
	p.138	p.95	p.95
		p.96	p.96
	p.139	p.97	p.97
	pp.140-142	pp.98-100	pp.98-100
1975.6.29号	p.27	p.101	p.101
	pp.28-29	-	
	pp.30-31	pp.102-103	pp.102-103
	p.32	-	
	pp.33-41	pp.104-112	pp.104-112
	p.43	p.113	p.113
	p.44とp.45	p.114	p.114
	pp.46-48	pp.115-117	pp.115-117
	pp.49-50	-	
	pp.51-52	pp.118-119	pp.118-119

週刊少年マガジン		KC325	MT108
1975.7.6号	p.92	p.120	p.120
	p.93	p.121	p.121
	p.94		
	pp.95-96	pp.122-123	pp.122-123
	p.94	p.124	p.124
	p.97	p.125	p.125
	pp.98-110		
第4章 ナバホ・ポイント			
1975.7.13号	p.94	p.140	p.140
	p.95	p.141	p.141
	p.96	p.142	p.142
	p.98		
	pp.99-105	pp.143-149	pp.143-149
	pp.106-108	-	
	pp.109-112	pp.150-153	pp.150-153
1975.7.20号	p.64	p.154	p.154
	p.65 と p.66	p.155	p.155
	p.67 と p.69	p.156	p.156
	p.70	p.157	p.157
	p.71	p.158	p.158
	p.77		
	pp.72-76	-	
1975.8.10号	pp.41-44	pp.159-162	
第5章 満月の奇蹟			
1975.7.20号	pp.78-82	pp.160-164	pp.164-168
1975.7.27号	pp.44-46	-	
	pp.47-54	pp.165-172	pp.169-176
	pp.57-63	pp.173-179	pp.177-183
1975.8.3号	pp.64-65	-	
	pp.66-79	pp.180-193	pp.184-197
	p.80	-	
	pp.81-82	pp.194-195	pp.198-199
1975.8.10号	p.83	-	
	pp.38-40	pp.196-198	pp.200-202
1975.8.10号	p.46	p.199	p.203
	pp.55-56	pp.200-201	pp.204-205

週刊少年マガジン		KC325	MT108
1975.8.10号	pp.47-48	pp.202-203	pp.206-207
	-	pp.204-205	pp.208-209
1975.8.10号	pp.49-51	pp.206-208	pp.210-212
	p.52 と p.53	p.209	p.213
	pp.54-56	-	-
	-	pp.210-212	pp.214-216
	-	pp.213-215	-
1975.8.17号	pp.108-112	-	pp.217-221
	p.113 と p.115		p.222
	pp.116-117		pp.223-224
	pp.119-120		pp.225-226
	p.121		p.227
	p.123		
			p.228
	p.124		
	pp.125-126		pp.229-230
1975.8.24号	pp.84-101		pp.231-248
	-	p.216	p.249
1975.8.24号	p.102	-	p.250
	-	pp.217-218	-
	-	p.219	p.251

ただし、本稿の手法はあくまで、作業者が実際に目で確認した「コマの絵柄の同一性」を、記述していくための手助けとなるに過ぎない。「コマの絵柄の同一性」を自動抽出するような画像ツールと組み合わせる、というのが一案なのだが、それを著作権法に違反せずにおこなえるのか、たとえおこなえたとして、それを公に発表することが可能なのか、今後その点が問題となるだろう。

なお、『三つ目がとおる』の異同に関しては、

http://ngt-the-knife.sakura.ne.jp/rep_mitusme.shtml

が非常に詳しく、本稿の執筆においても助けていただいた。作者の ngt 氏に謝意を述べる次第である。

TEI | テキスト・モデルの今昔

クリスティアン・ウィッテッルン(京都大学)
cwittern@gmail.com

要旨: TEI(Text Encoding Initiative)のテキスト・モデルは 1987 年以来幾多の変遷を経て今日の姿に変わった。TEI 本来のテキスト・モデルは、SGML の基本モデルに沿い、複数のツリーから成る構造を持っていたが、2002 年から 2007 年にかけて行った XML への移行の過程で、複数のツリーを可能にする構造が失われたので、テキスト・モデルもマークアップ言語の制約を受けて、基本的にはツリーを 1 つしか持たない構造になった。よって、マークアップのシンタックス・レイヤーではもはや複数の構造に対応できないので、その代替措置として、20 年以上に及ぶ運用の経験も踏まえ、セマンティクス・レイヤーでその構造を補う対策をとってきた。さらに、2011 年には、物理的な文献あるいはテキストの生成の表記のために、もう 1 つ、レイヤーを導入した。本発表では、元々のモデルの概要とそれ以来の変更、及びその変更の理由について報告する。

キーワード：テキスト・エンコーディング, テキスト・モデル, SGML, XML

1. SGML について

1.1 SGML の基本構造

SGML¹ (スタンダード ジェネラライズド マークアップ ランゲージ) は 二つの基本概念に基づいて定義された：

- マークアップは、記述的で有るべき：それは文書の構造と他の属性を記述することで、それに対して実行する処理を指定することでは必要ない。記述的マークアップは、より良い予測できない将来の処理のニーズと技術を見込んである。
- マークアップは、プログラムやデータベースのように厳密に定義されたオブジェクトを処理するために利用できる技術は、同様の文書を処理するために使用することができるように厳格なはずである。

SGML は 1986 年に ISO 8879 として国際標準になった。

¹ SGML の総合的な手引き書はゴールドファーブ(Goldfarb 1990)であるが、日本語ではブラヤンの『SGML 入門』(Bryan 1991)が便利。

SGMLには文書定義(DTD、ここではスキーマという)と文書そのもの(インスタンス)が別に行う、定義された要素(element)以外は使用不可能。さらに、その要素の構造も定義されている。例1では詩の構造が定義されて(行1から)、その後(行8から)実際のテキストが挙げられている。

```

1 <!DOCTYPE 七言絶句 [
2 <!ELEMENT 七言絶句 (作者, 題名, 絶句) >
3 <!ELEMENT 作者 (#PCDATA) >
4 <!ELEMENT 題名 (#PCDATA) >
5 <!ELEMENT 絶句 (句+) >
6 <!ELEMENT 句 (#PCDATA) >
7 ]>
8 <絶句>
9     <作者>張繼</作者>
10    <題名>楓橋夜泊</題名>
11    <本文>
12        <句>月落烏啼霜滿天, </句>
13        <句>江楓漁火對愁眠。 </句>
14        <句>姑蘇城外寒山寺, </句>
15        <句>夜半鐘聲到客船。 </句>
16    </本文>
17 </絶句>

```

例1：SGML文書定義と文書

ただ、実際の文書は複数の構造が持ち得る、その為にSGMLでも構造が複数で可能にする。例2は詩選集(anthology)を例して、その内の一つの詩を挙げている。しかし、詩(poem)の構造と別に、ページ(page)の構造もマークされている。詩の第一節始め行8から15となっているが、その間ページが変わるので、ページの構造は行3から11、と行12から23となっている、<(anthology)stanza>の要素と<p.anth>の要素はお互いに入れ子構造になってないので、XMLでは不可能な構造にはなっているが、SGMLではこのような複数の構造が可能。

```

1 <(anthology)anthology>
2 <(p.anth)p.anth>
3 <(p.anth)page>
4
5 <!--      other titles and lines on this page here  -->
6
7 <(anthology)poem><title>The SICK ROSE
8 <(anthology)stanza>
9     <line>O Rose thou art sick.
10    <line>The invisible worm,
11 </(p.anth)page>
12 <(p.anth)page>
13     <line>That flies in the night
14     <line>In the howling storm:
15 <(anthology)stanza>

```

```

16      <line>Has found out thy bed
17      <line>Of crimson joy:
18      <line>And his dark secret love
19      <line>Does thy life destroy.
20 </ (anthology)poem>
21
22 <!--      rest of material on this page here      -->
23 </ (p.anth)page>
24
25 </ (p.anth)p.anth)
26 </ (anthology)anthology>

```

例 2：複数の構造を持つ SGML 文書 (TEI P3 (1994), p. 25 より)

SGML は機能が満載されていたことにより、そのままでは全てを実装することは困難であった。また、要素の省略などが原因で、パーサのアルゴリズムが比較的複雑になることも難点だった。そこで後に、SGML を簡略化および「改良」した形の XML²が開発され、普及してゆくことになった。

1.2 順序を持つコンテンツオブジェクトの階層としてのテキスト

テキストは、切れ目のない語の連鎖ではない。ましてや切れ目のないバイトの連鎖でもない。異なる目的に応じて、テキストはそれぞれ異なる単位、種類、大きさに分けられる。この文書のような散文のテキストであれば、節、章、段落、文に分けることができる。韻文であれば、篇、スタンザ、行に分けられる。印刷されたものであれば、散文や韻文の連鎖は、巻、集、ページという単位に分けられる。こうした文書の階層的な構造は SGML、XML などのマークアップ言語の基本モデルである、記述的マークアップの本流となった。

記述的マークアップとそれを伴う方法論はあまりにも成功していたため、ある研究者はこの方法はテキスト処理についての便利な手法だけではなく、やはり根本的に《正しい》方法であると信じていた：「記述的なマークアップは一番合理的な方法だけではなく、想像できる以上に一番優れた方法である」(Coombs et. al, 1987)。この見解は記述的なマークアップの基にあるモデルだけが正しく「テキストの正しい理解」(DeRose et. al., 1990)を反映させる。

このモデルではテキストは、章、節、段落、文などといった階層を持つ論理的な構造によって決定している。ページ、列、行、フォントの変更、空白などといった物理的な構造のことではない。

この見解では、テキストは《順序を持つコンテンツオブジェクトの階層》

2 XML の入門書は沢山あるが、文系研究者には《入門 XML》(日本語版 Ray 2004)が勧められる。

('ordered hierarchy of content objects', OHCO)である。記述的マークアップは、この階層を明白に記述することで功を奏する。OHCOの見解はテキストエンコーディングの強力なモデルとなり、実在するテキストの多くの特徴の合理的な処理につながる(例えば、章の見出しは必ず章の始めにあり、途中にはない。詩の句は詩の中にあるから句となる、等)。

しかし、適用できない場合もある。例えば、詩の一文は必ず一句の中に入らない。句も文に入るとは限らない。つまり、文と句は違う階層に属する。同様に、小説などでは語り手が直接話法で割り込む場合もある。それにもかかわらず、OHCOはテキスト処理に於いて他のモデルより優れた処理を可能にするので(例えば、テキストは単なる文字列であるなどの見方に比べると)SGML、XML等のマークアップ言語の支配的な考えになり、テキスト・エンコーディングに広く適用されている。

2. TEIの文書構造

2.1 複数の階層を持つ文書の記述

TEIは当初SGMLに基づいて定義されました。P1³からP3まではSGMLのみが、P4ではSGMLとXMLは両方使用可能となり、P5以降はSGMLのサポートがなくなった。その結果はTEIがP1からP3までは、上の例のような、複数の階層を持つ文書は可能であったがXMLへの以降に伴い、P4からは不可能となった。ただ、理論上では可能であっても、こうした構造がかなり複雑なので、実際のSGML文書処理システムにはこの機能は実装されなかった。テキストの物理的な構造を示すためのTEI要素、つまり<page>と<line>もこの時点で削除された。理論的に可能であったが、実際に使用不可能になっていたので、以下の考察ではこの部分を省略する。

2.2 TEI P1 - P4の基本文書構造

TEI文書は基本的に<teiHeader>と<text>の二つの要素からなる。以下の例3⁴

3 TEIのバージョンは'P'の頭文字+バージョン番号で定義されている。P1(1990年7月)からP4(2002年)までのバージョン番号は整数であるが、P5以降は版管理のためマイナー番号付け加えられた、P5 1.0.0(2007年)のようになった。Pは草案の段階で付けられてた、'プロポーザル'の頭文字だが、1994年に'P3'で初めての決定版が公開された時点ではバージョンの名称は既に固定されたので、'P'はそのまま使い続けられた。本稿は2011年12月公開されたP5 2.0を対象とする。

4 この例では現在用いられているXMLシンタックスを採用する。厳密でいえば、TEI P2からTEI P4まではルート要素はTEIでは無く、TEI.2となった。また、XMLの名前空間

では TEI 文書の一番簡単な例の一つを挙げている。<teiHeader>のなかは三つの要素が必ず必要：<title>(行 6)、つまりこの電子テキストのタイトル、<publicationStmt>(行 9)、公開条件とライセンスの表記、と<sourceDesc>(行 12)、ここはソースになる文献の書誌情報などを書く。<text>の中身はそれに比べるとかなり自由ですが、ここは<body>(行 17)に一個の<p>のみ入れてある。

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>題名</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>公開条件</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>ソースに関する情報</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <p>テキストはここから...</p>
19    </body>
20  </text>
21 </TEI>
```

例 3：最低条件を果たす妥当な TEI 文書

簡単にいえば、<teiHeader>にはこのファイルに関するメタデータ、<text>のなかには電子化の対象となる文書の本文が入る。

その本文の内容を記述する為に<text>のなかに使用可能な、TEI に定義された要素は 500 個位がある。TEI の勧告では文書要素を「構造的」(structural)か「流動的」(floating)かに分類される。構造的要素は文書の中で現れる場所が制限される。例えば、<head>要素(=見出し)は、<p>の中には現れない。流動的要素は名前の通り、より制限を受けずに、ほぼ文章中何処でも現れる。例えば <note>や <date>要素などがそうである。

以上の二つの分類の中間的なもので、<list>や <cit>要素といった固有の構造を持っているが流動的な特徴を持つものを「クリスタル」(crystal)と呼んでいる。

テキストの本文は (<p>~ </p>でタグ付けされた) 一連の段落の集合、または章、節、小節などがある。後者の場合では、<body>要素は一般的に知ら

の指定は TEI P5 で導入された。

れている<div>要素で分けられる。与えられた文書の最も大きな集合には、<div>でタグ付けする。その際、<div>要素には「type」属性にどのような区分であるのかを明示し、ネストになったすべての区分に対して<div>要素でマークアップしてもかまわない。散文的な文章などは段落要素である<p>でさらに分割される可能性がある。韻文などに対して、行は<l>でタグ付けされ、韻文のまとまりには、<lg>がある。

先に述べたように、TEI 勧告は、多様な流動的特徴に対して名称や定義を提案している。例えば、タイトルやキャプション（一般的に、これらは特定の構造要素に結び付くので、流動的要素には適さない）に対する<head>要素、引用文や会話文には<q>、リストには<list>要素と<list>の中の項目に対する<item>要素、注釈などには<note>要素、文書作成者のテキストへの編集者の修正に対する<corr>要素、他にも、語彙的に扱いにくいものを省略形で書くための<abbr>要素、数字に<num>、名称に<name>、日付に<date>、文献的引用などに<cit>、住所に<address>、文書の言語と異なる単語や熟語に<foreign>といった流動的要素がある。

2.2.1 《道蔵輯要》の実例

道教経典の叢集である《道蔵輯要》⁵から一つの実際を見ましょう。図1は《重刊道蔵輯要》のページ5353の画像(ここではデジタル・ファクシミリと云う)である。同じページの一部に相当する《道蔵輯要》のデジタル・ファクシミリは図2に挙げている。

5 《道蔵輯要》と《道蔵輯要計画》については Esposito(2011)と Wittern(2011)を参照。

重刊道藏輯要

金華宗旨

李佑上帝純陽呂祖天師著

天心第一

呂祖曰自然曰道道無名相一性而已一元神而已性命不可見寄之天光天光不可見寄之兩目古來仙真皆口口相傳傳一得一自太上化現東華遞傳巖以及南北兩宗全真可為極盛盛者盛其徒眾衰者衰於心傳以至今日泛濫極矣凌替極矣極則返故昔日有許祖垂慈普度特立教外別傳之旨聞者千劫難逢受者一時法會皆當仰體許祖心先於人倫日用間

道藏輯要

金華宗旨

室集二

立定脚跟方可修真悟性我奉勅為度師今以太乙金華宗旨發明然後細為開說太乙者無上之謂丹訣甚多總假有為而臻無為非一超直入之旨我傳宗旨直提性功不落第二法門所以為妙金華即光也光是何色取象於金華亦祕一光字在內是天仙太乙之真炁水鄉鉛只一味者此也回光之功全用逆法方寸中具有鬱羅蕭臺之勝玉京丹闕之奇乃至虛至靈之神所注儒曰虛中釋曰靈臺道曰祖土曰黃庭曰玄關曰先天竅蓋天心猶宅舍一般光乃主人翁也故一回光則周身之氣皆上朝諸子只去回光便是無上妙諦光易動而難定回之既久此光凝結即是自然法身而凝神於九霄之上矣心印經

所謂默朝飛昇者此也

金華即金丹神明變化各師於心

元神識神第二

呂帝曰天地視人如蜉蝣大道視天地亦泡影惟元神真性則超元會而上之其精氣則隨天地而敗壞矣然有元神在即無極也生天生地皆由此矣學人但能護元神則超生在陰陽外不在三界中此見性方可所謂本來面目是也最妙者光已凝結為法身漸漸靈通欲動矣此千古不傳之秘也

識心如強藩悍將遙執紀綱久之太阿倒置矣今凝守元宮回光返照如英主在上大臣輔弼內政既肅自然強悍悞伏矣

道藏輯要

金華宗旨

室集二

丹道以精水神火意土三者為無上之訣精水云何乃先天真一之炁神火即光也意土即中宮天心也以神火為用意土為體精水為基凡人以意生身身不止七尺者為身也蓋身中有魄為魄附識而用識依魄而生魄陰也識之體也識不斷則生生世世魄之變形易質無已也惟有魂神之所藏也魂書寓于目夜舍于肝寓目而視舍肝而夢者神遊也九天九地剎那歷遍覺則冥冥焉淵淵焉拘于形也即拘於魄也故回光所以煉魂即所以保神即所以制魄即所以斷識古人出世法煉盡陰滓以返純乾不過消魄全魂耳回光者消陰制魄之訣也無返乾之功止有回光之訣光即乾也回之即返之也只守此法

圖 1 : 《重刊道藏輯要》p 5353

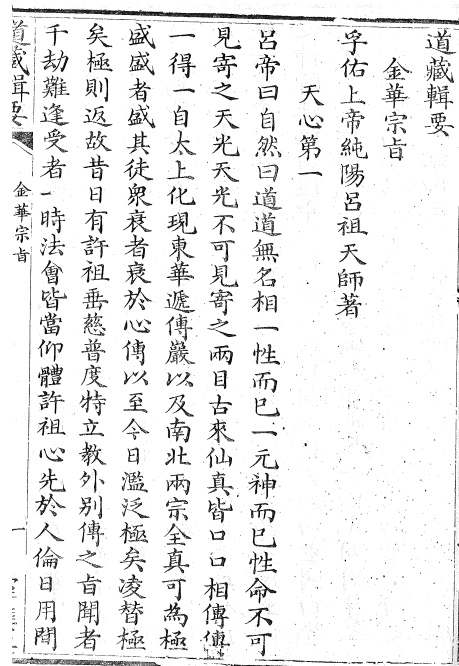


図2: 《道藏輯要》(人文研蔵本)
の図1に相当するページ

同じページのXML例は以下の例4のとおりとなる。ここでは先ずページの始めに<pb>(行1)要素で新しいページを記述する。それと別に、各行の開始も記述したい、そのために<lb>要素を用いる(行2)。このXMLテキストは二つのエディション、つまり《重刊道藏輯要》(CK)と《道藏輯要》の人文研蔵本(YP-Z)で作成されている、その両者の違いは<app>の要素で記述されている(行2、行11)。文書の構造は<div>要素(行3、6、28)、見だしは<head>(行4、7)でマークされている。

```

1 <pb n="JY161_02p001a"/>
2 <lb n="JY161_02p001a_01"/><fw><app><rdg wit="CK">重刊<
  /rdg><rdg wit="YP-Z"/></app>道藏輯要</fw>
3 <div>
4 <lb n="JY161_02p001a_02"/><head> 金華宗旨</head>
5 <lb n="JY161_02p001a_03"/><byline>孚佑上帝純陽呂祖天師著</byline>
6 <div>
7 <lb n="JY161_02p001a_04"/><head> 天心第一</head>
8 <lb n="JY161_02p001a_05"/><q>呂祖曰自然曰道道無名相一性而已一元神而
  己性命不可
9 <lb n="JY161_02p001a_06"/>見寄之天光天光不可見寄之兩目古來仙真皆口口
  相傳傳
10 <lb n="JY161_02p001a_07"/>一得一自太上化現東華遞傳巖以及南北兩宗全真
  可為極
11 <lb n="JY161_02p001a_08"/>盛盛者盛其徒眾衰者衰於心傳以至今日
  <app><rdg wit="CK">泛濫</rdg><rdg wit="YP-Z">濫泛</rdg></app>極
  矣凌替極

```

12 <lb n="JY161_02p001a_09"/>矣極則返故昔日有許祖垂慈普度特立教外別傳之
旨聞者

13 <lb n="JY161_02p001a_10"/>千劫難逢受者一時法會皆當仰體許祖心先於人倫
日用間

14 <pb n="JY161_02p001b"/>

15 <lb n="JY161_02p001b_01"/>立定脚跟方可修真悟性我奉勅為度師今以太乙金
華宗旨

16 <lb n="JY161_02p001b_02"/>發明然後細為開說太乙者無上之謂丹訣甚多總假
有為而

17 <lb n="JY161_02p001b_03"/>臻無為非一超直入之旨我傳宗旨直提性功不落第
二法門

18 <lb n="JY161_02p001b_04"/>所以為妙金華即光也光是何色取象於金華亦秘一
光字在

19 <lb n="JY161_02p001b_05"/>內是天仙太乙之真炁水鄉鉛只一味者此也回光之
功全用

20 <lb n="JY161_02p001b_06"/>逆法方寸中具有鬱羅蕭臺之勝玉京丹闕之奇乃至
虛至靈

21 <lb n="JY161_02p001b_07"/>之神所注儒日虛中釋曰靈臺道曰祖土曰黃庭曰
關曰先

22 <lb n="JY161_02p001b_08"/>天竅蓋天心猶宅舍一般光乃主人翁也故一回光則
周身之

23 <lb n="JY161_02p001b_09"/>氣皆上朝諸子只去回光便是無上妙諦光易動而難
定回之

24 <lb n="JY161_02p001b_10"/>既久此光凝結即是自然法身而凝神於九霄之上矣
心印經

25 <pb n="JY161_02p002a"/>

26 <lb n="JY161_02p002a_01"/>所謂默朝飛昇者此也</q>

27 <lb n="JY161_02p002a_02"/><p>金華即金丹神明變化各師於心</p>

28 </div><div>

29 <lb n="JY161_02p002a_03"/><head> 元神識神第二</head>

30 <lb n="JY161_02p002a_04"/><q>呂帝曰天地視人如蜉蝣大道視天地亦泡影惟
元神真性則

31 <lb n="JY161_02p002a_05"/>超元會而上之其精氣則隨天地而敗壞矣然有元神
在即無

32 <lb n="JY161_02p002a_06"/>極也生天生地皆由此矣學人但能護元神則超生在
陰陽外

33 <lb n="JY161_02p002a_07"/>不在三界中此見性方可所謂本來面目是也最妙者
光已凝

34 <lb n="JY161_02p002a_08"/>結為法身漸漸靈通欲動矣此千古不傳之秘也</q>

35 <lb n="JY161_02p002a_09"/><p>識心如強藩悍將遙執<q>[糸*巳]</q>綱久
之太阿倒置矣今凝守元宮回

36 <lb n="JY161_02p002a_10"/>光返照如英主在上大臣輔弼內政既肅自然強悍懼
伏矣</p>

例4：《道藏輯要》 p 5353 のXML 例

こうしたマークアップでは文書の構造は記述してあるが、<q>要素(行8、30)と<p>要素(行3、11)では段落レベルのテキストも記述された。しかしテキストの中身、つまり句読点、センテンスの区切り、人名、地名などのマーク

アップは未だ出来ていない。 それでも、ここまでマークアップを付けるだけでも既に大変な費用がかかるし、単に電子化するの何倍にも登る。そして御覧のとおり文字かされた電子テキストとデジタル・ファクシミリとしての電子テキストは両方例に出したが、実際には両者の関係はマークアップでは記述されていない。 こんなような関係が TEI P5 で初めてマークアップ出来るようになった。

2.2 TEI P5 1.0 (2007) : デジタルファクシミリの追加

デジタル・ファクシミリも TEI のテキスト・モデルに入れるべきだ、従来のテキスト・モデルが拡張する必要があるとの指摘を受けて、2007 年に出された、初めて完全に XML 対応出来るようになった TEI P5 勧告には <facsimile> という新しい要素が追加された。その <facsimile> は従来の <text> 要素と並列に並んで(例 5)、或いはその代わりに使用可能(例 6)。

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <facsimile> ( ...省略... ) </facsimile>
4 <text> ( ...省略... ) </text>
5 </TEI>
```

例 5 : ファクシミリとテキスト両方使用

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <facsimile> ( ...省略... ) </facsimile>
4 </TEI>
```

例 6 : ファクシミリのみ使用

これで TEI のテキスト・モデルに同じテキストを二通りの表現が可能になった。両者の関係は @fac という属性(attribute⁶)で記述できる :

```
<pb facs="CK-KZ_JY161_02p001a.png"/>
```

@facs の与は "URI" であるので、直接画像のファイル名を書くか、それでも <graphic> 要素へのリンク書くか、どちらでも可能である(下の例 XX も参照)。この @facs はグローバル(TEI で定義された全ての要素に使える)なので、制限なしでどんな要素にも、どこであつても付けられる。

P5 では <facsimile> の追加で従来のテキスト・モデルが拡張されていたが、本来の SGML 版で可能であった、物理的な構造を構造としてマークアップ化することは亦不可能。この点については TEI のメーリングリストには度々議

6 XPath や XSLT には属性は '@' で区別すれので、XML に関する文書もよく登場するので、本稿でも使う。

論が行ったと、早い段階でワーキンググループが発足された。2009年と2010年のTEIメンバーズ・ミーティングでは暫定的な報告があった上で、あらゆる議論を踏まえて、2011年末のTEI P5 2.0にその結果が追加された。

2.3 TEI P5 2.0 (2011) <sourceDoc>の追加

TEI P5では画像としてのテキストはテキスト・モデルに導入されたが、SGML時代のように、物理的な階層（つまりページ、行などの集合）と論理的な階層（章、節、段落、文などの集合）は未だ不可能である。しかし、例4のような、理論的な構造に基づいた記述は《道蔵輯要》のようなテキストでは手動で作成される、自動化は不可能である。でも実際に多くの場合はこうした手間の掛からない、或いはプロジェクトの後の段階で計画されているが、先ず物理的な記述だけをXML化でしたい。こうした要求を受けて、昨年の年末にリリースされたTEI P5 2.0にはTEIのテキスト・モデルがさらに拡張された。今回は<sourceDoc>という要素が、やはり<text>と<facsimile>と平行に使用する、テキストのもう一つの記述手段となる。

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <sourceDoc> ( ...省略... ) </sourceDoc>
4 </TEI>
```

例7 TEI P5 2.0 の新しい<sourceDoc>の使用例

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <sourceDoc> ( ...省略... ) </sourceDoc>
4 <facsimile> ( ...省略... ) </facsimile>
5 </TEI>
```

例8 <sourceDoc> と <facsimile>

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <facsimile> ( ...省略... ) </facsimile>
4 <sourceDoc> ( ...省略... ) </sourceDoc>
5 <text> ( ...省略... ) </text>
6 </TEI>
```

例9 <sourceDoc>、<facsimile> と <text>

例7～例9のとおり、それぞれのテキスト・モデルは他のテキスト・モデルと自由に合わせることが出来る。これでテキストのデジタルかの完全にTEIでの記述を使った新しいワークフローが可能になる：

- ステップ1：先ずは資料のスキャンして、デジタル・ファクシミレを

作成して、<facsimile>で記述する(自動作成可能)

- ステップ2：次はスキャンしたテキストを文字化して、そしてページと行毎に<sourceDoc>でマークアップを付ける(テキストがあれば自動化可能)
- ステップ3：最後に<text>を使って、従来のとおりのマークアップテキストを作成する(大抵手動)

その場合はそれぞれのテキスト・モデル間の参照が必要になる。TEIはその為はどんなような方法を提供するかは次に紹介したい。

2.3.1 <facsimile>と<sourceDoc>の実例

```
1 <TEI>
2 <teiHeader> ( ...省略... ) </teiHeader>
3 <facsimile>
4   <graphic url="CK-KZ_JY161_02p001a.png"
   xml:id="JY161_02p001a"/>
5   <graphic url="CK-KZ_JY161_02p001b.png"
   xml:id="JY161_02p001b"/>
6   <graphic url="CK-KZ_JY161_02p002a.png"
   xml:id="JY161_02p002a"/>
7 </facsimile>
8 </TEI>
```

例10 ステップ1で作成されたTEIの実例

例10はステップ1の一番簡単な例を紹介する。ここではJY161《金華宗旨》の文献の三つのページがスキャンされた、そして<graphic>要素としてはTEI文書のなかでアドレス可能になった。しかし、次のステップ2(例12を参照)のような使用が見込めば、今の段階も<surface>要素(これはページなどの文字が書かれている表面を抽象化した要素で、本来にあった<page>要素の代わりに使える)を使うのが勧められる:

```
<surface>
  <graphic url="CK-KZ_JY161_02p001a.png" xml:id="JY161_02p001a"/>
</surface>
```

例11には上記の例4に<facsimile>のセクションを追加された。今回は<surface>の中で<zone>要素を使用して、画像内の位置を記録する⁷。<zone>にも@xml:idでアドレス可能なID(行5、6、7)を与えるので、このIDでテキストからリンクを付けることが出来る(行13、25)。

7 この位置は右上(upper left == ul)と右下(lower right == lr)の二点で長方形を定義する。こうしたポイントの定義に使う座標系のゼロは画像の右上。この与は画像ファイルと等値である必要が無いので、違う解像度などのファイルに同じ座標系が使用する。

```

1 <TEI><teiHeader> ( ...省略... ) </teiHeader>
2 <facsimile
3   <surface ulx="0" uly="0" lrx="3633" lry="2645"
   xml:id="p5353">
4     <graphic url="p5353.png"/>
5     <zone ulx="93" uly="102" lrx="2535" lry="1716"
   xml:id="A"/>
6     <zone ulx="130" uly="1865" lrx="2544" lry="3471"
   xml:id="B"/>
7     <zone ulx="1195" uly="3490" lrx="1520" lry="3570"
   xml:id="C"/>
8   </surface>
9 </facsimile>
10 <text>
11 <body>
12 <p>
13   <pb n="JY161_02p001a" facs="#A"/>
14   重刊道藏輯要
15   金華宗旨
16   孚佑上帝純陽呂祖天師著
17   天心第一
18   呂祖曰自然曰道道無名相一性而已一元神而已性命不可
19   ( ...省略... )
20   千劫難逢受者一時法會皆當仰體許祖心先於人倫日用間
21   (ref:mid)
22   立定脚跟方可修真悟性我奉勅為度師今以太乙金華宗旨
23   ( ...省略... )
24   既久此光凝結卽是自然法身而凝神於九霄之上矣心印經
25   <pb n="JY161_02p002a" facs="#B"/>
26   所謂默朝飛昇者此也
27   金華卽金丹神明變化各師於心
28   元神識神第二
29   呂帝曰天地視人如蜉蝣大道視天地亦泡影惟元神真性則
30   ( ...省略... )
31   光返照如英主在上大臣輔弼內政既肅自然強悍慴伏矣
32   ( ...省略... )
33 </p></body></text></TEI>

```

例 11 <facsimile> と <text> のリンク

例 11 にはデジタル・ファクシミリとテキストの間にはリンクが付けれるようになったが、テキストそのものは従来のとおり、論理構造に随ったもの。しかし、上で述べたように、ステップ 2 には亦構造てきなマークアップが導入しないケースが多い。こうした場合には新たに追加された<sourceDoc>を使用する。

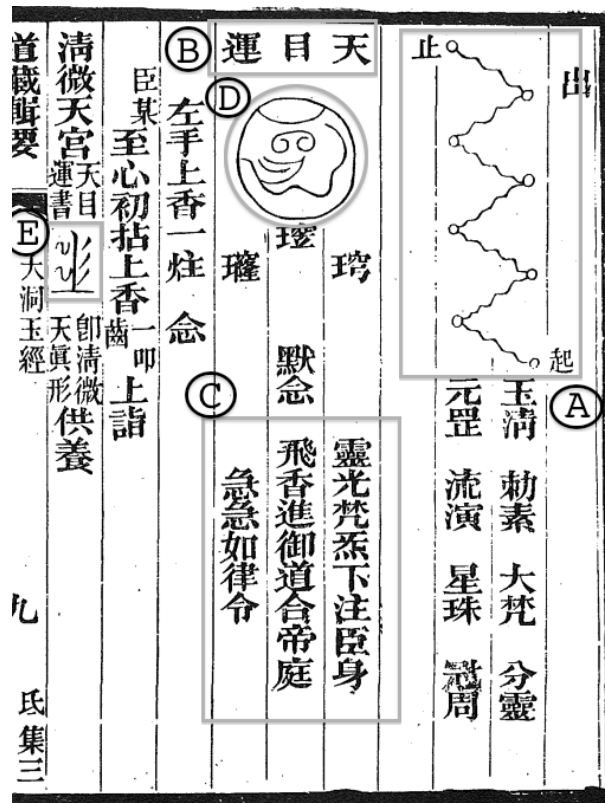


図3: 《道藏輯要》のJY006《元始大洞玉經》の一葉

例12には図3に対応するページにマークアップを付けた。このページには文字以外に図版が印刷されている、この図版の位置は図3で灰色の長方形とA○からC○とE○の文字で示されている。例12では行8からA○、行14からB、行18からCと行32からEへのリンクが実現されている。この例では画像の<zone>にしか座標を付けていないが、勿論行9のように<line>しか包含しない<zone>にも付けられる。また、行16のように長平系ではない、任意のポリゴンで記述できる形でも可能だが、その場合は@pointsでポリゴンの点を定義する。

```

1 <sourceDoc>
2   <surface ulx="0" uly="0" lrx="512" lry="680">
3     <graphic url="CK-KZ_JY006_02p009a.png"/>
4     <zone>
5       <line>出</line>
6     </zone>
7     <zone ulx="328" uly="16" lrx="480" lry="313">
8       <img xml:id="JY006_02p009a_A" type="image"/>
9     </zone>
10    <zone ulx="358" uly="308" lrx="453" lry="610">
11      <line>玉清 勅素 大梵 分靈</line>
12      <line>元罡 流演 星珠 冠周</line>
13    </zone>
14    <zone ulx="170" uly="11" lrx="310" lry="50">

```

```

14     xml:id="JY006_02p009a_B">天目運</zone>
15     <zone points="236,73 273,90 287,118 283,161 243,180
203,169 185,130 197,93"
16     xml:id="JY006_02p009a_D" type="image"/>
17     <zone> 默念</zone>
18     <zone ulx="170" uly="70" lrx="306" lry="235"
xml:id="JY006_02p009a_C">
19         <line>靈光梵炁下注臣身</line>
20         <line>飛香進御道合帝庭</line>
21         <line> 急急如律令</line>
22     </zone>
23     <zone>
24         <line>          左手上香一炷 念</line>
25     </zone>
26     <zone>
27         <line> <zone><line>臣某</line></zone>至心初拈上香
28         <zone><line>一叩</line><line>齒</line></zone>上詣<
/line>
29         <line>清微天宮
30         <zone><line>天目</line><line>運書</line></zone>
31         <zone lrx="30" lry="168" ulx="76" uly="231"
32         xml:id="JY006_02p009a_E" type="image"/>
33         <zone><line>卽清微</line><line>天真形</line></zone>供養
</line>
34     </zone>
35 </surface>
36 </sourceDoc>

```

例 1 2 <sourceDoc> と画像のリンク

必要であれば、<surface>か<zone>のなかには<surface>か<zone>をネストすることも可能だ。例えば例 1 2 にページ内の画像を切り抜いた、画質のよいバージョンにリンク付けたい場合には、上記の行 8 の<zone>は以下のように書き換える：

```

<zone ulx="328" uly="16" lrx="480" lry="313"
xml:id="JY006_02p009a_A" type="image">
<surface ulx="0" uly="0" lrx="152" lry="297">
<graphic url="JY006_02p009a_ZoneA.png"/>
</surface>
</zone>

```

3. 終わりに

TEI のマークアップ勧告は 1987 年以來の 25 年間に様々なテキストで使われて、人文科学における学術目的のデジタル・テキスト作成の基盤となった⁸。

8 例えばアメリカの現代語学会(MLA)は文献学的の編集のためのガイドラインでは TEI 使用を勧めている。そのガイドラインとそれに基づいた実践については Burnard 2006 を参照、昨年《人文学と電子編集—デジタル・アーカイブの理論と実践》のタイトルで日本語訳も出版された。

そのマークアップの実践で得られた経験を踏まえて、TEI のテキスト・モデルは何回拡張されたことがある、又は SGML から XML への変更の過程でもそのテキスト・モデルが多きな影響を受けた。しかし、最近のバージョンではとって基本的な、簡単なマークアップから、きめ細かな、複雑な付け方まで、幅広い適用可能な、多面的なモデルが定義された。これに基づいては完全に TEI を利用して、段階的にマークアップを付け加えるようなワークフローも可能になったので、これから日本でも更なる使用が期待される。

参考文献

Bryan, Martin (山崎俊一、福島誠 訳)、《SGML 入門》、アスキー出版局、1991.

Burnard, Lou; O'Brien O'Keeffe, Katherine; Unsworth, John, *Electronic textual editing*, Modern Language Association of America, 2006 [(明星聖子、神崎正英 訳) : 《人文学と電子編集—デジタル・アーカイブの理論と実践》、慶應義塾大学出版会、2011].

Coombs, James H.; Renear, Allen H.; DeRose, Steven J., “ Markup Systems and the Future of Scholarly Text Processing ”, *Communications of the ACM*, p.933-947, URL: <http://xml.coverpages.org/coombs.html> (2012-02-02 アクセス).

DeRose, Steven J.; Mylonas, Elli; Renear, Allen H.; Durand, David G., “ What is text, really? ”, *Journal of Computing in Higher Education*, p.3-26.

Esposito, Monica: 「清代道教における三教の寶庫としての『道藏輯要』」 in: 麥谷邦夫(編) 《三教交渉論叢續編》、京都 2011, p.431-445.

Goldfarb, Charles F and Rubinsky, Yuri, *The SGML handbook*, Oxford 1990.

Ray, Erik T. (宮下尚、牧野聡、立堀道昭 訳)、《入門 XML》、オライリージャパン、第 2 版 2004.

TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [バージョン 2.0.1]. [最終更新日 2011-12-22]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (2012-02-02 アクセス).

Wittern, Christian, 「道藏輯要の編纂と電子化をめぐる諸問題」, in: 麥谷邦夫(編) 《三教交渉論叢續編》、京都 2011, p.471-499.

漢字文献における電子的翻刻の課題

— 或いは翻刻者の使命 —

白須裕之

shirasu.hiroyuki@gmail.com

要旨: 従来より電子テキストと言え、情報の劣化を問題視されることが多いが、その劣化とはどのような意味で言われているのであろうか。そこには文献や電子的翻刻において対象が同一のものとして扱えるのはどのような時であるか、電子的翻刻とはどのような行為であるかということに対する暗黙の了解が存在している。本稿ではこのような暗黙の前提を明かにすることを通して、漢字文献における電子的翻刻の可能性について議論することを目的とする。情報抽出の根拠である電子テキストがどのような対象であるのかということについて、再考するきっかけになればと思う。

キーワード: 反覆可能性, 翻訳者の課題, 情報の流れ, 文字転写の理論

1 はじめに

人文学に限らず、資料を保存し、その資料について分析したり、情報を抽出したりという事柄を計算機を利用して行おうとする場合、どのような対象が保存され、また分析や情報抽出が対象のどのような面に向けられているのかを省察しない訳にはいかないであろう。計算機という手段を利用する場合には、このような省察は計算機の利用者に要請される切実な課題である。しかし、はたして計算機を利用する場合だけにこの課題は限られるものなのであろうか。

資料が言語的/非言語的なものであるに依らず、また、芸術・科学・日常的な体験などの分野に関わることなく、以上の課題を解決することが要求されるであろう。本稿では「漢字文献に対する電子的翻刻の課題」について、その予備的な省察を試みるつもりである。この課題を以下のような問いとして捉え直すことができる。

1. 電子的翻刻の対象は何如なるものであるか？
2. 電子的翻刻によって保存されるのは何であるか？
3. 数理的な分析における正当性の根拠は何処にあるのか？

電子的翻刻によって保存されるものは対象と無関係であっては意味がないし、分析されるものが保存されたものを媒介として、対象のどのような面を情報として抽出しているかというように、これらの議論は互に関連していると言える。ここでこれらの疑問に対して我々はすぐに答えを与えることができるだろうか。本稿ではこれらの問いを扱うための視座を用意することから始める。

まず、以下の J. Derrida の問い「文学的对象は何か」を出発点にする。Derrida は日本講演「私の

立場 — デリダは答える」 ([8] 所収) において、初期の仕事を振り返り、自身の仕事がこの問いを経由することによって成されたと述懐している。彼は自身の問い「文学的対象とは何か」を「文学的対象をしてイデア的対象たらしめるものは何か」という問いとして問い直している。Derrida が何故イデア性を問題にするのかというと、それは数学的対象のイデア性の特殊化として、文学的対象を以下のように捉えているからである。

イデア的対象とは自己との同一性をもつ対象のことですが、その同一性はその対象の諸種の経験的な書き込み、経験的な写しのいずれとも混同されないといったそういう同一性です。... 一つの文学作品をして、そのいわば物質的なもろもろの受肉のいずれとも混同されないイデア的対象たらしめるのは何か、と。要するに、一冊の書物の統一性、... は、書物の写しを通じてそれらを無限に再生産して、あるいはまたわれわれがその対象について無数の経験をもちうるにしても、同一のものとしてとどまるのです。(文献 [8] 邦訳 209 頁)

我々も電子的翻刻における対象として、まず最初にこのようなイデア的対象を取り上げるべきである。すなわち電子的翻刻に限らず、過去の様々な媒体、そして未来の媒体に対しても同一であるようなもの、これが解明できれば電子的な翻刻の対象が何であり、保存されるべきものがどういったものであるかを問うことができると考えられる。

Derrida はこのイデア的対象の解明の手掛かりを、以下のように『幾何学の起源』における Husserl の思索に求める。

数学的なイデア的対象は書かれた表記を必要とする。そして — この点がフッサールの独創的なところなのですが — イデア的対象がイデア的対象となり、したがって伝統のなかへ入り、いわば歴史的になるためには、それは書かれた表記を必要とする、と。これと同様に文学的対象もまた、おのれがあるところのものすなわち文学的対象となるためには、もしこう言っていれば数学的な対象の場合よりずっと多くの特殊な書き込みを必要とするのでした。(文献 [8] 邦訳 210 頁)

Derrida はこのような問題から出発して、エクリチュールの問題系に入っていく。

書くとは何を意味するのか。書くとは何か。何かを書きしるすとき、人は何をしているのか。そして、書かれたものの地位ないしは構造はどうであるのか。(文献 [8] 邦訳 211 頁)

この問いは、「書く」ということを「電子的に表現する」ことに置き換えれば、まさに我々のここでの問いである。勿論両者の違いも考察していかなければならない。Derrida のイデア的対象からエクリチュールの問題系への道を辿り直すことによって、我々は「電子的に表現する」ことの前提条件を明かにし、その過程で我々の問いを議論するための視座を得ることが本稿の目標である。

Derrida の思惟はイデア性を検討することによって、言語の可能性の条件を「反覆可能性」として捉えた。言語的なイデア的対象性が「反覆可能性」として捉えられるとき、表記としての電子的表現の原理は「反覆可能性のアポリア」を抱え込むことになる。本稿の以下の議論では、このアポリアを解きほぐす道を Benjamin の「翻訳者の使命」という思想に見出す。

2 電子的翻刻の条件

2.1 アイデア的对象性

前節で引用した「アイデア的对象は書かれた表記を必要とする」ということから始めよう。数学や論理学で扱われる概念、数、集合、命題、定理、証明などの概念はそれ自体が、何如なる時代、何如なる場所においても客観的なものとして現われる。それは特定の個人(発見者)の意識の中で見出されたとしても、それを受け入れようとする多くの人々には、その特定の状況に依存せずに同一のものとして認識されなければならない。このような発見者の主観から離れて、理念的客観性を獲得することが何故に可能であるかという問いは、以下のような「言語」の問題として捉えられる。

幾何学的意味の主観的明証がその理念的客観性を獲得する仕方について問う場合、われわれはまず理念的客観性がたんに幾何学的あるいは科学的真理の特性であるだけではないことに注意しなければならない。それは言語一般の要素である。(文献 [5] 邦訳 90 頁)

「理念的客観性はいかにして可能か」という問いは、「対象一般はいかにして可能か」と問いの模範であり、理念的客観性への階梯を登ることによって「対象一般はいかにして可能か」を問うことができる。その階梯は以下のような 3 つのアイデア性である*¹。

1. 表現の理念性 — 例えば、語の理念的客観性があげられる。その感性的、すなわち音声的あるいは表記的な受肉化は特定の言語内で自由である。
2. 意味の理念性 — Husserl が言う「志向的内容」*²あるいは「意味の統一性」であり、同一な内容が様々な国語から思念される。すなわちこの理念的同一性はその翻訳を可能にする。
3. 純粋な理念性 — 対象そのものの理念性

最も純粋で、あらゆる実在の束縛を離れたアイデア的对象性が、書かれるという具体的な経験(その後のエクリチュール概念)を存立の前提としていることを Derrida は以下のように述べる。

理念的客観性は、それが世界のなかに刻まれないかぎり、あるいはむしろ刻まれうるのでないかぎり、その純粋な意味において標識ないし着衣以上のものである受肉に身を委ねうるのでないかぎり、完全には構成されない。書く行為はそれゆえ、あらゆる「構成」の最高の可能性である。(文献 [5] 邦訳 134-135 頁)

次の小節ではアイデア的对象の存立の前提を、Derrida の「差延」以降の鍵語である「反覆可能性」に見ることにしたい。その課題に移る前に『幾何学の起源』と関連して、ここで電子的な翻刻においても重要な事柄である以下を指摘しておく必要がある。「言語活動が絶対的な理念的客観性を表現する限

*¹ この理念的客観性へと階梯を登ることは形相的還元に対応する。

*² Husserl は記号をその機能に従って「指標」と「表現」に分類する。記号から現実存在への指示機能である「指標」を還元したものが「表現」であり、志向的内容はその「表現」の意味に対応する。

り、この客観性を構成しているものをその中に認識するという事は、... 相互主観性が客観性の条件である」ということであり、それは以下の条件を一つの前提としている。

普遍的言語活動の — 間接的あるいは直接的 — 地平の可能性は、本質的な困難と限界に突き当たる危険をもっている。それはまず、言語活動の「純粹文法」および「アприオリな規範」の可能性という危険な問題が解決済みであることを仮定している。この可能性はフッサールにとって自明のものであること決してやめなかったのである。ついでそれは、「すべてのものは極めて広い意味で命名可能である。すなわち言語的に表現可能である」ということを前提している。(文献 [5] 邦訳 115-116 頁)

2.2 反覆可能性について

数学的なイデア的对象は常に同一性をもつものとして反復可能であった。ここでは Derrida の「差延」以降の鍵語である「反覆可能性」*3を手掛かりに、対象を可能ならしめる反復、及び一般の言語作用・記号作用一般の条件を考える。まず、彼は文献 [7] で以下のように述べる。

〈反復可能性〉は理想化＝イデア化を可能ならしめる — したがって、多様な事實的出来事から独立した一定の反復可能な独立性を可能ならしめる — けれども、それが可能ならしめる理想化＝イデア化そのものに制限を加える。すなわちそれは、理想化＝イデア化を〈開始しながら損なう〉のである。(文献 [7] 邦訳 124 頁)

〈開始しながら損なう〉ということが反復との違いを際立たせている*4。その内実は文献 [6][7] に示されているが、ここでは文献 [17] から言語作用を可能にする条件としての、反覆可能性の意味を読み取ることにしよう。

反覆可能性を取り上げねばならないのは、言語一般の成立条件として、「言語には、ある構造的な不在が必然的な可能性として内蔵している」からである。ここで言っている不在とは「送り手、受け手、意味、意図、規則、慣習、コード、発信のコンテクスト」等の不在のことである。ある言語記号(マーク)がそれらの不在においてもなお言語としての働きを失わないのは「反覆可能性」によってである。この不在は当然ながら電子的翻刻においても考慮すべき事柄である。「主観が規定したはずのコードの不在においてもなお、言語が言語として反復されうる」のは、以下のような〈開始しながら損なう〉ことによってである。

記号的同一性(意味のユニット)は、読み手が自らの解読コード(読解規則)を同時に措定することで創出されたものである。... これは言語の同一性を反復可能にするコードそのものをその都度同時に自己措定しつつ当の同一性を構成する(すなわちまた再破壊し脱構成する)と

*3 文献 [17] に従って、*répétition* と *itération* を各々反復、反覆と記す。但し、文献 [7] の邦訳からの引用ではそのまま反復、〈反復〉を使用する。

*4 「反覆可能性」は本節で簡単に述べただけでは汲み尽せない広い射程を持っている。ここでは本稿の主題である電子的翻刻に関連する内容のみに触れた。

いう条件、つまりは同一性の可能性の条件であると同時に不可能性の条件なのである。(文献 [17] 75 頁)

これらの条件を考慮すると、以下のような「反覆可能性のアポリア」は電子的翻刻において深刻な問題となって現われる。

「反覆可能性」と呼ばれている何ものかについて語るためには、我々はそれを名指し、同定し、記述しなければならず、そうすることであたかも、それが一つの対象であるかのように扱わねばならない。換言すれば、我々はそれを、それ自身が疑問を付している仕方で把握しなければならぬ。(文献 [17] 79 頁)

目的論的に要請された価値、規範、意図などの体系化を前提とすることなく、「反覆可能性のアポリア」の認識を電子的翻刻において形式化するにはどのようにすれば良いのか、次節ではこのアポリアを前提に電子的翻刻の可能性を議論する。

3 電子的翻刻の可能性

3.1 アーカイブズのテーゼ再考

筆者は以前、文献 [13] において、アーカイブズを記号体系によって捉えるという「アーカイブズのテーゼ」を提出した。

アーカイブズとは「指示機能の再現」を行なう記号体系であり、アーカイブズの「永久保存」という概念は、この「再現」をいつでも行なえる仕組みのことである。さらにアーカイブの活動とは「表記的体系」*5を求めることにほかならない。

電子的翻刻の場合には勿論電子化という操作が必要である。これは簡単に言えば「デジタルな表記的記号体系」の構成が必要であるということである*6。この提唱はアーカイブズを構築する上で今でも大きな指針であると思っている。特にアーカイブズされる対象が予め与えられている訳ではなく、アーカイブズの構築が一つの解釈(記号体系)を表現するものであるという、積極的な面を提示しているからである。

但し、「アーカイブズのテーゼ」はこれまで見てきたような「反覆可能性」の認識を考慮しているとは言えない。「反覆可能性」はコード/コンテキストの解体を示唆していた。一つにはメタな記号体系を予め与えることの困難である。これは電子的翻刻でも考慮しなければならない。また、デジタルアーカイブズに「永久保存」性を期待できないというアーカイブズ学からの批判は、「反覆可能性」のおける不在

*5 文献 [13] では「表示的体系」となっているが、「表記的体系」の誤り。

*6 「表記的体系」「デジタル」という概念については文献 [13] を参照してほしい。

を考慮していないという点で、原理的にはそれほど射っていないということが分かる*7。アーカイブズの基本概念への再考が必要である所以である。

3.2 翻訳論

前節で見た電子的翻刻の前提に対して、本節ではそれを克服する道を Benjamin の「翻訳論」の思想に探すことにする。通常は翻訳という行為を、異なる言語体系の間で意味を伝達するものであると考えているが、Benjamin は論文『翻訳者の課題』[4]において、翻訳を意味の伝達とは捉えていない。

二つの異なる言語体系が与えられたとき、翻訳という行為はその間の同一であるような意味を移し変えるような出来事として捉えることはできない。二つの言語体系の関係及び違いが始めから分っている訳でも、意味のような同一物が予め存在している訳でもないからだ。これは言語を可能にしている「反覆可能性」でも見てきたことである。それでは翻訳を Benjamin はどのように捉えているのだろうか。

翻訳は、原作の意味に自らを似せようとするのではなく、むしろ愛をこめて、そして細部に至るまで、原作での意図する仕方を自分自身の言語のなかで付け加えてゆかなければならない。そのようにして翻訳は、ちょうどあの破片が一つの器の断片であると認められるように、原作と翻訳の両者をもっと大きな一つの言語の断片として認めることになる。(文献 [4] 邦訳 103 頁)

この失なわれた器とは Benjamin の言う「純粹言語」のことである。器をつなぎ合わせるためには破片は細部まで互いに適合しなければならないが、双方が同じ形である必要はない。翻訳者の使命とは「翻訳において純粹言語の種子を成熟させるという課題」を負っているということである。

純粹言語は、もはや何も意図せず、何も表現しないが、表現を欠いた創造的な言葉となつて、あらゆる言語において意図されたものそのものである。この純粹言語のなかで、最終的には、あらゆる伝達、あらゆる意味、あらゆる志向が一つの層に達する。そこではそれらがすべて消滅すべく定められている。(文献 [4] 邦訳 105 頁)

純粹言語は Husserl が述べた「志向的意味」を実現するものであり、伝達、意味、志向を必要としない言語である。Husserl が『幾何学の起源』で述べていた「相互主観性を客観性の条件」と見るとき、Derrida がそこに指摘していた仮定、すなわち普遍的言語活動が遭遇する困難と限界に対する回答を翻訳者の課題に見ることができる。

翻訳についてのこのような見方は、異なる言語体系の間での翻訳に限らず、同一言語内或いは非言語的表現への変換に対しても有効であろう。翻訳者が自身の言語を豊富にすることで、「純粹言語」への憧れを語らなければならなかったと同様に、我々も電子的翻刻において、その憧れを語らなければならない*8。

*7 これはアーカイブズ学の原理原則を否定するものではない。文献 [12] の批判は当時のデジタルアーカイブに向けられたものであり、また、現在でもデジタルアーカイブズの価値、信頼性を高めるために有効な批判であろう。但し、本稿の意図はアーカイブズ学の前提にも、「反覆可能性」の認識が開くような再考の可能性があることを示唆したいだけである。

*8 次節で文字の semantics について議論するが、この翻訳論的な見方は Unicode の universality に対する相補的な面

3.3 現代における数学的対象

前節で見たように、数学的なアイデアの対象性についての考え方が「表記」を媒介として捉えられた。電子的翻刻の可能性を議論し、言語表現に対する意味論の諸相を見るためには、20世紀以降発展した数学の「表記」の変様が、数学的なアイデアの対象性についての考え方に影響を与える点を簡単に見ておこう。

Benacerrafの一連の議論 [2][3] によって、数学の対象や真理への捉え方が確実に変わってきている。文献 [2] では、数学が扱っているのは数のような数学的な対象ではなく、それら相互の関係であると主張する。文献 [3] で提出された Benacerraf のジレンマは、認識論と存在論をどのように調停するかに関わっている。これ以降、数学では数学的対象の存在について議論することを避けようとする傾向にある。従来から数学の主題は対象どうしの関係、すなわち構造であるとする数学的構造主義が提唱されてきた。数学の哲学としては構造主義に対しても多くの立場があるが*⁹。対象についての記述が与えられたとき、その記述を満足する様々な対象のどれが基本的であるのかを問う必要がないという意味で、数学的な構造主義は利点を持つ。

現在、数学の表記には集合と圏の二つが主に使われているが、圏論では構造自身をオブジェクトとして陽に扱え、更に命題や論理もオブジェクトとして記述できる。このような意味で、電子的翻刻の形式化には圏論的な表現が今後大きな比重を占めることになるであろう。本稿での形式化もそのような方向に進むことになる。

3.4 書記行為理論

これまでは主に言語テキストを電子的翻刻の対象としてきたと言えるが、言語表現の意味を補完するものとして、表記そのものを扱う「書記行為」も対象とすることができる。文献 [15] では、P. Shillingsburg の「書記行為理論」[10] の提唱に従って、テキストアーカイブズを書記行為の再現系として捉える試みについて議論した。しかし、書記行為の意味を表現可能とするためには、「反覆可能性」が大きな役割を果していることは否定できない。

一般に意味論を構成する方法としては、大きく字義主義と文脈主義の二つに分けられる。たとえば文の意味を定式化するとき、字義主義は文を構成する様々な単位の意味のみから文の意味が決められるという立場、文脈主義は文に付随する文脈も含めて文の意味が決められるという立場である。それらを補完する立場として、「反覆可能性」の認識が開くような立場もあると思う。また、数学的対象として構造を直接扱う立場では、書記行為理論を文脈主義を越えて形式化する道が可能であろう。本稿ではそのような可能性があることを指摘するに留め、今後の課題としたい。

に注目する試みであると言える。

*⁹ 圏論における構造主義の哲学については、例えば文献 [9] を参照。

4 電子的表現系の形式化の試み

電子的翻刻の可能性について、前節では現代における数学的対象の再考、翻訳論、及び書記行為理論について見た。これらを参考に新たな電子的表現系の形式化について本節で議論しよう。特に「反覆可能性のアポリア」を扱うには、Benjamin の翻訳論の考え方が有効であることをここで示したい。

現在、我々はその有効性を示すために十分な理論的道具を持っていない。ここでは「情報の流れ」理論的な形式化が我々の目的に合う解釈を許容することを見る。特に漢字文献に特有の問題である文字の問題を扱いたいと思う。電子テキストの表現については今後の課題としたい。

4.1 文字の semantics について

まず文字の意味論を議論することから始めよう。文献 [21] では、Unicode のおける semantics について議論している。その議論を参考に、Unicode 規格の特徴を電子的翻刻の前提という視点から扱ってみよう*¹⁰。まず Unicode 規格において、character の semantics が文字が対象化される前に与えられていることについて以下のように述べる。

Unicode に収録された character には自動的にプロパティが結びつけられるので、潜在的に(先天的に)すべての character には Unicode の言う semantics を有することになる。(文献 [21] 119 頁)

また、Unicode の設計原則には universality が重要な役割を果たしているが、その Universalism については以下のように述べる。

「世界共通のレパートリー *universal repertoire*」による Unicode モデルは普遍言語 (*universal language*) 的な立場に準えることもできよう。(文献 [21] 120 頁)

更に Unicode 規格の問題点として、Unicode における character の弁別方法が規格の内部ではなく、規格外にある提案者の解釈に依存しているとして、以下のように述べる。

Unicode においては、character の同一性を決定するためのルールが存在しておらず、semantics の決定が事実上、規格の外部に依存していることになっている。Unicode の semantics とは、規格の外部で決定された差異の構造のみを規格内に取り込むための方便のことを指す言葉なのかもしれない。(文献 [21] 121 頁)

このように Unicode 規格における文字の扱いは、文字概念が予め与えられている点、その文字概念の universality、及び文字における差異が記述されていないという点が指摘できる。以下ではこれらの点に注目しながら、電子的翻刻における文字の形式化を検討しよう。

*¹⁰ ここでの議論はあくまで、「漢字文献における電子的翻刻」に使用できる文字の形式化を検討するために、Unicode 規格を参考にしようということであって、交換符号としての文字コードの問題を扱う訳ではない。

4.2 IF-Chaon モデル

文字の表現形式を構築するために設計の一般論から始めよう。文字というものを「反覆可能性のアポリア」から眺めたときに、読み手が解読コードを措定することで文字を創出するというを、以下に示すようにある種の設計として捉えることができる。

ここでは一般設計学を「情報の流れ」理論^{*11}として定式化した抽象設計論^{*12}を参考にしよう。まず一般設計学の枠組みを簡単に説明する。実体概念集合 S に対して、その部分集合族として属性概念集合 T^0 、機能概念集合 T^1 を考える。 $\langle S, T^0 \rangle, \langle S, T^1 \rangle$ は位相空間であることが要請される。設計とは仕様 $F \subseteq T^1$ が与えられたとき、解 $A \subseteq T^0$ を求めるものことであると定式化される (図 1 参照)。但し、解 A は $\bigcap A \subseteq \bigcap F$ を満たす。

$$\begin{array}{ccccc}
 F \hookrightarrow & T^1 & \xrightarrow{id_S^{-1}} & T^0 & \longleftarrow A \\
 & \downarrow \in & & \downarrow \in & \\
 & S & \xleftarrow{id_S} & S &
 \end{array}$$

図1 一般設計学の枠組み

抽象設計論は一般設計論の逆理を避けるために定式化されたものである。その違いは機能概念空間の T^1 は与えられた集合であり、対応する実体概念集合は S でなく、 T^1 の冪集合 $\mathcal{P}(T^1)$ とされる。それらの集合は図 2 で示した情報射を構成する。

文献 [14] では、文字表現のモデルである Chaon モデル [18][19] の形式化を提出したが、図 2 の情報射はまさに形式化された Chaon モデルである。以下、これを IF-Chaon モデルと呼ぼう。

$$\begin{array}{ccc}
 T^1 & \xrightarrow{f} & T^0 \\
 \Downarrow \ni & & \downarrow \in \\
 \mathcal{P}(T^1) & \xleftarrow{g} & S
 \end{array}$$

図2 IF-Chaon モデル

このように Chaon モデルを「情報の流れ」理論的に定式化された設計理論を通して眺めてみると、

^{*11} 本稿で必要な「情報の流れ」理論の用語については [1]、或いは付録 A を参照してほしい。文献 [1] は「情報の流れ」理論の基本文献である。また、文字論への応用は文献 [14][16] を参照してほしい。

^{*12} 詳しくは、文献 [11] 等を参照してほしい。

読み手が解読コード(文字素性)を措定することで文字を創出しているような、「弁別システム」として捉えることができる。すなわち、読み手が解読に使用したコード(弁別素性)を使用して、痕跡から創出した対象がどのようなものであるかを表現したものになっている。このような解釈を一般化すると、情報射は「弁別システム」の一般化、つまり「痕跡の形式化」であると見ることができる。

4.3 翻刻の理論

文献 [16] では文字転写の理論を「情報の流れ」理論的に定式化した^{*13}。

文字転写の理論で用いた「表現系」における情報射には、IF-Chaon モデルを用いることができる。上で見たように IF-Chaon モデルを「痕跡の形式化」として捉えるならば、文字転写の「情報の流れ」理論的な定式化は「痕跡の痕跡」を表現していると解釈することができる。「反覆可能性」に対して、歴史的なものの積み重ねや慣習をどのように理解するかにも依存するが、文字転写の理論はある意味でそのような歴史的なものの形式化を与えている。

5 おわりに

本稿では Derrida の初期の仕事を手掛かりに「漢字文献における電子的翻刻の課題」について考察した。Derrida の著作は言語について考えようとする場面で常に新たな示唆を与えてくれるが、ここでは電子的翻刻の前提を理解するという立場から、「反覆可能性」の一部の面を参照した。哲学的な訓練を受けていない筆者には思わぬ誤りがあったかもしれない。専門家のご教示を賜りたい。

昨今、Digital Humanities においても、Linked Open Data(LOD) について議論される機会が増えてきた。LOD が果たす役割は様々なデータを保有する機関の相互運用性という面だけではなく、現時点で予想することは難しいが、その発展が「反覆可能性」の認識が開く新たな未来を持っているかもしれない。本稿の形式化はその点でもなんらかの見通しを与えていると思う。

本稿で述べた「情報の流れ」理論的な文字の表現は、「反覆可能性」の認識が開く表現の可能性という意味ではその極一部を実現しただけである。その可能性を充分に開く新たな形式化については今後の課題としたい。

電子的翻刻者の使命

最後に翻刻者の課題について本稿の立場から簡単に触れておこう。これまでの予備的な省察を行なった視点で、節 1 で述べた問いを改めて見直してみると、これらの問いはある意味で解消されたと思えてよいと思う。電子的翻刻という行為の前に予めその対象を前提することができず、その対象が電子的翻刻によって明かにされることを見れば、対象が何であり、保存されるものが何であるかは問いとして意味がない。また、数理的な分析が電子的翻刻によって創出される対象に対してなされるのである

^{*13} その概要については付録 B、または文献 [16] を参照してほしい。なお [16] には『説文解字』『廣韻』などの古辞書の例についても述べた。

から、電子的翻刻の行為そのものにその正当性の根拠があると見る事ができる^{*14}。

但し、どのような翻刻が行なわれても良いという訳ではない。翻訳者の課題が原作の言葉のある意味で変形する以上、そこには聖なるテキストを豊かにする翻訳が試みられなければならないと同じように、文献の翻刻も同様の「翻刻者の使命 (課題)」を帯びていることに変わりはない。

異質な言語のうちに呪縛された純粋言語を、自分自身の翻訳の言語のなかで救済すること、作品のうちにとらわれた言語を作品の改作において解放すること、それが翻訳者の課題なのである。この課題のために、翻訳者は自分自身の言語の朽ちてもろくなった障壁を打ち壊す。(文献 [4] 邦訳 106 頁)

Derrida の他者論的な転回 (他なるものへの思考) は、「支配的な理論の構築装置」、意味の起源を解体するものであろうが、電子的翻刻はあくまで表現形式を追求するものであり、ある意味でそれに背く行為である。しかし、ここで議論した電子的翻刻者の課題を、我々自身の絶えざる使命として引き受けていかねばならない。Derrida が書くことによって「書くとは何か」を問い続けたように。

付録 A 「情報の流れ」理論

本稿に必要な範囲で、「情報の流れ」理論について纏めておく。詳しくは文献 [1] を参照されたい。

定義 1 (分類 Classification, 制約 Constraint) 分類 $\mathbf{A} = \langle \text{tok}(\mathbf{A}), \text{typ}(\mathbf{A}), \models_{\mathbf{A}} \rangle$ とは、トークンの集合 $\text{tok}(\mathbf{A})$ 、タイプの集合 $\text{typ}(\mathbf{A})$ 、その上の二項関係 $\models_{\mathbf{A}}$ からなる。

$\Gamma, \Delta \subseteq \text{typ}(\mathbf{A})$ とする。任意の $a \in \text{tok}(\mathbf{A})$ に対して、 $\forall \alpha \in \Gamma. a \models_{\mathbf{A}} \alpha$ ならば $\exists \alpha \in \Delta. a \models_{\mathbf{A}} \alpha$ が成り立つとき、 Γ と Δ は \mathbf{A} に対する制約と言い、 $\Gamma \vdash_{\mathbf{A}} \Delta$ と記す。

定義 2 (情報射 Infomorphism) 分類 \mathbf{A} から分類 \mathbf{B} への情報射 $f : \mathbf{A} \rightarrow \mathbf{B}$ とは、二つの関数 $f^* : \text{typ}(\mathbf{A}) \rightarrow \text{typ}(\mathbf{B})$ と $f_* : \text{tok}(\mathbf{B}) \rightarrow \text{tok}(\mathbf{A})$ の対で、任意の $b \in \text{tok}(\mathbf{B})$ と $\alpha \in \text{typ}(\mathbf{A})$ に対して、 $f_*(b) \models_{\mathbf{A}} \alpha \Leftrightarrow b \models_{\mathbf{B}} f^*(\alpha)$ を充すものを言う (以下、適宜 * を省略する)。

$$\begin{array}{ccc} \text{typ}(\mathbf{A}) & \xrightarrow{f^*} & \text{typ}(\mathbf{B}) \\ \vdots \Big\| \models_{\mathbf{A}} & & \Big\| \models_{\mathbf{B}} \vdots \\ \text{tok}(\mathbf{A}) & \xleftarrow{f_*} & \text{tok}(\mathbf{B}) \end{array}$$

定義 3 (シークエント Sequent, 分割 Partition) 集合 Σ に対して、 Σ の部分集合の組 $\langle \Gamma, \Delta \rangle$ を Σ のシークエントと言う。また、シークエント $\langle \Gamma, \Delta \rangle$ が $\Gamma \cup \Delta = \Sigma', \Gamma \cap \Delta = \emptyset$ を充すとき、 Σ' の分割と言う。

^{*14} 数理的な分析の正当性については更なる考察が必要であろう。翻訳が純粋言語への相補性の表現であることを考慮すれば、文献に対する電子的翻刻の相補性も本質的に不完全であるからである。

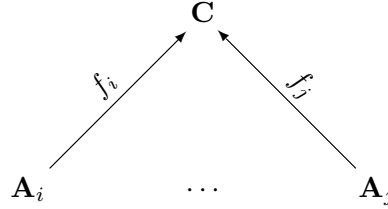
定義 4 (帰結 Consequence, 理論 Theory) $\mathcal{P}(\Sigma)$ 上の二項関係 \vdash を Σ 上の帰結関係と言う。このとき $T = \langle \Sigma, \vdash \rangle$ を理論と言う。また、 $\text{typ}(T) = \Sigma$ と記す。

定義 5 (正則理論 Regular theory) 理論 $\langle \Sigma, \vdash \rangle$ が以下を充すとき正則であると言う。

- Identity: $\alpha \vdash \alpha$,
- Weakening: $\Gamma \vdash \Delta \Rightarrow \Gamma, \Gamma' \vdash \Delta, \Delta'$,
- Global Cut: Σ' の任意の分割 $\langle \Sigma_0, \Sigma_1 \rangle$ に対して、 $\Gamma, \Sigma_0 \vdash \Delta, \Sigma_1$ ならば $\Gamma \vdash \Delta$ 。

定義 6 (局所論理 Local logic) 分類 \mathbf{C} 上の局所論理 $\mathcal{L} = \langle \mathbf{C}, \vdash_{\mathcal{L}}, N_{\mathcal{L}} \rangle$ とは、正則理論 $\text{th}(\mathcal{L}) = \langle \text{typ}(\mathbf{C}), \vdash_{\mathcal{L}} \rangle$ 、 $N_{\mathcal{L}} \subseteq \text{tok}(\mathbf{C})$ からなる。但し、 $N_{\mathcal{L}}$ は $\text{th}(\mathcal{L})$ の全ての制約を充たすトークンの集合である。 $N_{\mathcal{L}}$ の要素を正規トークン normal tokens と呼ぶ。

定義 7 (チャンネル Channel) $\{\mathbf{A}_i\}_{i \in I}$ を分類の族とする。チャンネル $\mathcal{C} = \{f_i : \mathbf{A}_i \rightarrow \mathbf{C}\}_{i \in I}$ とは、 \mathbf{C} への情報射の族であり、 \mathbf{C} をその核と言う。 $I = \{1, \dots, n\}$ のとき、 \mathcal{C} を n 項チャンネルと言う。



定義 8 (表現系 Representation system) 表現系^{*15} $\mathcal{R} = \langle \mathcal{C}, \mathcal{L} \rangle$ とは、二項チャンネル $\mathcal{C} = \{f : \mathbf{A} \rightarrow \mathbf{C}, g : \mathbf{B} \rightarrow \mathbf{C}\}$ と、核 \mathbf{C} 上の局所論理 \mathcal{L} からなる。このとき \mathbf{A} のトークンを \mathcal{R} の表現と言う。

- $a \in \text{tok}(\mathbf{A})$ が $b \in \text{tok}(\mathbf{B})$ の表現 representation であるとは、 $\exists c \in \text{tok}(\mathbf{C}). f(c) = a \wedge g(c) = b$ のときを言い、 $a \rightsquigarrow_{\mathcal{R}} b$ と書く。また、 $c \in \text{tok}(\mathbf{C})$ が正規トークンのとき、的確な表現 accurate representation と言う。
- $\Gamma \subseteq \text{typ}(\mathbf{A}), \beta \in \text{typ}(\mathbf{B})$ に対して、 Γ が β を表示する indicate とは、 $f(\Gamma) \vdash_{\mathcal{L}} g(\beta)$ のときを言い、 $\Gamma \Rightarrow_{\mathcal{R}} \beta$ と書く。

付録 B 文字転写の理論

「情報の流れ」理論を応用した「文字転写の理論」について簡単に述べる^{*16}。これは複数の文字体系の関係を記述するものである。文字体系の関係は文字の間のマッピングを拡張したものであり、文字だけを対象にするのではなく、文字の性質をどのように関係付けるかということも、その記述に含まれている。

^{*15} “Representation system” なる用語は元々は F. Dretske の構想に基づくもので、訳語としては「表象系」が適切かもしれない。「表象」という用語は表象主義を連想させるので、ここでは「表現系」を使用する。

^{*16} 実際の応用については文献 [16] を参照してほしい。

文字体系は文字をトークン、文字の性質をタイプとする分類で与える。ここでは簡単のために、文字体系を表現する分類はそのトークンが文字であるという制限を加える*17。

表現に利用する文字体系を分類 **A**、翻刻の対象となる文字体系を分類 **B** で与えたとする。このとき **B** から **A** への転写は、表現系 $\mathcal{R} = \langle \mathcal{C}, \mathcal{L} \rangle$ でモデル化する。但し、二項チャンネル $\mathcal{C} = \{f: \mathbf{A} \rightarrow \mathbf{C}, g: \mathbf{B} \rightarrow \mathbf{C}\}$ 、及び理論 \mathcal{L} は、以下のように **B** の言葉で書かれた事象や性質を、**A** の言葉で表現するように決められる*18。

- ・ **B** の文字 b を **A** の文字 a に転写するとき、 a が b の表現になる ($a \rightsquigarrow_{\mathcal{R}} b$)。
- ・ **B** の性質 β が、**A** の性質 α に反映されるとき、 $f(\alpha) \vdash_{\mathcal{L}} g(\beta)$ とし、それを理論に自然に拡張する。

体系 **B** には元々なかった制約が、体系 **A** の制約を通して、体系 **B** に押しつけられるということも、表現系を使用した理論では記述できる。これは文字コードでの翻刻において、文字コードでは区別できない釈文の文字があったり、文字コードでは区別があるが、釈文では区別しない文字があるなど、常に付き纏う現象を記述できるということである。

■謝辞 本シンポジウムの錚々たる研究者に比して未熟なる著者をそのメンバーに加えて下さり、本発表の機会を与えて下さったことを共同プロジェクト代表の山崎直樹さんに感謝致します。文献 [20] は本稿の問題を考察する上で重要なきっかけを与えてくれました。著者である師茂樹さんには人文情報学への考え方から多くを学んでおります。ここに記して感謝致します。本稿では電子的翻刻の課題に対して拙い考察しか提出できませんでしたが、人文学と情報学双方の専門家の方々が少しでもこの方面に目を向けて下さるような、きっかけが提供できていれば幸いです。最後にいつも支えてくれる妻留美と娘に感謝します。

参考文献

- [1] J. Barwise, J. Seligman: Information Flow: The Logic of Distributed Systems, Cambridge University Press, 1997.
- [2] P. Benacerraf: What Numbers Could Not Be, The Philosophical Review, 74:47-73, 1965.
- [3] P. Benacerraf: Mathematical Truth, The Journal of Philosophy 70, 661-679, 1973.
- [4] W. Benjamin: Die Aufgabe des Übersetzers, 1923. (邦訳 山口裕之訳: 翻訳者の課題;「ベンヤミン・アンソロジー」, 河出文庫, 2011. 所収)
- [5] J. Derrida: E. Husserl, L'Origine de La Géométrie, Traduction et Introduction, Presse Universitaires de France, 1962. (邦訳 田島節夫他訳:『幾何学の起源』序説, 青土社, 1980.)
- [6] J. Derrida: Signature événement contexte, in *Marges — de la philosophie*, Editions de Minuite, 1972. (邦訳 高橋哲哉訳: 署名 出来事 コンテクスト; 理想「デリダ — 言語行為とコミュニケーション」, 青土社, 1988. 所収)

*17 分類によって表現できる一般の文字体系では、トークンが文字であるとは限らない。このような文字体系の一般的な構成法については、文献 [14] を参照してほしい。

*18 この表現系では **A** が表現、**B** が対象である。チャンネルの左右が転写の向きと逆になっていることに注意。

- [7] J. Derrida: Limited Inc. abc ... in *GLYPH 2*, supplement, the Johns Hopkins University Press, 1977. (邦訳 高橋哲哉, 増田一夫訳: 有限責任会社 abc; 理想「デリダ — 言語行為とコミュニケーション」, 青土社, 1988. 所収)
- [8] J. Derrida: 高橋允昭編訳, 他者の言語 — デリダの日本講演, 法政大学出版局, 1989.
- [9] E. Landry, J.-P. Marquis: Categories in Context: Historical, Foundational, and Philosophical, *Philosophia Mathematica III*, 13(1), 1-43, 2005.
- [10] P. Shillingsburg: From Gutenberg to Google — *Electronic Representations of Literary Texts*, Cambridge University Press, 2006. (邦訳 明星聖子, 大久保讓, 神崎正英訳: グーテンベルクからグーグルへ — 文学テキストのデジタル化と編集文献学, 慶應義塾大学出版会, 2009.)
- [11] 菊池誠, 長坂一郎: 言語学の設計学への応用に関する研究, *精密工学会誌* 70(3), 2004; 72(1), 2006.
- [12] 五島敏芳: 「デジタルアーカイブ」における永久保存の概念, *情報処理学会研究報告* 2008-CH-79, 2005.
- [13] 白須裕之: 記号機能としてのアーカイブズ, *情報処理学会「人文科学とコンピュータシンポジウム」論文集*, 2008.
- [14] 白須裕之: 文字の指示概念に関する試論, *情報処理学会「人文科学とコンピュータシンポジウム」論文集*, 2008.
- [15] 白須裕之: 書記行為の再現系としてのテキストアーカイブズ, *情報処理学会「人文科学とコンピュータシンポジウム」論文集*, 2010.
- [16] 白須裕之: 古辞書のテキストアーカイブズ構築について — 文字転写の理論とその応用, *情報処理学会「人文科学とコンピュータシンポジウム」論文集*, 2011.
- [17] 宮崎裕助: 反覆可能性の法 — デリダ『有限責任会社』と行為遂行性の問題, *哲学・科学史論叢* 3, 71-93, 2001.
- [18] 守岡知彦: 文字オントロジーに基づく文字処理について, *情報処理学会研究報告* 2006-CH-72, 2006.
- [19] 守岡知彦, 師茂樹: 文字素性に基づく文字処理, *情報処理学会研究報告* 2004-CH-62, 2004.
- [20] 師茂樹: 一般キャラクター論としての文字論の可能性, 文字研究会編: 「新常用漢字表の文字論」, 勉誠出版, 2009. 所収
- [21] 師茂樹: 携帯電話の絵文字における semantics の問題, *東洋学のコンピュータ利用* 第 21 回研究セミナー, 京都大学人文科学研究所, 2010.

CiNii のメタデータ・デザイン

大向 一輝（国立情報学研究所）

i2k@nii.ac.jp

要旨：近年の学術情報サービスは、本来の目的である専門家に向けた情報提供だけでなく、広く一般に知識を伝える役割が求められている。また、データの入手性を高めることは、分野横断的な研究を促進する上でも有効であると思われる。現在、さまざまな機関が所持するデータをオープン化し、それぞれをつなぐ試みが進められている。本稿では、CiNii を中心とした学術情報サービスにおける応用とその可能性について議論する。

キーワード：書誌，メタデータ，Linked Open Data，学術情報流通

1. はじめに

研究活動を支援するためのインフラとして、学術情報サービスは重要な位置を占めている。学術情報サービスの主要な機能は目録・所在情報と文献情報の提供である。前者については電子化・ネットワーク化が完了し、現在では電子図書館システムを通じて各種の情報を利用することができる。また、文献情報のうち論文については、電子ジャーナルの仕組みによって情報の生産から入手に到るまでのすべての過程がデジタル化され、物理的実体を伴わない情報流通が実現している。

電子ジャーナルの普及によって研究者の利便性は確実に向上したが、影響はそれだけにとどまらない。文献の検索・発見・入手がリアルタイムで行えるようになったことで、研究コミュニティの情報共有のスピードが飛躍的に向上し、その結果として研究成果を得るための競争が激化している。研究の先鋭化あるいは多様化によって、研究者が読むべき文献の数は大幅に増加しており、個人の努力を超えるレベルに達しつつある。

こういった背景の下で、学術情報サービスは研究者個人の活動を支援するとともに、コンピュータによるアクセスや情報の2次利用を念頭に置く必要がある。そのためには、コンピュータが理解可能なメタデータを付加して情報提供を行わなければならない。

このような問題意識は学術情報分野だけでなく、ウェブそのものにも存在している。人間のためのウェブからコンピュータに開かれたウェブへという考え方は、ウェブの創始者であるティム・バーナーズ＝リーが提唱するセマンティックウェブの思想に合致し

ている。

セマンティックウェブではメタデータを用いてウェブ情報を構造化し、コンピュータによる自動処理を促進することを目指している。セマンティックウェブの構想では、メタデータの基盤として個別の概念に対応する URI を導入し、概念間の関係を XML / RDF やオントロジー言語によって記述する。こういったメタデータを多数用意することで、複数サービス間の横断検索・マッシュアップ(サービスの組み合わせ)から推論・人工知能までさまざまな応用が可能であるとされている。

後に述べるように、学術情報サービスとセマンティックウェブで用いられるメタデータの親和性は非常に高い。そこで、筆者が設計・開発・運用に関わる論文情報ナビゲータ「CiNii」におけるメタデータの提供に際し、セマンティックウェブ技術の積極的な活用を試みた。本論文では、CiNii のメタデータがどのように設計され、提供されているかについて詳説する。

2. 学術情報とメタデータ

一般に、情報システムにおける実用的なメタデータの設計を行う際には、以下の 3 つの課題を解決する必要があると思われる。

- 1) 何をメタデータとするか
- 2) どのようにメタデータを作成するか
- 3) どのようにメタデータを表現するか

学術情報サービスにおいては、上記の 2 つの課題はあらかじめ解決されていると見てよい。1 について、メタデータは書誌情報そのものである。何を書誌情報とするかは論文と書籍で異なる点はあるものの、タイトルや著者名などの基本的な項目に関しては見解が統一されている。また Dublin Core などの組織的な活動の成果として、一般的に用いられるメタデータの項目が定められている。2 についても、研究者・大学・学会・図書館などによって組織的かつ継続的にメタデータが作成される体制が構築されている。

一方、3 のメタデータの表現については、明快な標準は存在せず、さまざまな表現形式が併存しているのが現状である。いずれを選択するにせよ、利用目的によって望ましい形式が大きく異なるため、サービスごとに詳細な検討が求められる。用途によっては独自のフォーマットを定義することや、同じ情報を複数のフォーマットで提供することを考える必要がある。

メタデータフォーマットを定義する際には、構文・構造・語彙の 3 要素を検討する必要がある。

構文はメタデータの表記方法を規定する。ここでは、XML や CSV (カンマ区切りテキスト) といったファイル形式の中から選択することになる。また、ファイル形式とは水準が異なるが、データの形態もここで選択する。整形形式の XML であればツリー構造 (木構造) となり、CSV であれば表形式となる。また、RDF (Resource Description Framework) ではグラフ (ネットワーク) 構造のデータとなる。

構造は各メタデータ間の関係を規定する。XML を例に取ると、「人物」という概念の中に「氏名」「所属」「電話番号」がフラットに並ぶという構造と、「人物」に「氏名」「所属」が含まれ、「所属」の中に「電話番号」が含まれるという構造は似て非なるものである。このような階層構造の定義は非常に重要である。

語彙は各種メタデータの名前と詳細な定義を規定する。人物の項目に「People」と名付けるか「Individual」と名付けるかで、背景となる概念は微妙に異なるであろう。語彙は、さまざまなコミュニティが用途に応じて定義しており、それを利用することが多い。代表的なものが Dublin Core や知人関係を記述するための FOAF (Friend of a Friend) である。既存の語彙では表現できない概念がある場合には、新たに語彙を作ることもできる。

あらゆるメタデータはこれらの要素の組み合わせとして表現されているが、新たにメタデータフォーマット定義する場合には何を考慮し、どういった方針のもとで各要素を選択していけばいいのだろうか。実際には答は無数にあり、どれか一つが正解というわけではない。何のための、あるいは誰のためのメタデータであるかを問い続け、その上で適切と思われる選択をしていくことが重要である。

学術情報サービスにおけるメタデータの役割は 2 つに大別できるものと思われる。ひとつは「知識表現のためのメタデータ」である。ここでは、元データである書誌情報の意味構造をできるだけ保持し、受け手であるユーザやコンピュータに対してそのまま伝達できるようなフォーマットであることが望まれる。

もうひとつの役割は「道具としてのメタデータ」である。学術情報サービスが持っている膨大なデータを第 3 者が利用するための手段としてメタデータを定義するというアプローチである。ここでは、できるだけ利用者の理解が容易で使いやすいフォーマットであることが求められる。

この 2 つの観点には相反する要求が含まれている。情報を厳密に表現するためには、XML であれば多段階の階層を持つ深いツリー、RDF であれば複雑なグラフを記述すればよい。しかしながら、メタデータを利用する立場からすると、複雑な構造を持つデータは再解釈することが困難であり、これを処理するためのソフトウェア開発のコストもコンピュータによる処理コストも増大する傾向がある。

このように、厳密性とコストはトレードオフの関係にある。メタデータの設計においては、想定される用途をできるだけ明確にし、これに基づいてコストパフォーマンスの高い設計をする必要がある。

3. CiNii のメタデータ・デザイン

3.1 CiNii の概要

CiNii は国立情報学研究所（NII）が 2005 年から運営している国内最大規模の学術論文検索・提供サービスである。学協会や大学との連携によって電子化を行った約 400 万件の論文を提供するとともに、国立国会図書館の雑誌記事索引データベース、科学技術振興機構の電子ジャーナルデータベース J-Stage、各大学・研究機関が提供する機関リポジトリ上の論文情報を合わせて約 1500 万件の論文・記事を検索することができる。また、国立情報学研究所が別途構築している引用文献索引データベースに基づき、論文の引用・被引用関係を表示することができる。

3.2 ウェブ API とメタデータ

2 節で述べたように、メタデータのフォーマットは知識表現を目的とするのかデータの再利用を目的とするのかによってその姿が大きく異なる。CiNii のメタデータ設計においても、どちらを主目的にするのかを最初に決定する必要がある。結論としては、データの再利用性を主眼に置きつつ、知識表現としてもできるかぎり整合性を持たせるというアプローチを取るようになった。

こうした決定を行うにあたっての背景として、ウェブ API の普及がある。ウェブ API とは、ウェブサービスの提供者が外部の開発者に対して HTML 以外の構造化フォーマットでデータを提供するものである。ウェブ API によって、既存のデータの新たな応用や他のサイトの情報と組み合わせるマッシュアップを促進する効果がある。

そこで、CiNii のメタデータはウェブ API として利用しやすい形式で提供することを目指し、論理的厳密性を追求するよりもシンプルな構造であることを重視した。また、独自性をなるべく排除することで、既存のウェブ API を活用しているプログラムが少しの改変で CiNii のデータにも対応できるようにすることを心がけた。語彙の選択についても同様に独自の提案を避け、できるかぎり標準的なものを採用することとした。

3.3 CiNii のメタデータ

CiNii の検索機能、書誌提供機能で表示される情報にはタイトル・著者・刊行物名・巻・号・ページ・出版年月日・抄録などが含まれる。

これらの情報に対し、適切な語彙を用いてメタデータの構造を定義していく。書誌や文献のための語彙としては、Dublin Core が代表的な存在である。Dublin Core は多くの分野で利用されている語彙セットだが、語彙が 15 種類に限られており、すべての書誌情報を表現することができない。そこで、Dublin Core を補完する語彙セットとして PRISM を用いている。PRISM は論文情報に特化した語彙セットとして、Nature Publishing など海外の学術情報サービスでも利用されている。

CiNii では、検索機能と書誌提供機能で同じメタデータを用いるのではなく、ユーザの要求に合わせて語彙・構造を変えている。CiNii の検索機能は、キーワードを入力して検索結果の一覧を得るといった点では標準的なウェブ検索エンジンと変わらない。よって、CiNii のメタデータを利用する開発者にとっては、多くの検索エンジンのウェブ API と同じ挙動をすることが望ましい。

そこで、検索エンジンのウェブ API の振る舞いを規定する OpenSearch と呼ばれる規格に沿ってメタデータの設計を行った。OpenSearch は検索結果を RSS (ブログやニュースサイトの新着情報を配信するフォーマット) 等で提供するプロトコルである。OpenSearch を用いることで、ブラウザ内蔵の検索ボックスから直接検索を行うことや、定期的に検索を行って最新のコンテンツを自動的に入手することができるようになる。

複数の語彙セットを用いる場合、同じような情報を表現するための語彙が各セットに含まれていることがある。例えば、CiNii の OpenSearch で論文のタイトルを表現するには、RSS に含まれる rss:title、Dublin Core の dc:title、PRISM の prism:title の 3 種類がある。このうちのどれを採用するかが問題となるが、CiNii ではよりポピュラーな語彙セットである RSS のものを選択し、RSS にない語彙は Dublin Core を、いずれにもない語彙に限って PRISM を選択するようにしている。意味的には rss:title よりも prism:title の方がより厳密な定義だが、多くのブラウザは rss:title のみを解釈するため、使い勝手や利便性を優先してこのような設計となっている。

検索機能と異なり、個別の書誌情報を提供する際には、他のサービスとの類似性を考える必要がなく、できるだけ正確な書誌構造をメタデータとして表現することが優先される。そこで、書誌メタデータでは RSS の語彙は利用せず、Dublin Core と PRISM を用いてモデル化を行っている。

3.4 メタデータの利用促進

CiNii ではメタデータの利用促進策として、2009 年度と 2010 年度の 2 回に渡ってウェブ API コンテストを実施し、メタデータを利用したアプリケーションやサービスを募った。成果としては初年度の応募が 24 件、2 年目は 33 件であった。提案としては書誌

情報を新たなインターフェイスで表示・操作させるものや、既存の図書館システムや文献検索システムの中に CiNii の情報を取り込むものなど、実用的な作品が見られた。

国立情報学研究所では、携帯電話向けのインターフェイスの提供や、研究者の成果公開サービス Researchmap において論文情報の自動取り込みにメタデータを活用している。

これらの施策の結果、メタデータへのアクセスは順調に増加し、2011 年 12 月現在では CiNii に対する総アクセス数の 40%~50%を占めるに至っている。

4. 著者情報とメタデータ

4.1 著者情報の現状

書誌情報に含まれる著者情報については、図書のような著者名典拠が整備されておらず、データベースには著者名の文字列が格納されているにすぎない。この中には同名異人や、旧姓あるいはデータベース入力時のタイプミスによる同一人物の分離などが多数生じており、結果として正確に著者が捕捉されているとはいえない。研究者あるいは研究機関の成果を厳密に管理する手法として、研究者 ID の導入が注目されており、トムソン・ロイターが提供している Researcher ID を筆頭にさまざまな規格が生まれるとともに、ID の国際連携の枠組みである ORCID の活動が始まっている。

一方、これまでに蓄積してきた書誌情報については、研究者 ID は付加されておらず、根本的な解決にはならない。CiNii に限ってもデータベースに含まれている名前の異なり数は 800 万を超える。この 1 つ 1 つについて同一人物かどうかを手で判定していくことは困難である。

4.2 機械処理による著者 ID の生成

CiNii では、この問題の解決策として、著者名に対する機械処理によって ID を付与することを行っている。機械処理による名寄せ処理は下記のように行われる。

- 1) すべての書誌情報（1500 万件）について、著者名を取り出す。
- 2) 任意の書誌中の著者名について、同じ著者名が含まれる書誌を検索する。
- 3) 元の書誌と検索で選ばれた書誌を比較し、共著者の類似性やタイトル・所属の文字列の類似性をもとに同一人物かどうかを判定する。同一人物であればそれぞれの著者名をリンクする。
- 4) これらの処理を繰り返すと、書誌集合の中の同一人物がネットワーク化される。最終的にこのネットワークを単一の著者とみなして ID を発行する。

なお、同一人物判定には機械学習アルゴリズムの一種であるサポートベクタマシン

(SVM) を用いる。機械処理による名寄せにおいては、本来は同一であるべき人物が複数に分離してしまう事象（未統合）や、別の人物であるべき情報が同一化されるといふ事象（誤統合）が起こりうる。この2つの事象はトレードオフの関係にあるが、CiNii では極力誤統合を避けるようチューニングを行ったうえで、未統合についてはフィードバック機能を用意し、ユーザからの報告に基づいて統合を進めるというアプローチを取っている。

上記のアルゴリズムでは、機械処理の結果として得られたネットワークに連番を振ったという形になり、ID 自体は特定の意味を持たない。ただし、別のデータベースである科研費データベースには研究者に ID（科研費番号）が付加されており、このデータベースが格納している報告書の中に CiNii 掲載の論文情報が含まれている場合には、この情報をキーにしてクラスタの ID を科研費番号に置き換えることで信頼性の高い ID 体系との紐付けを可能にしている。

一連の処理によって生成された著者 ID と、この ID に紐付けされた書誌情報は、CiNii 著者検索の機能として提供される。また、これらの情報はメタデータとして提供されている。

5. 課題と展望

本論文では CiNii で提供しているメタデータについて紹介してきた。2 節で述べたとおり、学術情報については組織的かつ継続的にメタデータが作成されているため、提供の容易さは他の情報源とは比較にならない。しかしながら、著者情報など構造化がなされていないものについては今後さまざまな努力が必要になる。著者情報の他にも所属情報やキーワードなど、統制されていない情報は数多い。こういった情報の価値がますます高まることは間違いなく、何らかの対応方針を持つことが急務であると思われる。

こういった内部的な構造化とは別に、他の分野で構造化されている情報とのマッチングも重要な課題である。近年、セマンティックウェブの分野ではあらゆる情報をメタデータ化し、他の情報源のメタデータと接続して巨大な情報空間を構築していく **Linked Open Data** が注目されている。

とくに、Wikipedia の情報を RDF 化して提供している DBpedia を中心に、さまざまな情報が相互にリンクされつつある。手持ちのメタデータを **Linked Open Data** に参加させるためには、他の情報源上に存在する同一概念のメタデータとのリンクが必要である。文字列の同一性でマッチングすることも可能だが、大量かつ精密にリンクさせるためには何らかの機械処理やユーザ参加の仕組みが必要であると思われる。いずれにせよ、学術情報分野で行われてきた組織的なデータ作成と合わせて、メタデータの世界を豊かに

する活動は今後も重要である。

参考文献

[1] 大向一輝. 学術情報サービスのユーザモデルとファインダビリティ. 情報の科学と技術. Vol.58, No.12, pp.595-601, 2008.

[2] 大向一輝. 学術情報プラットフォームとしての CiNii. カレントアウェアネス. No.301, CA1691, pp.2-4, 2009.

[3] Linked Data - Connect Distributed Data across the Web. <http://linkeddata.org/>

情報の構造とメタデータ

発行日: 2012年2月24日

発行者: 全国共同利用・共同研究拠点「人文学諸領域の複合的共同研究
国際拠点」

住所: 〒606-8501 京都大学人文科学研究所

印刷:

